Electronic Colloquium on Computational Complexity, Report No. 64 (2007)

Direct Product Theorems for Communication Complexity via Subdistribution Bounds

(Preliminary version)

Rahul Jain *
U. Waterloo

Hartmut Klauck †
Goethe-Universität Frankfurt

Ashwin Nayak ‡
U. Waterloo & Perimeter

June 19, 2007

Abstract

A basic question in complexity theory is whether the computational resources required for solving k independent instances of the same problem scale as k times the resources required for one instance. We investigate this question in various models of classical communication complexity.

We define a new measure, the *subdistribution bound*, which is a generalization of the well-studied rectangle or corruption bound in communication complexity. We prove that the one-way version of this bound tightly captures the one-way public-coin randomized communication complexity of any relation. More importantly, we show that the bound satisfies the strong direct product property under product distributions, for both one- and two-way communication. This way we recover and generalize, in one shot, several recent results on the direct product question, including those due to Klauck *et al.* [KvdW04], Beame *et al.* [BPSW07], Gavinsky [Gav06], and de Wolf [dW06].

The simplicity and broad applicability of our technique is perhaps an indication of its potential to solve yet more challenging questions regarding the direct product problem.

*School of Computer Science, and Institute for Quantum Computing, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada. Email: rjain@cs.uwaterloo.ca. Research supported in part by ARO/NSA USA.

†Institut für Informatik, Goethe-Universität Frankfurt, 60054 Frankfurt am Main, Germany. E-mail: hklauck@gmail.com. Supported by DFG grant KL 1470/1.

‡Department of Combinatorics and Optimization, and Institute for Quantum Computing, University of Waterloo, 200 University Ave. W., Waterloo, ON N2L 3G1, Canada. E-mail: anayak@math.uwaterloo.ca. Research supported in part by NSERC Canada, CIFAR, MITACS, CFI, OIT, and an ERA from the Province of Ontario. A.N. is also Associate Member, Perimeter Institute for Theoretical Physics, Waterloo, Canada. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

ISSN 1433-8092

1 Introduction

Consider two parties, Alice and Bob, who wish to communicate (classically) to solve several instances of the *same* computational problem. The problem is modeled as a relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, and Alice receives an input $x \in \mathcal{X}$, and Bob an input $y \in \mathcal{Y}$. The goal is to find an element $z \in \mathcal{Z}$ that satisfies the relation, i.e., z is such that $(x, y, z) \in f$. Given a communication protocol to solve f, a straightforward method for solving k instances of f is to run the protocol independently on each problem instance. This method has complexity that scales as k times the complexity of the original protocol. Moreover, when the protocol is randomized, and is guaranteed to succeed with probability at least $2/3$, then the probability of simultaneously succeeding on all k instances is only guaranteed to be at least $(2/3)^k$. A basic question in complexity theory is whether this method of solution is essentially optimal. A proof of its optimality is called a *strong direct product* theorem.

Direct product results and their variants appear in many different areas of complexity theory, ranging from hardness amplification in the theory of pseudo-randomness (see, e.g., [GNW95]), to parallel repetition in interactive proof systems (see, e.g., [Raz98, CSUU07]), to time-space tradeoffs in concrete models of computation (for some recent examples, see [Aar04, KvdW04]).

Although they seem highly plausible, it is well-known that strong direct product results fail to hold for several modes of communication and computation. We concentrate on the setting of communication complexity. For example, testing the equality of $k = \log n$ pairs of n-bit strings with a constant-error private-coin communication protocol has complexity $\mathrm{O}(k \log k + \log n) = \mathrm{O}(\log n \log \log n)$ (see, e.g., [KN97, Example 4.3, page 43]), where we might expect a complexity of $\Omega(k \log n) = \Omega(\log^2 n)$. Similarly, Shaltiel [Sha03] gives an example for which a strong direct product result fails to hold for average case (i.e., distributional) communication complexity.

Notwithstanding the abovementioned counterexamples, various forms of direct product result have been discovered in special cases. Early attempts at the question can be found in [IRW94], and the references therein. Parnafes, Raz, and Wigderson [PRW97] prove a direct product result for "collections" of protocols. In their result the bound on the success probability, however, is only shown to behave like $2^{-\Omega(k/c)}$ for the communication complexity c of the problem at hand. Shaltiel [Sha03] proves a strong direct product property in cases where the discrepancy method is used under the uniform distribution; Klauck, Špalek, and de Wolf [KvdW04] prove it for the quantum communication complexity of Set Disjointness; Beame, Pitassi, Segerlind, and Wigderson [BPSW07] prove it in cases where the *rectangle* or *corruption bound* is tight under product distributions; and Gavinsky [Gav06] proves it for the one-way complexity of a certain class of relational problems. The result by Beame *et al.* for instance allows the conclusion that solving k instances of Set Disjointness with communication complexity $\mathrm{o}(k\sqrt{n})$ has success probability $2^{-\Omega(k)}$. Recently, de Wolf [dW06] proved a strong direct product theorem for the one-way public-coin randomized communication complexity of the Index function. This can also be used to handle the one-way complexity of Set Disjointness via a reduction.

Whether the strong direct product theorem holds in general for public-coin randomized protocols remains a frustrating open question in communication complexity theory. Research on weaker types of property, namely the *direct sum* or the *weak direct product* property, has met with better success.

A direct sum theorem states that solving k instances with constant probability of success incurs at least k times the cost of solving 1 instance. (A strong direct product theorem would show that even with probability of success that is exponentially small in k, the cost would be k times the cost of solving one instance.) Direct sum results have met with better success than general strong direct product theorems.

For deterministic protocols it is known that k times the square root of the deterministic complexity of a function f is needed to compute k instances of f (see, e.g., [KN97, Exercise 4.11, page 46]). It is also

straightforward to show that the deterministic *one-way* communication complexity of every function f has the direct sum property. For randomized protocols, Chakrabarti, Shi, Wirth, and Yao [CSWY01] give a lower bound for the direct sum problem in the simultaneous message (SMP) model in terms of "information cost". This has also been extended to two-way classical and quantum communication [BYJKS04, JRS03b].

Jain, Radhakrishnan, and Sen [JRS05b] show a tight direct sum theorem for the one-way and SMP models for both quantum and randomized classical communication, along with a weak direct sum result for two-way communication. In other work, Jain, Radhakrishnan, and Sen [JRS03a] give a direct sum type lower bound for bounded round private-coin protocols in terms of the average case communication complexity under product distributions. Harsha, Jain, McAllester, and Radhakrishnan [HJMR07] have strengthened the latter lower bound by reducing to a large extent its dependence on the number of rounds. Recently, Pătraşcu and Thorup [PT06] used direct sum type results to prove improved lower bounds for approximate near-neighbour search in the cell probe model.

Another type of result is a *weak direct product theorem.* In such a result one shows that the success probabilty of solving k instances of a problem with the resources needed to solve one instance (with probability $2/3$) goes down exponentially with k. Klauck [Kla04] shows such a result for the rectangle/corruption bound under arbitrary distributions, leading to the conclusion that solving k instances of Set Disjointness with communication complexity $o(n)$ is possible only with success probability $2^{-\Omega(k)}$.

In this article, we define a new measure of hardness of computing a function (and more generally a relation) in a distributed fashion, which we call its *subdistribution complexity*. In fact, subdistribution complexity is a relaxation of the well-studied rectangle/corruption bound from communication complexity. It gives us a better handle on the direct product problem without weakening the rectangle bound. In the setting of public-coin randomized *one-way* communication, we show that this measure tightly characterizes the communication complexity of any relation. More importantly, we show that subdistribution complexity satisfies the strong direct product property under product distributions. In particular, we recover strong direct product theorems for problems whose complexity is captured by the rectangle/corruption bound under product distributions.

Our proof of the strong direct product property belongs to a line of work based on the powerful *substate theorem* due to Jain, Radhakrishnan, and Sen [JRS02], and a closely related notion, the *relative co-min-entropy* of two distributions (we present formal definitions and statements in Section 2.4). It provides a simple and uniform information-theoretic explanation of recent works due to Beame *et al.* [BPSW07] (and a consequence of their result independently due to Klauck *et al.* [KvdW04, Theorem 20]), Gavinsky [Gav06], and de Wolf [dW06], simultaneously improving and generalizing the last two. Our methods also extend to give a lower bound for the simultaneous message passing model. These consequences are perhaps an indication of the wider applicability of the subdistribution approach.

Finally we investigate some applications of our results. Gavinsky [Gav06] used a strong direct product theorem to show that quantum protocols for relations such as one based on the hidden matching problem [BYJK04] need a certain minimum amount of entanglement to be computable optimally by one-way protocols with classical communication. This shows that unlike public randomness, the amount of entanglement cannot be decreased to being sublinear in the input length without either worsening the communication complexity or changing the protocol massively. By employing our theorem for one-way communication we strengthen this result and remove some logarithmic factors in the bounds. Beame *et al.* [BPSW07] use their direct product theorem to establish a bound for the disjointness problem in a restricted *number on the forehead* model of multiparty communication complexity. A larger lower bound for Set Disjointness in the one-way version of this model follows from a result due to Wigderson [BHK01, Section 9.3] on a layered pointer jumping problem. We give a simpler proof based on our direct product result, in terms of a larger class of functions.

Organization of the paper

We follow standard terminology and notation in communication complexity. For completeness, this is summarized in Section 2.1. We define our notation, and review basic information theory in Section 2.4. In Section 3, we present a characterization of one-way communication complexity in terms of what we call the one-way subdistribution bound. In Section 4 we present direct product results in the setting of two-way communication, and describe how this extends to one-way and SMP protocols as well. In Section 5 we present some applications of our direct product theorems. Some proofs are deferred to the appendix.

2 Preliminaries

2.1 Communication complexity

In this section we briefly describe the model of communication complexity. For a comprehensive introduction to the subject we refer the reader to the text by Kushilevitz and Nisan [KN97].

We consider the two-party model of communication. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be finite sets, and let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. In a two-party communication protocol the parties, say Alice and Bob, get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. They alternately send messages to each other with the goal of determining an element $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. We assume that for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ given as input, there is at least one $z \in \mathcal{Z}$ such that $(x, y, z) \in f$.

2.2 One-way communication

We first consider the *one-way* model of communication, in which there is a single message, from Alice to Bob at the end of which Bob determines the answer z from the single message from Alice, and his input y. (In the one-way protocols we consider, the single message is always from Alice to Bob.) Let $0 \leq \epsilon < 1/3$, and let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. We let $\mathsf{D}_\epsilon^{1,\mu}(f)$ represent the *distributional one-way communication complexity* of f under μ with expected error ϵ, i.e., the communication of the best private-coin one-way protocol for f, with *distributional error* (average error over the coins and the inputs) at most ϵ under μ. We note that $\mathsf{D}_\epsilon^{1,\mu}(f)$ is achieved by a deterministic one-way protocol, and will henceforth restrict ourselves to deterministic protocols in the context of distributional communication complexity. We let $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(f)$ represent the public-coin *randomized one-way communication complexity* of f with worst case error ϵ, i.e., the communication of the best public-coin randomized one-way protocol for f with error for each input (x, y) being at most ϵ. The analogous quantity for private coin randomized protocols is denoted by $\mathsf{R}_\epsilon^1(f)$. The following is a consequence of the *min-max* theorem in game theory [KN97, Theorem 3.20, page 36].

Lemma 2.1 (Yao principle) $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(f) = \max_\mu \mathsf{D}_\epsilon^{1,\mu}(f)$.

The communication complexity of a relation may reduce significantly when μ is restricted to product distributions over $\mathcal{X} \times \mathcal{Y}$. We define $\mathsf{R}_\epsilon^{1,\square}(f) \stackrel{\Delta}{=} \max_{\mu \text{ product}} \mathsf{D}_\epsilon^{1,\mu}(f)$.

The VC-dimension of a boolean function f is an important combinatorial concept and has close connections with the one-way communication complexity of f.

Definition 2.1 (Vapnik-Chervonenkis (VC) dimension) *A set S is said to be shattered by a set $\mathcal{G}$ of boolean functions from S to $\{0,1\}$, if $\forall R \subseteq S, \exists g_R \in \mathcal{G}$ such that $\forall s \in S, (s \in R) \iff (g_R(s) = 1)$.*

Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a boolean function. For all $x \in \mathcal{X}$ let $f_x : \mathcal{Y} \to \{0,1\}$ be defined as $f_x(y) \stackrel{\Delta}{=} f(x,y), \forall y \in \mathcal{Y}$. Let $\mathcal{F} \stackrel{\Delta}{=} \{f_x : x \in \mathcal{X}\}$. Then the Vapnik-Chervonenkis dimension of f is defined as $\mathsf{VC}(f) \stackrel{\Delta}{=} \max_{S \subseteq \mathcal{Y}} \{|S| : S \text{ is shattered by } \mathcal{F}\}$.

Kremer, Nisan, and Ron [KNR99, Theorem 3.2] relate VC-dimension to communication complexity. The tighter dependence (stated below) of the communication complexity on the error ϵ in the communication protocol appears in Ambainis, Nayak, Ta-Shma, and Vazirani [ANTSV99, Theorem 1.1].

Theorem 2.2 ([KNR99, ANTSV99]) *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a boolean function, and let $\epsilon \in (0, 1/2)$. Then there is a universal constant κ_0 such that*

$$(1 - \mathsf{H}_2(\epsilon)) \cdot \mathsf{VC}(f) \quad \leq \quad \mathsf{R}^{1,[]}_{\epsilon}(f) \quad \leq \quad \kappa_0 \cdot \frac{1}{\epsilon} \log \frac{1}{\epsilon} \cdot \mathsf{VC}(f),$$

where $\mathsf{H}_2(p) = -p \log_2 p - (1-p) \log_2 (1-p)$ is the binary entropy function defined on $[0,1]$.

2.3 Two-way communication and the SMP model

Next we consider two-way protocols, which are defined analogously. These allow communication between Alice and Bob over multiple rounds at the end of which both parties output the *same* element $z \in \mathcal{Z}$ *that depends upon the transcript of the protocol alone*. Following Kushilevitz and Nisan [KN97], we assume Alice and Bob disregard their inputs when they determine their output. This is unlike in one-way protocols, where we (necessarily have to) allow Bob to determine his output from Alice's message *and his input*. The relevant complexity measures for this model are denoted $\mathsf{D}^{\mu}_{\epsilon}(f)$, $\mathsf{R}^{[]}_{\epsilon}(f)$, $\mathsf{R}^{\mathsf{pub}}_{\epsilon}(f)$, $\mathsf{R}_{\epsilon}(f)$ etc. (without the superscript '1'). Lemma 2.1 holds for two-way protocols *mutatis mutandis*.

A two-way communication protocol in which the two parties consider their inputs for the computation of their respective outputs may be converted into the form above. One party may send an additional message consisting of his/her output. The consequent increase in communication complexity is at most $\log |\mathcal{Z}|$.

We also consider the *Simultaneous message passing* (SMP) model of communication. In this model, Alice and Bob receive inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. They each send one message to a third party, called the "referee". The goal of the referee is to output an element $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. In the SMP model, by public coin protocols we mean protocols in which Alice, Bob and the referee all have access to the same source of random coins. The relevant complexity measures are denoted $\mathsf{D}^{\|,\mu}_{\epsilon}(f)$, $\mathsf{R}^{\|,\mathsf{pub}}_{\epsilon}(f)$ etc. Lemma 2.1 also holds for SMP protocols *mutatis mutandis*.

2.4 Information theory

In this section we present some information theoretic notation, definitions and facts that we use in our proofs. For an introduction to information theory, we refer the reader to the text by Cover and Thomas [CT91]. Most of the facts stated in this section without proof may be found in this book.

All logarithms in the article are taken with base 2. For an integer $t \geq 1$, $[t]$ represents the set $\{1, \ldots, t\}$. For square matrices P, Q, by $Q \geq P$ we mean that $Q - P$ is positive semi-definite. For a matrix A, $\|A\|_1$ denotes its ℓ_1 norm.

Specializing from the quantum case, we view a discrete probability distribution P as a positive semi-definite trace one diagonal matrix indexed by its (finite) sample space. For a distribution P with support on set $\mathcal{X}$, and $x \in \mathcal{X}$, $P(x)$ denotes the (x,x) diagonal entry of P, and $P(\mathcal{E}) = \sum_{x \in \mathcal{E}} P(x)$ denotes the probability of the event $\mathcal{E} \subseteq \mathcal{X}$. For a random variable X, we sometimes also let X represent its distribution.

Let $\mathcal{X}, \mathcal{Y}$ be sets and let P be a distribution with support on $\mathcal{X} \times \mathcal{Y}$. For $x \in \mathcal{X}$, we define $P(x) = \sum_{y \in \mathcal{Y}} P(x, y)$, the probability of x in the marginal distribution on $\mathcal{X}$; $P(y)$ is similarly defined for $y \in \mathcal{Y}$. Further, if $y \in \mathcal{Y}$ occurs with probability $P(y) > 0$, we define $P(x|y) = \frac{P(x,y)}{P(y)}$, the conditional probability given the event $\mathcal{X} \times \{y\}$. The distribution P is said to be a *product* distribution if there are distributions P_X, P_Y on $\mathcal{X}, \mathcal{Y}$ respectively such that $P = P_X \otimes P_Y$, where $\otimes$ denotes the tensor product operation. Equivalently, for a product distribution, $P(x, y) = P(x) \cdot P(y)$.

For distributions P, Q, $\mathsf{S}(P \,\|\, Q) \stackrel{\Delta}{=} \mathrm{Tr}(P \log P - P \log Q)$ is called the *relative entropy* or the *Kullback-Leibler divergence* between them. It is known that relative entropy is jointly convex in its arguments.

Lemma 2.3 *Let P_1, P_2, Q_1, Q_2 be probability distributions. Then for $r \in [0, 1]$,*

$$\mathsf{S}(rP_1 + (1-r)P_2) \,\|\, rQ_1 + (1-r)Q_2) \quad \leq \quad r\mathsf{S}(P_1 \,\|\, Q_1) + (1-r)\mathsf{S}(P_2 \,\|\, Q_2).$$

Relative entropy satisfies the following *chain rule*:

Lemma 2.4 (Chain rule for relative entropy) *Let $M_1, \ldots, M_k$ and $N_1, \ldots, N_k$ be collections of random variables. For $1 \leq i \leq k$, let $\tilde{M}_i$ represent the random variable $M_1 \ldots M_{i-1}$. Similarly define $\tilde{N}_i$. Then*

$$\mathsf{S}(M_1 \ldots M_k \,\|\, N_1 \ldots N_k) \quad = \quad \sum_{i=1}^{k} \mathbb{E}_{m \sim \tilde{M}_i} [\mathsf{S}(M_i | \tilde{M}_i = m \,\|\, N_i | \tilde{N}_i = m)].$$

Lemma 2.5 *Let $M_1 M_2$ be random variables and let $N_1 N_2$ be mutually independent random variables. Then*

$$\mathsf{S}(M_1 M_2 \,\|\, N_1 N_2) \quad \geq \quad \mathsf{S}(M_1 \,\|\, N_1) + \mathsf{S}(M_2 \,\|\, N_2).$$

Proof: Using the chain rule (Lemma 2.4), the independence of N_1 and N_2, and finally convexity (Lemma 2.3), we have

$$\begin{aligned} \mathsf{S}(M_1 M_2 \,\|\, N_1 N_2) &= \mathsf{S}(M_1 \,\|\, N_1) + \mathbb{E}_{m \sim M_1} [\mathsf{S}(M_2 | M_1 = m \,\|\, N_2 | N_1 = m)] \\ &= \mathsf{S}(M_1 \,\|\, N_1) + \mathbb{E}_{m \sim M_1} [\mathsf{S}(M_2 | M_1 = m \,\|\, N_2)] \\ &\geq \mathsf{S}(M_1 \,\|\, N_1) + \mathsf{S}(M_2 \,\|\, N_2), \end{aligned}$$

as claimed. ∎

For distributions P, Q, with support on set $\mathcal{X}$, we define

$$\mathsf{S}_\infty(P \,\|\, Q) \quad \stackrel{\Delta}{=} \quad \inf\{c : Q \geq P/2^c\},$$

as the *relative co-min-entropy* of P with respect to Q. This quantity measures what scaling of a distribution "sits inside" another. Note that the relative co-min-entropy of P with respect to the uniform distribution on $\mathcal{X}$ is precisely $\log |\mathcal{X}| - \mathsf{H}_\infty(P)$, where $\mathsf{H}_\infty(P) = \min_x \log \frac{1}{P(x)}$ is the *min-entropy* of P.

The following fact is a special case of Theorem 1(7) in [JRS05a]. It follows directly from the monotonicity of the logarithm function.

Lemma 2.6 *Let P, Q be distributions. Then $\mathsf{S}(P \,\|\, Q) \leq \mathsf{S}_\infty(P \,\|\, Q)$.*

The *substate theorem* [JRS02, Proposition 1] gives us a powerful operational characterization of relative entropy.

Lemma 2.7 (Substate theorem) *Let P, Q be probability distributions over the same finite sample space such that $\mathsf{S}(P \,\|\, Q) \leq c$. Then for all $r > 1$, there exist distributions P_r such that $\|P - P_r\|_1 \leq \frac{2}{r}$ and*

$$(1 - \tfrac{1}{r})\tfrac{P_r}{2^{r(c+1)}} \quad \leq \quad Q \qquad \Leftrightarrow \qquad \mathsf{S}_\infty(P_r \,\|\, Q) \quad \leq \quad r(c+1) + \log \tfrac{r}{r-1}.$$

The following fact is readily verified:

Lemma 2.8 *If P, Q are distributions on the same sample space such that $\|P - Q\|_1 \leq \epsilon$, then for any event $\mathcal{E}$, we have $|P(\mathcal{E}) - Q(\mathcal{E})| \leq \epsilon/2$.*

The following fact may be verified from the definition of relative co-min-entropy.

Lemma 2.9 *Let X_1, X_2, Y_1, Y_2 be random variables. Then $\mathsf{S}_\infty(X_1 \,\|\, Y_1) \leq \mathsf{S}_\infty(X_1 X_2 \,\|\, Y_1 Y_2)$.*

Random variables X, Y, Z form a *Markov chain*, represented as $X \to Y \to Z$, iff for all x, y, the conditional random variable $Z|(XY = xy)$ is equal to $Z|(Y = y)$. The following lemma may be verified readily from this definition.

Lemma 2.10 *If $X \to Y \to Z$ is a Markov chain, then so is $Z \to Y \to X$.*

We use various forms of the *Markov inequality* from probability theory [CT91] in our arguments without proof.

3 A characterization of one-way communication complexity

In this section we present new, tight upper and lower bounds for randomized one-way communication complexity in terms of what we call the one-way subdistribution bound.

We start with some definitions. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 \leq \epsilon \leq 1/3$.

Definition 3.1 *Let λ, μ be distributions on $\mathcal{X} \times \mathcal{Y}$.*

1. **ϵ-monochromatic:** *We say that the distribution λ is* ϵ-monochromatic *for f if there exists $z \in \mathcal{Z}$ such that $\Pr_{XY \sim \lambda}[(X, Y, z) \in f] \geq 1 - \epsilon$.*
2. **one-way ϵ-monochromatic:** *We call λ* one-way ϵ-monochromatic *for f if there is a function $g : \mathcal{Y} \to \mathcal{Z}$ such that $\Pr_{XY \sim \lambda}[(X, Y, g(Y)) \in f] \geq 1 - \epsilon$.*
3. **one-message-like:** *We call λ* one-message-like *for μ if for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, whenever $\lambda(x) > 0$, we have $\mu(x) > 0$ and $\lambda(y|x) = \mu(y|x)$.*

This definition is motivated by properties of distributions that arise in (one-way) communication protocols. The distribution λ is one-way ϵ-monochromatic precisely when there is a one-way communication protocol for f with *zero* communication cost and distributional error at most ϵ under λ. Suppose $\mathcal{P}$ is a deterministic one-way protocol for f, with a single message from Alice to Bob. Let X, Y denote random variables with joint distribution μ, corresponding to Alice and Bob's inputs respectively. For any message string m in $\mathcal{P}$, we may readily verify that the conditional distribution $XY|(M = m)$ is one-message-like for μ. Furthermore, suppose the distributional error made by $\mathcal{P}$ is at most ϵ. Then, for any $\delta \in (0, 1]$, the distribution of $XY|(M = m)$ is one-way $\frac{\epsilon}{\delta}$-monochromatic for f with probability at least $1 - \delta$ over the messages m.

We now define the *subdistribution bounds* on one-way communication complexity.

Definition 3.2 (One-way subdistribution bound & one-way product subdistribution bound) *For a distribution μ over $\mathcal{X} \times \mathcal{Y}$, let $\mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon, \mu) \stackrel{\Delta}{=} \min_{\lambda} \mathsf{S}_{\infty}(\lambda \,\|\, \mu)$, where λ ranges over all distributions which are both one-message-like for μ and one-way ϵ-monochromatic for f. We define the* one-way subdistribution bound *as $\mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon) \stackrel{\Delta}{=} \max_{\mu} \mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon, \mu)$, where μ ranges over all distributions on $\mathcal{X} \times \mathcal{Y}$. When the maximization is restricted to product distributions μ, we refer to the quantity as the* one-way product subdistribution bound $\mathsf{sub}^{1,\square}_{\mathsf{B}}(f, \epsilon)$.

Remark: First, in the above definition, the subscript B is used to emphasize the fact that that in the definition of one-way ϵ-monochromatic, we allow for different values of output depending upon Bob's input in a zero-communication protocol for f under distribution λ. Second, note that a distribution λ which is one-way for a product distribution μ is itself a product distribution. Third, if we took the distribution λ to range over μ conditioned upon one-way rectangles (i.e., rectangles of the form $S \times \mathcal{Y}$, where $S \subseteq \mathcal{X}$), then we would get a one-way variant of the rectangle or corruption bound in communication complexity as introduced by Yao, and applied by Razborov and others. (See, e.g., Beame *et al.* [BPSW07] for a formal definition of the bound.) We elaborate on the precise connection between the rectangle bound and subdistribution bound in Section 4.

Moving on to our characterization theorem, we now show that the one-way communication complexity of a relation is always larger than the subdistribution bound.

Lemma 3.1 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 \leq \epsilon \leq 1/3$ and $k > 0$ be non-negative real numbers. Then*

$$\mathsf{R}^{1,\mathsf{pub}}_{\epsilon(1-2^{-k})}(f) \quad \geq \quad \mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon) - k.$$

Proof: For any distribution μ on $\mathcal{X} \times \mathcal{Y}$, we show

$$\mathsf{D}^{1,\mu}_{\epsilon(1-2^{-k})}(f) \quad \geq \quad \mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon, \mu) - k. \tag{1}$$

Maximizing over μ, and appealing to the Yao min-max principle (Lemma 2.1) and the definition of $\mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon)$ we get our bound.

Let $c \stackrel{\Delta}{=} \mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon, \mu)$. If $\lfloor c - k \rfloor \leq 0$, Eq. (1) holds vacuously. Otherwise, let $\mathcal{P}$ be a deterministic one-way protocol with communication $\lfloor c - k \rfloor$. Let the random variables X, Y with joint distribution μ represent the inputs of Alice and Bob respectively. Let M represent the correlated random variable corresponding to Alice's message. For a message string m with $p_m \stackrel{\Delta}{=} \Pr[M = m] > 0$ let ϵ_m denote the probability of error of $\mathcal{P}$ conditional on $M = m$. Let $\mathcal{M}$ be the set of messages m such that $p_m > 2^{-c}$. Since there are at most 2^{c-k} messages, we get that $\sum_{m \notin \mathcal{M}} p_m \leq 2^{-k}$. Let λ_m be the distribution of $XY|(M = m)$. For $m \in \mathcal{M}$, we have $\mathsf{S}_{\infty}(\lambda_m \,\|\, \mu) < c$. Since λ_m is one-message-like for μ, from the definition of $\mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon, \mu)$ we have $\epsilon_m > \epsilon$. Hence the overall error of the protocol $\mathcal{P}$ is $> \epsilon(1 - 2^{-k})$. Therefore, by its definition $\mathsf{D}^{1,\mu}_{\epsilon(1-2^{-k})}(f) > \lfloor c - k \rfloor$, which is the communication in $\mathcal{P}$. ∎

For the other direction, we first show that for a relation f with low subdistribution complexity, any distribution μ may be decomposed into a small number of one-message-like distributions that are one-way ϵ-monochromatic for f.

Lemma 3.2 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation, $0 \leq \epsilon < 1$, and $c \stackrel{\Delta}{=} \mathsf{sub}^{1}_{\mathsf{B}}(f, \epsilon)$. For any distribution μ on $\mathcal{X} \times \mathcal{Y}$, and $\delta \in (0, 1]$, there exists an integer $r \geq 1$, distributions $\{\lambda_j, j \in [r+1]\}$ on $\mathcal{X} \times \mathcal{Y}$ and numbers $\{p_j, j \in [r+1], 0 \leq p_j \leq 1\}$ such that:*

1. *$\forall j \in [r+1], \lambda_j$ is one-message-like for μ and one-way ϵ-monochromatic for f,*

2. $\mu = \sum_{j=1}^{r+1} p_j \lambda_j$,
3. $p_{r+1} \leq \delta$, *and*
4. *For* $j \in [r], p_j > 2^{-c}\delta$ *and* $r < \frac{2^c}{\delta}$.

Proof: By hypothesis, $c = \mathsf{sub}^1_{\mathsf{B}}(f, \epsilon) = \max_\nu \mathsf{sub}^1_{\mathsf{B}}(f, \epsilon, \nu)$. This means for every distribution ν there exists a distribution θ_ν with the properties that θ_ν is one-message-like for ν, one-way ϵ-monochromatic for f, and $\mathsf{S}_\infty(\theta_\nu \,\|\, \nu) \leq c$.

We obtain the distributions $\lambda_1, \ldots, \lambda_{r+1}$ in the decomposition of μ inductively:

- Let $\mu_1 \stackrel{\Delta}{=} \mu, \lambda_1 \stackrel{\Delta}{=} \theta_{\mu_1}$ and $p_1 \stackrel{\Delta}{=} 2^{-\mathsf{S}_\infty(\lambda_1 \,\|\, \mu)}$.
- Suppose for some $j \geq 1$, the distributions $\lambda_1, \ldots, \lambda_j$ have been obtained. Let $q_{j+1} \stackrel{\Delta}{=} \left\| \mu - \sum_{k=1}^{j} p_k \lambda_k \right\|_1$, and $\mu_{j+1} \stackrel{\Delta}{=} \frac{1}{q_{j+1}} \left(\mu - \sum_{k=1}^{j} p_k \lambda_k \right)$.

  In case $q_{j+1} > \delta$, we let $\lambda_{j+1} \stackrel{\Delta}{=} \theta_{\mu_{j+1}}$ and $p_{j+1} \stackrel{\Delta}{=} q_{j+1} 2^{-\mathsf{S}_\infty(\lambda_{j+1} \,\|\, \mu_{j+1})}$ and move to $j+2$.

  In case $q_{j+1} \leq \delta$ we stop the process and let $\lambda_{j+1} \stackrel{\Delta}{=} \mu_{j+1}, p_{j+1} \stackrel{\Delta}{=} q_{j+1}$ and $r \stackrel{\Delta}{=} j$.

Part 1 of the lemma is immediate from our construction and the following properties of the 'one-message-like' relation. Let ν, σ, τ be distributions over $\mathcal{X} \times \mathcal{Y}$.

- If σ is one-message-like for ν, and $p \geq 0$ is such that $p\sigma \leq \nu$, the distribution $\frac{\nu - p\sigma}{\|\nu - p\sigma\|_1}$ is also one-message-like for ν.
- The distributions that are one-message-like for a fixed distribution ν form a convex set. I.e., if σ, τ are one-message-like for ν, the distribution $p\sigma + (1-p)\tau$ is also one-message-like for ν for any $0 \leq p \leq 1$.
- The 'one-message-like' relation is transitive. I.e., if σ is one-message-like for τ, and τ is one-message-like for ν, then σ is one-message-like for ν.

Parts 2 and 3 of the lemma may be verified from our construction. For Part 4 we note that for any $1 \leq j \leq r$,

$$p_j \quad = \quad q_j \, 2^{-\mathsf{S}_\infty(\lambda_j \,\|\, \mu_j)} \quad > \quad \delta 2^{-c}.$$

Since $\sum_{j=1}^{r} p_j \leq 1$, we get $r < 2^c/\delta$. ∎

Using the above decomposition of distributions, we can design efficient protocols for relations with small subdistribution complexity.

Lemma 3.3 *Let* $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ *be a relation, and* $0 \leq \epsilon \leq 1/6$ *and* $0 < \delta \leq 1/6$. *Then,*

$$\mathsf{R}^{1,\mathsf{pub}}_{\epsilon+\delta}(f) \quad \leq \quad \mathsf{sub}^1_{\mathsf{B}}(f, \epsilon) + \log \frac{1}{\delta} + 2.$$

Proof: We show that for every distribution μ on $\mathcal{X} \times \mathcal{Y}$,

$$\mathsf{D}^{1,\mu}_{\epsilon+\delta}(f) \quad \leq \quad \mathsf{sub}^1_{\mathsf{B}}(f, \epsilon) + \log \frac{1}{\delta} + 2. \tag{2}$$

The result then follows from the Yao min-max principle (Lemma 2.1).

We exhibit a private coin protocol $\mathcal{P}$ for f whose distributional error under μ is at most $\epsilon + \delta$ and communication is at most $c + \log(1/\delta) + 2$, where $c \stackrel{\Delta}{=} \mathsf{sub}^1_{\mathsf{B}}(f, \epsilon)$. From $\mathcal{P}$ we also get a deterministic protocol with the same communication and distributional error. This implies Eq. (2).

In the protocol $\mathcal{P}$, Alice and Bob start with their inputs XY in distribution μ. Using the decomposition of μ as given by Lemma 3.2, we define a random variable M that is correlated with XY. We then argue that M may be produced from the knowledge of X alone, and therefore be used as a message to derive a protocol with small distributional error.

Let $\mu = \sum_{j \in [r+1]} p_j \lambda_j$ with p_j, λ_j and r as given by Lemma 3.2 for δ as in the statement of this lemma. The random variable M has support in $[r+1]$. The joint distribution of XYM is defined by

$$\Pr[XYM = (x, y, j)] \quad = \quad p_j \, \lambda_j(x, y),$$

for $(x, y, j) \in \mathcal{X} \times \mathcal{Y} \times [r+1]$. Note that $\Pr[M = j] = p_j$ and the distribution of $XY|(M = j)$ is λ_j. Since for all j, the distribution λ_j is one-message-like for μ, we have $Y|(X = x, M = m) \; = \; Y|(X = x)$ for all x, m. Hence $M \to X \to Y$ is a Markov chain. From Lemma 2.10, $Y \to X \to M$ is also a Markov chain. Therefore, the random variable M is a function of X alone, and Alice can generate it using private coins.

To summarize the protocol $\mathcal{P}$, on input x, Alice generates message M as above using private coins, and sends it to Bob. From the construction of XYM, on receiving message j, Bob knows that the conditional distribution on XY is λ_j. On each λ_j with $j \in [r]$ we can ensure that the error of $\mathcal{P}$ is at most ϵ since λ_j is one-way ϵ-monochromatic. On message $r+1$, which occurs with probability at most δ, the error may be as large as 1. Therefore $\mathcal{P}$ has distributional error at most $\epsilon + \delta$ on μ. The communication in $\mathcal{P}$ is bounded by $\lceil \log(r+1) \rceil \leq c + \log(1/\delta) + 2$. ∎

Combining the bounds in Lemmata 3.1 and 3.3 with standard probability amplification techniques, we get our characterization of one-way communication complexity in terms of the subdistribution bound.

Theorem 3.4 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $0 \leq \epsilon \leq 1/6$. There are universal constants κ_1, κ_2 such that*

$$\mathsf{sub}^1_{\mathsf{B}}(f, \epsilon) - 1 \quad \leq \quad \kappa_1 \cdot \mathsf{R}^{1,\mathsf{pub}}_{\epsilon}(f) \quad \leq \quad \kappa_2 \left[\mathsf{sub}^1_{\mathsf{B}}(f, \epsilon) + \log \frac{1}{\epsilon} + 2 \right].$$

Remark: From proofs of Lemma 3.3 and Lemma 3.1, we also conclude that for a distribution μ such that $\mathsf{sub}^1_{\mathsf{B}}(f, \epsilon) = \mathsf{sub}^1_{\mathsf{B}}(f, \epsilon, \mu)$ we have $\mathsf{D}^{1,\mu}_{\epsilon}(f) = \Theta(\mathsf{sub}^1_{\mathsf{B}}(f, \epsilon, \mu))$ (for a constant ϵ). However for other distributions, $\mathsf{sub}^1_{\mathsf{B}}(f, \epsilon, \mu)$ may be much smaller than $\mathsf{D}^{1,\mu}_{\epsilon}(f)$. As an example consider the function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined as $f(x, y) \stackrel{\Delta}{=} x_1 \vee \bigoplus_{i=2}^{n} x_i \wedge y_i$. While the one-way communication required for computing this function with distributional error at most $1/5$ under the uniform distribution U is $\Omega(n)$, we have $\mathsf{sub}^1_{\mathsf{B}}(f, 0, \mathsf{U}) \leq 1$. This is because the distribution with $x_1 = 1$ and remaining bits uniform has 0 error and sits in U with a scaling of $1/2$.

The proof of Theorem 3.4 readily adapts to give a similar relationship between $\mathsf{R}^{1,[]}_{\epsilon}(f)$ and $\mathsf{sub}^{1,[]}_{\mathsf{B}}(f, \epsilon)$.

Theorem 3.5 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $0 \leq \epsilon \leq 1/6$. There are universal constants κ_1, κ_2 such that*

$$\mathsf{sub}^{1,[]}_{\mathsf{B}}(f, \epsilon) - 1 \quad \leq \quad \kappa_1 \cdot \mathsf{R}^{1,[]}_{\epsilon}(f) \quad \leq \quad \kappa_2 \left[\mathsf{sub}^{1,[]}_{\mathsf{B}}(f, \epsilon) + \log \frac{1}{\epsilon} + 2 \right].$$

Since the one-way distributional communication complexity under product distributions of a boolean function is captured by its VC-dimension (Theorem 2.2) both quantities in the above theorem are of the same order as the VC-dimension of f (for constant ϵ). The precise dependence on ϵ may be inferred from the preceding theorems.

Corollary 3.6 *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a boolean function. Let $0 \leq \epsilon \leq 1/6$ be a constant. Then $\mathsf{R}^{1,[]}_{\epsilon}(f) = \Theta(\mathsf{sub}^{1,[]}_{\mathsf{B}}(f, \epsilon)) = \Theta(\mathsf{VC}(f))$.*

4 Direct product theorems for classical communication

4.1 Two-way protocols

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. We define a two-way subdistribution bound under product distributions in a manner analogous to the one-way bound.

Definition 4.1 (Two-way product subdistribution bound) *Let μ be a product distribution on $\mathcal{X} \times \mathcal{Y}$, and $\epsilon \in [0,1]$. Let $\mathsf{sub}^{\square}(f, \epsilon, \mu) \stackrel{\Delta}{=} \min_{\lambda} \mathsf{S}_{\infty}(\lambda \,\|\, \mu)$, where λ ranges over all product distributions that are ϵ-monochromatic for f. We define the* two-way product subdistribution bound *as $\mathsf{sub}^{\square}(f, \epsilon) \stackrel{\Delta}{=} \max_{\mu \text{ product}} \mathsf{sub}^{\square}(f, \epsilon, \mu)$.*

Remark: It is important to restrict λ to product distributions in the definition of $\mathsf{sub}^{\square}(f, \epsilon, \mu)$, even when μ is product. Otherwise, the quantity is at most 1 for boolean functions, even with $\epsilon = 0$: consider the possibly non-product distribution λ that results from conditioning upon $f^{-1}(1)$ or $f^{-1}(0)$, whichever has higher probability under μ. The distribution λ is 0-monochromatic for f, and sits well inside μ, since the event on which we condition has probability $\geq \frac{1}{2}$.

To state the precise connection between the subdistribution bound and the rectangle/corruption bound from communication complexity, we define the latter bound precisely. A rectangle in $\mathcal{X} \times \mathcal{Y}$ is a subset of the form $S \times T$, where $S \subseteq \mathcal{X}, T \subseteq \mathcal{Y}$. For a distribution μ, and an event R, let μ_R denote the conditional distribution of μ given the event R. For a (possibly non-product) distribution μ, define $\mathsf{rec}(f, \epsilon, \mu) \stackrel{\Delta}{=} \min_R \mathsf{S}_{\infty}(\mu_R \,\|\, \mu)$, where R ranges over all rectangles in $\mathcal{X} \times \mathcal{Y}$ such that μ_R is ϵ-monochromatic for f. The rectangle bound maximized over all distributions is well-known to be a lower bound for two-way randomized communication complexity (see [BPSW07, Section 3] for a precise formulation of this bound). When the maximization is restricted to product distributions μ, we get the two-way *product* rectangle bound $\mathsf{rec}^{\square}(f, \epsilon)$. This may be substantially smaller than the unrestricted bound, but is still known to give strong bounds for certain functions. For example, for $f = \mathsf{DISJ_n}$, the set disjointness problem on n-bit inputs, $\mathsf{R}^{\mathsf{pub}}_{1/3}(f) = \Theta(n) = \mathsf{rec}(f, 1/3)$ [KN97, Section 4.6, Lemma 4.49], whereas $\mathsf{rec}^{\square}(f, 1/3)$ and $\mathsf{R}^{\square}_{1/3}(f)$ are both $\mathrm{O}(\sqrt{n} \log n)$ and at least $\Omega(\sqrt{n})$ [BFS86]. It is open whether randomized communication complexity may be super-polynomially larger than distributional communication complexity under product distributions [KN97, Open Problem 3.26], as is the corresponding question between the two rectangle bounds.

The two-way product subdistribution bound is a relaxation of the two-way product rectangle bound. Nevertheless, the rectangle bound $\mathsf{rec}(f, \epsilon, \mu)$ is approximately equal to $\mathsf{sub}^{\square}(f, \epsilon, \mu)$, for any product distribution μ.

Lemma 4.1 *Let μ be a product distribution on $\mathcal{X} \times \mathcal{Y}$ and let $\delta \in (0,1)$. Then*

$$\mathsf{rec}(f, \epsilon, \mu) \quad \geq \quad \mathsf{sub}^{\square}(f, \epsilon, \mu) \quad \geq \quad \mathsf{rec}\Big(f, \frac{\epsilon}{\delta^2}, \mu\Big) - \log \tfrac{1}{(1-\delta)^2} \ .$$

We defer the proof of the lemma to Appendix A.

Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. We define the *k-fold product* of f, $f^{\otimes k} \subseteq \mathcal{X}^k \times \mathcal{Y}^k \times \mathcal{Z}^k$ as $f^{\otimes k} \stackrel{\Delta}{=} \{(x_1, \ldots x_k, y_1, \ldots, y_k, z_1, \ldots, z_k) : \forall i \in [k], (x_i, y_i, z_i) \in f\}$. This relation captures k independent instances of the relation f. We show that the two-way product subdistribution bound satisfies the direct product property by considering f and its k-fold product.

Theorem 4.2 *Let $\epsilon, \delta \in (0, 1/6)$, k be a positive integer, and let $q \stackrel{\Delta}{=} (1-\epsilon/2)^{(1-\delta)k}$. Let $\mu \stackrel{\Delta}{=} \mu_A \otimes \mu_B$ be any product distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathsf{sub}^{[]}(f, \epsilon, \mu) > \frac{48}{\delta\epsilon}$. Then,*

$$\mathsf{sub}^{[]}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) \quad > \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{[]}(f, \epsilon, \mu).$$

Proof: Let $c \stackrel{\Delta}{=} \mathsf{sub}^{[]}(f, \epsilon, \mu)$ and $l \stackrel{\Delta}{=} \frac{\delta\epsilon}{16} \cdot k \cdot c$. Let $\lambda \stackrel{\Delta}{=} \lambda_A \otimes \lambda_B$ be a product distribution on $\mathcal{X}^k \times \mathcal{Y}^k$, such that $\mathsf{S}_\infty(\lambda \parallel \mu^{\otimes k}) \leq l$. Let XY be joint random variables distributed according to λ. For $i \in [k]$, let X_i, Y_i represent the components of X, Y respectively in the ith coordinate. The symbol $\mathbf{1}$ denotes a sequence of appropriate length of ones (that is implied by the context).

We show that for any output string $z = z_1 \ldots z_k \in \mathcal{Z}^k$, the distributional error under λ is greater than $1-2q$. Formally, define boolean random variables S_i such that $S_i = 1$ iff the output in the ith coordinate is correct, i.e., $(X_i, Y_i, z_i) \in f$; $S_i = 0$ otherwise. We show the following.

Lemma 4.3 $\Pr_{XY\sim\lambda}[S_1 \ldots S_k = \mathbf{1}] \leq 2q$.

This lemma directly implies our theorem. ∎

Proof of Lemma 4.3: Let $t \stackrel{\Delta}{=} \lceil (1-\delta)k \rceil$. Our goal is to identify t indices $i_1, \ldots, i_t \in [k]$ such that for each sucessive index i_j in this sequence, the probability, conditioned upon success on the previous $j-1$ coordinates, that the protocol succeeds with output z_{i_j} for the coordinate i_j is bounded by $1 - \frac{\epsilon}{2}$. (This implies our lemma.) We do this by choosing the coordinate i_j such that the marginal distribution of XY in that coordinate "sits well" inside μ, and is a product distribution. We ensure that this property holds even when we condition on success in the previous coordinates. Ensuring a product distribution involves conditioning on the inputs to one party (say, Bob) in the previous coordinates. As a consequence, we only identify the required t coordinates for all but a small fraction of "atypical" values for the conditioned input variables. We elaborate on this below.

For a string $y \in \mathcal{Y}^k$ and $i \in [k]$, let y_i denote the substring in the ith coordinate of y. We extend this notation to a subset of coordinates $I = \{i_1, \ldots, i_j\} \subseteq [k]$ as $y_I = y_{i_1} \ldots y_{i_j}$ (where the coordinates in the subset are always taken in a canonical order). Similarly for $x \in \mathcal{X}^k$.

In the interest of readability, we sometimes use non-standard notation in our arguments below. For a subset $I \subseteq [k]$, we abbreviate $X_I Y_I$ as XY_I. Similarly, we write XY_i for $X_i Y_i$. The subscript (I, w), where $I \subseteq [k]$ and $w \in \mathcal{Y}^{|I|}$, indicates conditioning on the event $Y_I = w$. For example, $XY_{i,(I,w)}$ is the random variable $X_i Y_i$ conditioned upon the event $Y_I = w$.

Let $X'Y'$ be distributed according to $\mu^{\otimes k}$. We identify a set $\mathcal{B}_I \subseteq \mathcal{Y}^{|I|}$ of "atypical" inputs substrings for Bob for each subset I. Let $w \in \mathcal{B}_I \subseteq \mathcal{Y}^{|I|}$, iff

$$\mathsf{S}_\infty(XY_{(I,w)} \parallel X'Y'_{(I,w)}) \quad > \quad l + 2k.$$

In Appendix A we bound the probability that Bob's input has an atypical substring.

Lemma 4.4 $\Pr_{XY\sim\lambda}[\ (\exists I \subset [k])\ Y_I \in \mathcal{B}_I] < 2^{-k}$.

Inputs with substrings in a set $\mathcal{B}_I$ are precisely the ones for which we are not able to carry out the line of argument outlined above.

We also identify a set $\mathcal{L}_I \subseteq \mathcal{Y}^{|I|}$ of "lucky" input substrings for Bob, for each $I \subseteq [k]$ of size less than t. Let $w \in \mathcal{L}_I$ iff $\Pr[S_I = \mathbf{1} | Y_I = w] < 2^{-k}$. Since $2^{-k} \leq q$, for such lucky substrings we already have $\Pr[S_1 \ldots S_k = \mathbf{1} | Y_I = w] < q$.

The following lemma captures the main step in our proof.

Lemma 4.5 *Let $I \subseteq [k]$ be of size less than t,and let $w \in \mathcal{Y}^{|I|}$. Then, either*

1. *The substring $w \in \mathcal{B}_I$, i.e., $\mathsf{S}_\infty(XY_{(I,w)} \,\|\, X'Y'_{(I,w)}) > l + 2k$, or*
2. *The substring $w \in \mathcal{L}_I$, i.e., $\Pr[S_I = \mathbf{1} \mid Y_I = w] < 2^{-k}$, or*
3. *There exists an $i \in [k] - I$, such that $\Pr[S_i = 1 \mid S_I = \mathbf{1}, Y_I = w] < 1 - \frac{\epsilon}{2}$.*

Below we sketch how this implies Lemma 4.3; the technical details are deferred to Appendix A. Lemma 4.5 allows us to select t indices on which the success probability of the protocol is bounded appropriately, so long as parts 1 and 2 are not satisfied. Part 1 is satisfied only for a 2^{-k} fraction of inputs, and we ignore these. As we successively add indices to $I = \{j_1, j_2, \ldots, j_m\}$, if for any value of $m \leq t$, part 2 of the Lemma 4.5 holds, then, in that "branch of conditioning" on the value of Y_I, the probability of success on all k coordinates is bounded by 2^{-k}. If part 2 does not hold for any $m \leq t-1$, then we keep choosing the indices as given by part 3. As long as Bob's input Y does not contain an atypical substring, either part 2 or 3 hold. Therefore we get that the probability of success on all k instances is at most $q + 2^{-k}$. Along with Lemma 4.4 this implies that $\Pr[S_1 \cdots S_k = \mathbf{1}] < 2q$. ■

For the final piece of the argument we prove a key property of sub-distributions.

Lemma 4.6 *Let $0 < \eta < 1/2$ and $\zeta \leq 1$. Let $\mu \stackrel{\Delta}{=} \mu_A \otimes \mu_B$ and $\omega \stackrel{\Delta}{=} \omega_A \otimes \omega_B$ be product distributions on $\mathcal{X} \times \mathcal{Y}$. If $\mathsf{S}(\omega \,\|\, \mu) < \eta \cdot \mathsf{sub}^{[]}(f, \zeta, \mu)$, and $\mathsf{sub}^{[]}(f, \zeta, \mu) > \frac{9}{\eta}$, then any zero-communication protocol for f with output $u \in \mathcal{Z}$ has error at least $\zeta - 4\eta$ under ω, i.e., $\Pr_{XY \sim \omega}[(X, Y, u) \notin f] \geq \zeta - 4\eta$.*

Proof: Suppose $\mathsf{sub}^{[]}(f, \zeta, \mu) = d$, $\mathsf{S}(\omega_A \,\|\, \mu_A) = s_A$ and $\mathsf{S}(\omega_B \,\|\, \mu_B) = s_B$. Note that $\mathsf{S}(\omega \,\|\, \mu) = s_A + s_B < \eta d$. Let $r \stackrel{\Delta}{=} 1/2\eta$. Applying Lemma 2.7 to ω_A and ω_B separately, we get a distribution $\omega' = \omega'_A \otimes \omega'_B$ with $\|\omega - \omega'\|_1 \leq 4/r$ and $\mathsf{S}_\infty(\omega' \,\|\, \mu) \leq r(s_A + s_B + 2) + 2\log\frac{r}{r-1} < d$. This implies, from definition of $\mathsf{sub}^{[]}(f, \zeta, \mu) = d$, that any zero-communication protocol with output $u \in \mathcal{Z}$ has error $> \zeta$ under ω'. Since $\|\omega - \omega'\|_1 \leq 4/r = 8\eta$, Lemma 2.8 tells us that the protocol has error at least $\zeta - 4\eta$ under ω. ■

Proof of Lemma 4.5: We follow the previously described non-standard notation for conditional random variables. In addition, a superscript '$\mathbf{1}$' indicates conditioning on the event $S_I = \mathbf{1}$, with I and S_I as in the statement of the lemma.

To prove the lemma, we show that when parts 1 and 2 are false, part 3 holds. By hypothesis, we have

$$\begin{array}{rrcll}
 & \mathsf{S}_\infty(XY_{(I,w)} \,\|\, X'Y'_{(I,w)}) & \leq & l + 2k & \\
\Rightarrow & \mathsf{S}_\infty(XY^{\mathbf{1}}_{(I,w)} \,\|\, X'Y'_{(I,w)}) & \leq & l + 3k, & \text{since } \Pr[S_{I,(I,w)} = \mathbf{1}] \geq 2^{-k}; \\
\Rightarrow & \mathsf{S}_\infty(XY^{\mathbf{1}}_{(I,w),[k]-I} \,\|\, X'Y'_{[k]-I}) & \leq & l + 3k, & \text{from Lemma 2.9}; \\
\Rightarrow & \mathsf{S}(XY^{\mathbf{1}}_{(I,w),[k]-I} \,\|\, X'Y'_{[k]-I}) & \leq & l + 3k, & \text{from Lemma 2.6}; \\
\Rightarrow & \sum_{i \in [k]-I} \mathsf{S}(XY^{\mathbf{1}}_{i,(I,w)} \,\|\, X'Y'_i) & \leq & l + 3k, & \text{from Lemma 2.5}; \\
\Rightarrow & \exists (i \in [k] - I) \quad \mathsf{S}(XY^{\mathbf{1}}_{i,(I,w)} \,\|\, X'Y'_i) & \leq & \frac{l+3k}{k-(1-\delta)k} < \frac{\epsilon c}{8} &
\end{array} \tag{3}$$

In the third inequality, we also used the independence of $X'Y'_I$ and $X'Y'_{[k]-I}$. The last inequality follows from $l = \frac{\delta\epsilon}{16}kc$ and the assumption that $\mathsf{sub}^{[]}(f, \epsilon, \mu) = c > \frac{48}{\delta\epsilon}$.

We show in Appendix A that:

Lemma 4.7 *The distribution of the random variables $XY^{\mathbf{1}}_{i,(I,w)}$ is product on $\mathcal{X} \times \mathcal{Y}$.*

Lemma 4.6 tells us that the error in the ith coordinate is therefore at least $\epsilon - \frac{\epsilon}{2} \geq \frac{\epsilon}{2}$. This implies part 3 of the lemma. ■

The direct product property of the subdistribution bound tranlates to a similar result for the communication complexity of two-way protocols. Its proof appears in Appendix A.

Theorem 4.8 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\epsilon, \delta \in (0, 1/6)$ and k be a positive integer. Let $q \triangleq (1-\epsilon/2)^{(1-\delta)k}$. Suppose $\mathsf{sub}^{\square}(f, \epsilon) > \frac{48}{\delta\epsilon}$. Then,*

$$\mathsf{R}^{\mathsf{pub}}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{\square}_{1-3q}(f^{\otimes k}) \quad > \quad k \cdot \left[\frac{\delta\epsilon}{16} \cdot \mathsf{sub}^{\square}(f, \epsilon) - 1 \right].$$

The two-way product rectangle bound (for constant error) for Set Disjointness, and therefore the product subdistribution bound, is $\Omega(\sqrt{n})$ [BFS86]. As a consequence, there is a constant κ such that any two-way protocol for its k-fold product with communication at most $\kappa k\sqrt{n}$ has success probability at most $2^{-\Omega(k)}$.

4.2 One-way protocols

We now explain how the same ideas as in the previous section lead to a direct product result for one-way communication. The primary difference in this case is that the output of the protocol cannot in general be inferred from the single message sent by Alice. To handle this, we define a variant of the product subdistribution bound which is symmetric with respect to Alice and Bob.

Definition 4.2 (One-way symmetric product subdistribution bound) *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\mu \triangleq \mu_A \otimes \mu_B$ be a product distribution on $\mathcal{X} \times \mathcal{Y}$. Let $\mathsf{sub}^{1,\square}(f, \epsilon, \mu) \triangleq \min_\lambda \mathsf{S}_\infty(\lambda \,\|\, \mu)$, where λ ranges over all (product) distributions that are* one-message-like *for μ and ϵ-monochromatic for f. We define the* one-way symmetric product subdistribution bound *as $\mathsf{sub}^{1,\square}(f, \epsilon) \triangleq \max_\mu \mathsf{sub}^{1,\square}(f, \epsilon, \mu)$, where μ ranges over all* product distributions *on $\mathcal{X} \times \mathcal{Y}$.*

Note that a distribution that is one-message-like for a product distribution is itself a product distribution.

The following relationships between the one-way symmetric product subdistribution bound and the one-way product subdistribution bound are straightforward and we state them without proof.

Lemma 4.9 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $\mu \triangleq \mu_A \otimes \mu_B$ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Then*

1. $\mathsf{sub}^{1,\square}(f, \epsilon, \mu) \geq \mathsf{sub}^{1,\square}_{\mathsf{B}}(f, \epsilon, \mu), \quad \mathsf{sub}^{1,\square}(f, \epsilon) \geq \mathsf{sub}^{1,\square}_{\mathsf{B}}(f, \epsilon).$
2. $\mathsf{sub}^{1,\square}_{\mathsf{B}}(f, \epsilon, \mu) + \log|\mathcal{Z}| \geq \mathsf{sub}^{1,\square}(f, \epsilon, \mu), \quad \mathsf{sub}^{1,\square}_{\mathsf{B}}(f, \epsilon) + \log|\mathcal{Z}| \geq \mathsf{sub}^{1,\square}(f, \epsilon).$

We arrive at the following direct product result for one-way symmetric product subdistribution bound along the lines of Theorem 4.2.

Theorem 4.10 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \triangleq (1-\epsilon/2)^{(1-\delta)k}$. Let $\mu \triangleq \mu_A \otimes \mu_B$ be any product distribution on $\mathcal{X} \times \mathcal{Y}$ such that $\mathsf{sub}^{1,\square}(f, \epsilon, \mu) > \frac{48}{\delta\epsilon}$. Then*

$$\mathsf{sub}^{1,\square}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) \quad > \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{1,\square}(f, \epsilon, \mu).$$

This implies the following direct product result for one-way communication.

Corollary 4.11 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \triangleq (1-\epsilon/2)^{(1-\delta)k}$. Suppose $\mathsf{sub}^{\square}(f, \epsilon) > \frac{48}{\delta\epsilon}$. Then,*

$$\mathsf{R}^{1,\mathsf{pub}}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{1,\square}_{1-3q}(f^{\otimes k}) \quad > \quad k \cdot \left[\frac{\delta\epsilon}{16} \cdot \mathsf{sub}^{1,\square}_{\mathsf{B}}(f, \epsilon) - \log|\mathcal{Z}| - 1 \right].$$

This combined with Theorem 3.6 subsumes the strong direct product result due to de Wolf [dW06] for the one-way randomized communication complexity of Index. Similar results for other functions like Set Disjointness and Inner product, whose one-way communication complexity is captured by their VC-dimension, then follow.

In fact for a complete function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, we can avoid the loss of the $\log |\mathcal{Z}|$ term.

Theorem 4.12 *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a complete function. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \stackrel{\Delta}{=} (1 - \epsilon/2)^{(1-\delta)k}$. There are universal constants $\gamma_0, \gamma_1 > 0$ such that if $\mathsf{sub}^{\emptyset}(f, \epsilon) > \frac{\gamma_0}{\delta\epsilon}$. Then,*

$$\mathsf{R}^{1,pub}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{1,\emptyset}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{sub}^{1,\emptyset}_{\mathsf{B}}(f^{\otimes k}, 1-2q) - k \quad > \quad k \cdot \left[\frac{\delta\epsilon}{\gamma_1} \cdot \mathsf{sub}^{1,\emptyset}_{\mathsf{B}}(f, \epsilon) - 1 \right].$$

The proof will be included in the full version of this article.

4.3 SMP protocols

Subdistribution bounds are also relevant for the SMP model, as sketched in this section.

Definition 4.3 *For distributions θ, μ with support in $\mathcal{X} \times \mathcal{Y}$, we say θ is one-message-like for μ* with respect to Alice *if for all $x \in \mathcal{X}$ if $\theta(x) > 0$, then $\mu(x) > 0$, and $(\forall y)\mu(y|x) = \theta(y|x)$. Similarly we say θ is one-message-like for μ* with respect to Bob *if for all $y \in \mathcal{Y}$, if $\theta(y) > 0$, then $\mu(y) > 0$, and $(\forall x)\mu(x|y) = \theta(x|y)$.*

Definition 4.4 (SM-like) *We call a distribution λ with support in $\mathcal{X} \times \mathcal{Y}$ SM-like for μ, if there exists a distribution θ such that θ is one-message-like for μ with respect to* Alice *and λ is one-message-like for θ with respect to* Bob.

We define the *SM-subdistribution bound* as follows.

Definition 4.5 (SM-subdistribution bound) *Let $\mathsf{sub}^{\|}(f, \epsilon, \mu) \stackrel{\Delta}{=} \min_\lambda \mathsf{S}(\lambda \,\|\, \mu)$, where λ ranges over all distributions which are SM-like for μ and ϵ-monochromatic for f. Define $\mathsf{sub}^{\|}(f, \epsilon) \stackrel{\Delta}{=} \max_\mu \mathsf{sub}^{\|}(f, \epsilon, \mu)$, where μ ranges over all distributions on $\mathcal{X} \times \mathcal{Y}$. When the maximization is over only product distributions we get the SM product subdistribution bound $\mathsf{sub}^{\|,\emptyset}(f, \epsilon)$.*

We now state a direct product theorem for SMP protocols in terms of this SM-product subdistribution bound.

Theorem 4.13 *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Let $0 < \epsilon, \delta < 1/6$ and k be a positive integer. Let $q \stackrel{\Delta}{=} (1 - \epsilon/2)^{(1-\delta)k}$. There are universal constants $\gamma_0, \gamma_1 > 0$ such that if $\mathsf{sub}^{\|,\emptyset}(f, \epsilon) > \frac{\gamma_0}{\delta\epsilon}$. Then,*

$$\begin{aligned} \mathsf{R}^{\|,\mathsf{pub}}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{R}^{\|,\emptyset}_{1-3q}(f^{\otimes k}) \quad &\geq \quad \mathsf{sub}^{\|,\emptyset}(f^{\otimes k}, 1-2q) - k \\ &> \quad k \cdot \left[\frac{\delta\epsilon}{\gamma_1} \cdot \mathsf{sub}^{\|,\emptyset}(f, \epsilon) - 1 \right]. \end{aligned}$$

The proof follows along the lines of Theorem 4.2 and Theorem 4.8 and is omitted.

5 Applications

5.1 Entanglement versus communication

Some of the most important questions in quantum communication concern the power of entanglement. Here we consider quantum communication complexity, as introduced by Yao [Yao93], and investigated extensively thereafter. For definitions concerning quantum computing we refer the reader to Nielsen and Chuang's monograph [NC00].

There are several models of quantum communication complexity: with entanglement and quantum communication, with entanglement and classical communication, and without entanglement but with quantum communication. Due to the phenomenon of quantum teleportation [BBC$^+$93], any protocol with shared entanglement and c qubits of *quantum* communication may be converted to a protocol with an additional c shared EPR-pairs, and a total of $2c$ classical communication.

We are interested in the question whether the quantum communication complexity can be reduced drastically by allowing prior entanglement. So far only small savings in the quantum communication complexity are known when entanglement is allowed. Basically, superdense coding [BW92] allows us to compress classical messages by a factor of 2 when entanglement is available, hence saving a factor of 2 in the quantum communication complexity for the model with entanglement. Also, entanglement can be used like public randomness, leading to additive $\Theta(\log n)$ savings for some functions, e.g., $\mathsf{Equality}$. This gives rise to the question as to how much entanglement is actually necessary to compute a function optimally.

In the analogous situation for public randomness, Newman [New91] shows that $O(\log n)$ public random bits are enough to compute any function with optimal communication complexity. His proof is a black box simulation, in the sense that is does not change the protocol, but rather chooses uniformly at random from a polynomial-size set of strings and runs the protocol with this randomness. Can the amount of entanglement be reduced in the same way for quantum protocols? Jain *et al.* [JRS05b] showed that in fact such a black box approach does not work. Recently, Gavinsky [Gav06] showed a stronger statement. He showed that there is a relation that can be computed with $\mathrm{O}(k \log n)$ communication and entanglement in a simultaneous message passing protocol, while every one-way protocol with $\mathrm{o}(k/\log n)$ entanglement and only classical messages needs communication $\Omega(k\sqrt{n}/\log n)$. Hence trying to work with less entanglement increases the communication complexity, or requires drastic changes to the protocol, e.g., going from classical to quantum messages.

Gavinsky derives his result using a direct product theorem for the one-way communication complexity of a certain class of relations. Here we follow the same approach, but use our direct product theorem for one-way communication complexity to get stronger trade-offs.

We begin by defining the relation used in the result. Recall that a perfect matching is an undirected graph in which there is one and only one edge incident on each vertex.

Definition 5.1 (Boolean Hidden Matching) *In the* boolean hidden matching *problem* BHM_n*,* Alice *gets a string* $x \in \{0,1\}^{2n}$*, and* Bob *gets a perfect matching* M *on* $2n$ *vertices.* Bob *is required to output an edge* $\{j,k\}$ *in* M *along with the bit* $x_j \oplus x_k$*.*

Bar-Yossef, Jayaram, and Kerenidis [BYJK04] show that there is a large gap between the classical and quantum one-way complexity of the relation BHM_n.

Theorem 5.1 ([BYJK04]) *The one-way quantum communication complexity (with no error and with no prior shared entanglement) of the boolean hidden matching relation* BHM_n *is* $\mathrm{O}(\log n)$*. Moreover,* $\mathsf{R}^{1,[]}_{1/3}(\mathsf{BHM}_n) = \Omega(\sqrt{n})$.

As mentioned above, with quantum teleportation we can implement the quantum protocol for BHM_n as a one-way protocol with $O(\log n)$ shared EPR-pairs and $O(\log n)$ classical communication.

Like Gavinsky, we show that a certain amount of entanglement is necessary to preserve the optimal communication complexity of the k-fold product of Boolean Hidden Matching.

Theorem 5.2 *The relation $\mathsf{BHM}_n^{\otimes k}$, with input length $\Theta(kn \log n)$, can be computed exactly (with no error) by a one-way quantum protocol with prior entanglement in the form of $O(k \log n)$ shared EPR-pairs, and (classical) communication $O(k \log n)$. There is a constant $\gamma > 0$ such that any one-way quantum protocol which has an entangled state on γk qubits needs classical communication $\Omega(k\sqrt{n})$.*

Proof: By Theorem 5.1 and the remark following it, $\mathsf{BHM}_n^{\otimes k}$ can be computed by a one-way protocol with no error with $O(k \log n)$ EPR-pairs and using $O(k \log n)$ bits of classical communication.

From our direct product theorem for one-way classical protocols, Theorem 4.11, there is a constant $d > 0$ such that

$$\mathsf{R}^{1,\mathsf{pub}}_{1-2^{-dk}}(\mathsf{BHM}_n^{\otimes k}) \quad \geq \quad \mathsf{R}^{1,[]}_{1-2^{-dk}}(\mathsf{BHM}_n^{\otimes k}) \quad > \quad \Omega(k(\sqrt{n} - 2\log n - 2)) \quad = \quad \Omega(k\sqrt{n}). \tag{4}$$

Suppose we are given a one-way protocol for $\mathsf{BHM}_n^{\otimes k}$ with entanglement ρ over $dk/2$ qubits, classical communication c, and error at most $1/3$. The initial state of the protocol is the entangled state given to Alice and Bob, in tensor product with their inputs. The entire computation of the protocol (unitary operations and measurements) followed by the acceptance criterion is captured by a POVM element E that depends upon the input alone, and acts on the entangled state. The probability of acceptance is then $\mathrm{Tr}(E\rho)$. We replace the entangled state by the *maximally mixed state* over $dk/2$ qubits. This decreases the success probability of the protocol to no worse than $(2/3) \cdot 2^{-dk/2} > 2^{-dk}$. This holds because any quantum state ρ on l qubits (formally a positive semidefinite $2^l \times 2^l$ matrix with trace 1) "sits inside" the maximally mixed state U_l with probability at least 2^{-l}, i.e., $\mathsf{U}_l - 2^{-l}\rho \geq 0$.

An l-qubit maximally mixed state is physically and computationally equivalent to the uniform distribution on l-bit strings. This is a product distribution. As a result, we are left with a private-coin randomized protocol for $\mathsf{BHM}_n^{\otimes k}$ with classical communication c and success probability $> 2^{-dk}$. From Eq. (4) we conclude that $c = \Omega(k\sqrt{n})$. ∎

If we choose $k = n^p$ for some constant $p > 0$, we get a polynomial gap between the two bounds in the theorem, when the entanglement used is reduced only slightly (from $\Theta(n^p \log n)$ to $O(n^p)$).

5.2 Multiparty communication

Beame *et al.* [BPSW07] apply their strong direct product theorem for the corruption/rectangle bound under product distributions to get a lower bound in the *number on the forehead* model of multiparty communication complexity. In the number on the forehead (NOF) model with three players, called Alice, Bob, and Charlie, each player receives exactly two of inputs from amongst the three $x, y, z \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. They wish to compute an output $w \in \mathcal{W}$ such that (x, y, z, w) satisfy a relation f. Player Alice receives inputs x, z, Bob y, z, and Charlie x, y. In effect, each player has an input written on her/his forehead, which is not visible to her/him. However, the player can see the inputs written on the other players' foreheads. The three parties may have access to different patterns of communication channel.

Consider the case when $x, y, z \in \{0,1\}^n$ and the players wish to compute three-player Set Disjointness, i.e., they would like to accept iff $\bigvee_{i=1}^n (x_i \wedge y_i \wedge z_i) = 0$. Beame *et al.* show that this task needs communication $\Omega(n^{1/3})$ for randomized protocols with error at most $1/3$, when the communication is restricted as follows: Charlie sends one message to Alice or Bob, and the remaining communication is only between Alice

and Bob. No superlogarithmic bounds are known in the case in which all pairs of players can communicate with each other throughout the protocol. Such lower bounds would have strong consequences, as described in [BPSW07]. In the one-way version of this model there are two messages, the first from Charlie to Alice, and the second from Alice to Bob, who then produces the output. A lower bound of $\Omega(\sqrt{n})$ in this model for Set Disjointness follows from a result due to Wigderson [BHK01, Section 9.3]. De Wolf [dW06] shows how to infer this lower bound from a strong direct product theorem for the one-way communication complexity of Index.

Our direct product results give us such lower bounds for general relations. Below, we describe the particular case boolean functions for the one-way NOF model (although still not in the full generality in which this kind of argument applies).

Let g_n be a family of functions on $\{0,1\}^n$. We consider the three-party NOF communication complexity of $g_n(x \wedge y \wedge z)$, where $u \wedge v$ is the n-bit string whose ith coordinate is $u_i \wedge v_i$. We say that g_n is *l-self-reducible*, if its n-bit input can be partitioned into $l+1$ blocks $w_1, \ldots, w_{l+1}$, each of size $\lfloor n/(l+1) \rfloor$, so that for each $i \in [l]$, $g_{\lfloor n/(l+1) \rfloor}(w_i)$ equals the value of the function g_n when evaluated on an input $\tilde{w}_i$ described next. The input $\tilde{w}_i$ consists of 0s in all blocks except the ith and the $(l+1)$th. The ith block equals w_i, and the $(l+1)$th block is some fixed string, possibly with 1s and may depend upon the block number i.

Clearly, $g_n = \mathsf{OR}_n$, the logical OR of n bits, gives us three-player Set Disjointness, and is l-self-reducible for every l. Other well-studied functions like an AND of $\sqrt{n}$ ORs of $\sqrt{n}$ variables each (called Tribes_n) also fall in this class.

Theorem 5.3 *Let $f_n(x,y,z) = g_n(x \wedge y \wedge z)$, and $h_n(x,y) = g_n(x \wedge y)$, where g_n is l-self-reducible. Every bounded error randomized one-way three-party NOF protocol (in which* Charlie *sends a message to* Alice*, who sends a message to* Bob*) for f_n needs communication $\Omega(\min\{l, \mathsf{VC}(h_{\lfloor n/(l+1) \rfloor})\})$.*

Proof: The idea behind the proof is the same as in the work of De Wolf [dW06] and Beame *et al.* [BPSW07], and we only sketch it here.

Given a bounded error one-way three-party randomized NOF protocol Π for f_n, we derive a one-way protocol Π' for computing the l-fold product of $h_{\lfloor n/(l+1) \rfloor}$. The protocol Π' is such that for every sequence of l inputs, it correctly computes at least $(1-\nu)l$ instances with probability $2^{\Omega(-l)}$, where ν is a small constant. Its communication complexity is l times that of Π.

Suppose Alice and Bob get l instances $\{(u_i, v_i)\}$ of $h_{\lfloor n/(l+1) \rfloor}$. To compute the function value for all of these simultaneously, they concatenate their inputs into n-bit strings x, y the first l blocks of which are given by the l respective inputs, and the remaining block in each is set to the all 1s string.

If in protocol Π, Charlie sends a message of length $\geq \alpha l$ (for some small constant α) to Alice, the theorem holds. Otherwise, this communication is at most l. Charlie's message only depends upon x, y, but not on z. To solve the ith instance in the protocol Π', Alice and Bob construct $z = z^i$ as follows. They set the ith block of $z^i \in \{0,1\}^n$ to the all 1s string, the $(l+1)$th block according to the l-self-reduction, and the remaining blocks to 0. Using the inputs x, y, z^i to Π, they compute an output for the ith instance u_i, v_i. The net communication from Alice to Bob is at most l times the communication c in Π.

By the l-self-reducibility property of g_n, we have $h_{\lfloor n/(l+1) \rfloor}(u_i, v_i) = g_n(x \wedge y \wedge z^i)$. Since for every input the protocol Π makes an error with probability at most a small constant ν, the expected number of instances out of the given l for which Π' *incorrectly* computes the answers is $\leq \nu l$. By the Markov inequality, the probability that more than $2\nu l$ instances are incorrectly computed is at most $1/2$. Replacing Charlie's message by a uniformly random l-bit string, we get that with probability at least $2^{-(\alpha l+1)}$, Alice and Bob solve at least $(1-2\nu)l$ instances correctly.

Corollary 4.11 implies that even for the relaxed notion of correctness described above, the probability of success is bounded by $2^{-\Omega(l)}$. (The corallary implies that for any *fixed* subset of size $(1-2\nu)l$, the probability of being simultaneously correct on all indices in the subset is at most $2^{-\Omega(l)}$. A union bound over all subsets of size at least $(1-2\nu)l$ now gives us the required variant of the theorem, when ν is small enough.)

Combining all these pieces of argument, provided α is sufficiently small, we get that $lc \geq \Omega(l \cdot \mathsf{R}^{1,[]}_{1/3}(h_{\lfloor n/(l+1)\rfloor})) = l \cdot \Omega(\mathsf{VC}(h_{\lfloor n/(l+1)\rfloor}))$. ∎

The function $g_n = \mathsf{OR}_n$ is $\sqrt{n}$-self-reducible, and the corresponding bivariate function h_n has VC-dimension n. Hence the one-way randomized NOF complexity of three-party Set Disjointness is $\Omega(\sqrt{n})$. Similarly, Tribes_n is $\sqrt{n}$-self-reducible, and the corresponding VC-dimension is $n - \sqrt{n}$. Therefore, the one-way randomized NOF complexity of Tribes_n is also $\Omega(\sqrt{n})$.

Similar results can be shown in the model where Charlie sends a message to Alice, and then Alice and Bob engage in a two-party protocol.

Theorem 5.4 *Let $f_n(x,y,z) = g_n(x \wedge y \wedge z)$, and $h_n(x,y) = g_n(x \wedge y)$, where g_n is l-self-reducible. Every bounded error randomized three-party NOF protocol as described above for f_n needs communication $\Omega(\min\{l, \mathsf{sub}^{[]}(h_{\lfloor n/(l+1)\rfloor}, 1/3)\})$.*

In particular, we get the $\Omega(n^{1/3}$ lower bound for Set Disjointness.

Acknowledgements

We thank Ronald de Wolf for sharing with us his write-up [dW06] on the strong direct product theorem for Index and for many useful discussions.

References

[Aar04] Scott Aaronson. Limitations of quantum advice and one-way communication. In *Proceedings of the 19th Annual IEEE Conference on Computational Complexity*, pages 320–332, 2004.

[ANTSV99] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, pages 376–383. ACM Press, 1999.

[BBC+93] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899, 1993.

[BFS86] László Babai, Peter Frankl, and Janos Simon. Complexity classes in communication complexity theory. In *Proceedings of the 27th Annual IEEE Symposium on Foundations of Computer Science*, pages 337–347, 1986.

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

[BPSW07] Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A direct sum theorem for corruption and a lower bound for the multiparty communication complexity of Set Disjointness. *Computational Complexity*, 2007. To appear.

[BW92] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884, 1992.

[BYJK04] Ziv Bar-Yossef, T. S. Jayram, and Iordanis Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 128–137, 2004.

[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. Special issue on FOCS 2002.

[CSUU07] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, 2007. To appear.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew C.-C. 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. John Wiley & Sons, New York, NY, USA, 1991.

[dW06] Ronald de Wolf. Random access codes, direct product theorems, and multiparty communication complexity. Private communication, 2006.

[Gav06] Dmitry Gavinsky. On the role of shared entanglement. Technical Report quant-ph/0604052, ArXiv.org Preprint Archive, `http://www.arxiv.org/abs/quant-ph/`, April 2006.

[GNW95] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR-lemma. Technical Report TR95-050, Electronic Colloquium on Computational Complexity, `http://http://eccc.hpi-web.de/eccc/`, 1995. Revision 1, January 1999.

[HJMR07] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. In *Proceedings of the 22nd Annual IEEE Conference on Computational Complexity*, 2007. To appear.

[IRW94] Russell Impagliazzo, Ran Raz, and Avi Wigderson. A direct product theorem. In *Proceedings of the Ninth Annual IEEE Structure in Complexity Theory Conference*, pages 88–96, 1994.

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

[JRS03a] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In *Proceedings of the Thirtieth International Colloquium on Automata Languages and Programming*, volume 2719 of *Lecture notes in Computer Science*, pages 300–315. Springer, Berlin/Heidelberg, 2003.

[JRS03b] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of Set Disjointness. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 220–229, 2003.

[JRS05a] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. On divergence, relative entropy and the substate property. Technical Report quant-ph/0506210, ArXiv.org Preprint Archive, `http://www.arxiv.org/abs/quant-ph/`, June 2005.

[JRS05b] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 285–296, 2005.

[Kla04] Hartmut Klauck. Quantum and classical communication-space tradeoffs from rectangle bounds. In *Proceedings of the 24th Annual IARCS International Conference on Foundations*

of Software Technology and Theoretical Computer Science, volume 3328 of *Lecture notes in Computer Science*, pages 384–395. Springer, Berlin/Heidelberg, 2004.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.

[KNR99] Ilan Kremer, Noam Nisan, and Dana Ron. On randomized one-round communication complexity. *Computational Complexity*, 8(1):21–49, 1999. Corrected version available at `http://www.eng.tau.ac.il/ danar/Public-pdf/KNR-fix.pdf`.

[KvdW04] Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 12–21, 2004.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.

[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 Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 363–372, 1997.

[PT06] Mihai Pătraşcu and Mikkel Thorup. Higher lower bounds for near-neighbor and further rich problems. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 646–654. IEEE Computer Society Press, Los Alamitos, CA, USA, 2006.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

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

[Yao93] Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, 1993.

A Proofs of some lemmas and theorems

Proof of Lemma 4.1: Let $\mu = \mu_A \otimes \mu_B$ be a product distribution on $\mathcal{X} \times \mathcal{Y}$. By definition, $\mathsf{rec}(f, \epsilon, \mu) \geq \mathsf{sub}^{\square}(f, \epsilon, \mu)$. For the second inequality we argue as follows. Consider any ϵ-monochromatic product distribution $\lambda = \lambda_A \otimes \lambda_B$ along with an output $z \in \mathcal{Z}$ which makes it ϵ-monochromatic. View λ as a convex combination of distributions λ_x on $\{x\} \times \mathcal{Y}$. Note that the marginal distribution of λ_x on $\mathcal{Y}$ is λ_B for all x. Using the Markov Inequality, we get a subset $S \subseteq \mathcal{X}$ such that $\lambda_A(S) \geq 1 - \delta$ and for each $x \in S$, the distribution λ_x is (ϵ/δ)-monochromatic for the same output z. Therefore, the distribution $\pi = \mu_{A,S} \otimes \lambda_B$, where $\mu_{A,S}$ is the distribution on $\mathcal{X}$ conditioned upon event S, is also (ϵ/δ)-monochromatic for f with output z. Similarly we identify a subset $T \subseteq \mathcal{Y}$ such that $\lambda_B(T) \geq 1 - \delta$, and each distribution π_y on $\mathcal{X} \times \{y\}$ with marginal $\mu_{A,S}$ on $\mathcal{X}$ is (ϵ/δ^2)-monochromatic for every $y \in T$.

Thus, we get a rectangle $R = S \times T$ with probability $\lambda(R) \geq (1 - \delta)^2$ such that the distribution μ_R, the distribution μ conditioned on R, is (ϵ/δ^2)-monochromatic. Moreover, if $\mathsf{S}_\infty(\lambda \| \mu) = c$, then $\mu(R) \geq \lambda(R) \cdot 2^{-c} \geq (1 - \delta)^2 \cdot 2^{-c}$. In other words, $\mathsf{rec}(f, \epsilon/\delta^2, \mu) \leq c + \log \frac{1}{(1-\delta)^2}$. Minimizing over all such λ, we see that $\mathsf{rec}(f, \epsilon/\delta^2, \mu) \leq \mathsf{sub}^{\square}(f, \epsilon, \mu) + \log \frac{1}{(1-\delta)^2}$. ∎

Proof of Lemma 4.4: Let $w \in \mathcal{B}_I$ for some $I \subseteq [k]$. Since

$$\mathsf{S}_\infty(XY_{(I,w)} \parallel X'Y'_{(I,w)}) \quad > \quad l + 2k,$$

there exist $x, y \in \mathcal{X}^k \times \mathcal{Y}^k$ with $y_I = w$ such that

$$\frac{1}{2^{l+2k}} \cdot \Pr[X = x, Y = y \mid Y_I = w] \quad > \quad \Pr[X' = x, Y' = y \mid Y'_I = w].$$

Since $\mathsf{S}_\infty(\lambda \parallel \mu^{\otimes k}) \leq l$ we have, $\Pr[X = x, Y = y] \leq 2^l \cdot \Pr[X' = x, Y' = y]$. Combining these, we get

$$\Pr[Y_I = w] \quad < \quad 2^{-2k} \cdot \Pr[Y'_I = w].$$

Summing up over all possibilities for w, we get

$$\Pr[Y_I \in \mathcal{B}_I] \quad < \quad 2^{-2k}.$$

Therefore, by the union bound over subsets I,

$$\Pr[\,(\exists I \subseteq [k])\; Y_I \in \mathcal{B}_I] \quad < \quad \sum_{I \subseteq [k]} 2^{-2k} \quad < \quad 2^{-k}.$$

■

Proof of Lemma 4.3: Here we rigorously complete the proof of this lemma following the informal sketch in Section 4.1.

In order to bound $\Pr[S_1 \dots S_k = \mathbf{1}]$, we recursively define a subset $J = \{j_1, \dots, j_t\} \subseteq [k]$ of size t for every $y \in \mathcal{Y}^k$. The set J depends upon Bob's input y, and therefore is a random variable correlated with XY. For the purposes of analysis, we also introduce boolean random variables A_m, L_m, for $m \in [t]$.

Since $\mathsf{S}_\infty(\lambda \| \mu) \leq l$, parts 1 and 2 of Lemma 4.5 are false (with the $I = \emptyset$ and w set to the null string). Let j_1 be the smallest index given by part 3 of the lemma. We set $J = \{j_1\}$, $A_1 = 0 = L_1$.

Suppose indices $I = \{j_1, \dots, j_m\}$ have been defined for input y for some $m \in [t]$. If $y_I \in \mathcal{B}_I \cup \mathcal{L}_I$, i.e., part 1 or 2 of Lemma 4.5 is satisfied with $w = y_I$, then we extend I arbitrarily to a subset J of size t containing I. If part 1 is satisfied we define $A_p = 1, L_p = 0$ for all $p > m$. If part 2 is, then we set $L_p = 1, A_p = 0$ for all $p > m$. Otherwise, we let j_{m+1} be the smallest index i given by part 3 of Lemma 4.5 for I as above and $w = y_I$, and set $A_{m+1} = 0 = L_{m+1}$. Thus, the random variables A_p, L_p are monotonically non-decreasing functions that indicate if parts 1 or 2 were satisfied at any point in the recursive definition of J. In particular, they indicate if the input y is atypical or is lucky.

Lemma 4.4 tells us that $\Pr[A_t = 1] \leq \Pr[\,(\exists I \subseteq [k])\; Y_I \in \mathcal{B}_I] < 2^{-k}$. Since

$$\begin{aligned}
&\Pr[S_1 \dots S_k = \mathbf{1}] \\
&\quad = \quad \Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 1] + \Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0] \\
&\quad < \quad 2^{-k} + \Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0],
\end{aligned}$$

if we show that

$$\Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0] \quad < \quad q + 2^{-k}, \tag{5}$$

we would get a bound of $q + 2^{-k+1} \leq 2q$ as required to prove our lemma.

Now,

$$\begin{aligned}
&\Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0] \\
&\quad = \quad \Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0,\; L_t = 1] + \Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0,\; L_t = 0] \\
&\quad < \quad 2^{-k} + \Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0,\; L_t = 0],
\end{aligned} \tag{6}$$

since $L_t = 1$ implies that there is a subset J as defined above such that

$$\Pr[S_1 \dots S_k = \mathbf{1},\; A_t = 0,\; L_t = 1] \quad \leq \quad \mathbb{E}_{Y_J} \Pr[S_J = \mathbf{1} | Y_J] \quad < \quad 2^{-k}.$$

We bound the second term in Eq. (6) by an inductive argument. We show that for all $m \in [t]$,

$$\Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_m = 0,\ L_m = 0] \quad < \quad (1-\epsilon/2)^m. \tag{7}$$

This is true for $m = 1$ by virtue of Lemma 4.5. Assume that Eq. 7 holds for some $m \geq 1$. Then,

$$\begin{aligned}
&\Pr[S_{j_1} \dots S_{j_{m+1}} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0] \\
&= \sum_{w \in \mathcal{Y}^m} \Pr[S_{j_{m+1}} = 1 \mid S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0,\ Y_{j_1} \cdots Y_{j_m} = w] \\
&\quad \times \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0,\ Y_{j_1} \cdots Y_{j_m} = w] \\
&< (1-\epsilon/2) \cdot \sum_{w \in \mathcal{Y}^m} \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0,\ Y_{j_1} \cdots Y_{j_m} = w] \\
&= (1-\epsilon/2) \cdot \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_{m+1} = 0,\ L_{m+1} = 0] \\
&\leq (1-\epsilon/2) \cdot \Pr[S_{j_1} \dots S_{j_m} = \mathbf{1},\ A_m = 0,\ L_m = 0] \\
&< (1-\epsilon/2)^{m+1}.
\end{aligned}$$

Here, we invoked part 3 of Lemma 4.5 in the first inequality, the monotone non-decreasing property of A_p, L_p in the penultimate step, and the induction hypothesis in the final step. This proves that the second term in Eq. (6) is bounded by q, and therefore Eq. (5) holds. ■

Proof of Lemma 4.7: Recall that $XY \sim \lambda = \lambda_A \otimes \lambda_B$, and therefore are in a product distribution. Therefore, $XY_{(I,w)} = XY|(Y_I = w) = X \otimes (Y|(Y_I = w))$ are in a product distribution. Also $S_I|(Y_I = w) = \mathbf{1}$ is the event $(X_I, w, z_I) \in f^{\otimes |I|}$. So $XY^{\mathbf{1}}_{(I,w)} = XY|(Y_I = w, S_I = \mathbf{1})$ are also in a product distribution. Consequently, the marginal of these random variables on the ith coordinate is also in a product distribution. ■

Proof of Theorem 4.8: The first inequality follows from the definitions. For the second inequality consider a product distribution μ such that $\mathsf{sub}^{[]}(f,\epsilon) = \mathsf{sub}^{[]}(f,\epsilon,\mu)$. Arguing as in the proof of Lemma 3.1, and noting that the the conditional distribution of the inputs given any message is still a product distribution, we get

$$\mathsf{D}^{\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \quad > \quad \mathsf{sub}^{[]}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k.$$

Since $q \geq 2^{-k}$, we get:

$$\begin{aligned}
\mathsf{R}^{[]}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{D}^{\mu^{\otimes k}}_{1-3q}(f^{\otimes k}) \quad &\geq \quad \mathsf{D}^{\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \\
&> \quad \mathsf{sub}^{[]}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k \\
&> \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{[]}(f,\epsilon) - k.
\end{aligned}$$

The last inequality above follows from Theorem 4.2. ■

Proof of Theorem 4.11: The first inequality follows from the definitions. For the second inequality consider a product distribution μ such that $\mathsf{sub}^{\mathbf{1},[]}(f,\epsilon) = \mathsf{sub}^{\mathbf{1},[]}(f,\epsilon,\mu)$. Arguing as in the proof of Lemma 3.1, we get

$$\mathsf{D}^{1,\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \quad > \quad \mathsf{sub}^{1,[]}_{\mathsf{B}}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k.$$

Now since $q \geq 2^{-k}$, we get

$$\begin{aligned}
\mathsf{R}^{1,[]}_{1-3q}(f^{\otimes k}) \quad \geq \quad \mathsf{D}^{1,\mu^{\otimes k}}_{1-3q}(f^{\otimes k}) \quad &\geq \quad \mathsf{D}^{1,\mu^{\otimes k}}_{1-2q-2^{-k}}(f^{\otimes k}) \\
&> \quad \mathsf{sub}^{1,[]}_{\mathsf{B}}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k \\
&\geq \quad \mathsf{sub}^{1,[]}(f^{\otimes k}, 1-2q, \mu^{\otimes k}) - k \log |\mathcal{Z}| - k \qquad (8) \\
&> \quad \frac{\delta\epsilon}{16} \cdot k \cdot \mathsf{sub}^{1,[]}(f, \epsilon, \mu) - k \log |\mathcal{Z}| - k \qquad (9) \\
&= \quad \frac{\delta\epsilon}{8} \cdot k \cdot \mathsf{sub}^{1,[]}(f, \epsilon) - k \log |\mathcal{Z}| - k \\
&\geq \quad k \cdot \left[\frac{\delta\epsilon}{8} \cdot \mathsf{sub}^{1,[]}_{\mathsf{B}}(f, \epsilon) - \log |\mathcal{Z}| - 1 \right]. \qquad (10)
\end{aligned}$$

The Eq. (9) follows from Theorem 4.10. Eq. (8) and (10) follow from Lemma 4.9. ■

ECCC ISSN 1433-8092

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