Theoretical Computer Science 410 (2009) 2463–2477

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

# New bounds on classical and quantum one-way communication complexity

Rahul Jain [a,*], Shengyu Zhang [b]

[a] School of Computer Science and Centre for Quantum Technologies, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077, Singapore
[b] Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

ARTICLE INFO

*Article history:*
Received 30 April 2008
Received in revised form 22 September 2008
Accepted 3 October 2008
Communicated by Y. Shi

*Keywords:*
Communication complexity
Information theory
Quantum
Rectangle bound

ABSTRACT

In this paper we provide new bounds on classical and quantum distributional communication complexity in the two-party, one-way model of communication.

In the classical one-way model, our bound extends the well known upper bound of Kremer, Nisan and Ron [I. Kremer, N. Nisan, D. Ron, On randomized one-round communication complexity, in: Proceedings of The 27th ACM Symposium on Theory of Computing, STOC, 1995, pp. 596–605] to include non-product distributions. Let $\epsilon \in (0, 1/2)$ be a constant. We show that for a boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ and a non-product distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$,

$$\mathsf{D}_{\epsilon}^{1,\mu}(f) = O((I(X : Y) + 1) \cdot \mathsf{VC}(f)),$$

where $\mathsf{D}_{\epsilon}^{1,\mu}(f)$ represents the one-way distributional communication complexity of $f$ with error at most $\epsilon$ under $\mu$; $\mathsf{VC}(f)$ represents the *Vapnik–Chervonenkis* dimension of $f$ and $I(X : Y)$ represents the mutual information, under $\mu$, between the random inputs of the two parties. For a non-boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{1, \ldots, k\}$ ($k \geq 2$ an integer), we show a similar upper bound on $\mathsf{D}_{\epsilon}^{1,\mu}(f)$ in terms of $k$, $I(X : Y)$ and the *pseudo-dimension* of $f' \stackrel{\text{def}}{=} \frac{f}{k}$, a generalization of the VC-dimension for non-boolean functions.

In the quantum one-way model we provide a lower bound on the distributional communication complexity, under product distributions, of a function $f$, in terms of the well studied complexity measure of $f$ referred to as the *rectangle bound* or the *corruption bound* of $f$. We show for a non-boolean total function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and a product distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$,

$$\mathsf{Q}_{\epsilon^3/8}^{1,\mu}(f) = \Omega(\mathsf{rec}_{\epsilon}^{1,\mu}(f)),$$

where $\mathsf{Q}_{\epsilon^3/8}^{1,\mu}(f)$ represents the quantum one-way distributional communication complexity of $f$ with error at most $\epsilon^3/8$ under $\mu$ and $\mathsf{rec}_{\epsilon}^{1,\mu}(f)$ represents the one-way rectangle bound of $f$ with error at most $\epsilon$ under $\mu$. Similarly for a non-boolean partial function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \cup \{*\}$ and a product distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$, we show,

$$\mathsf{Q}_{\epsilon^6/(2\cdot 15^4)}^{1,\mu}(f) = \Omega(\mathsf{rec}_{\epsilon}^{1,\mu}(f)).$$



* Corresponding author. Tel.: +65 15195733541; fax: +65 15198851208.
*E-mail addresses:* rahul.jain.75@gmail.com (R. Jain), syzhang@cse.cuhk.edu.hk (S. Zhang).

0304-3975/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.10.014

## 1. Introduction

Communication complexity studies the minimum amount of communication that two or more parties need to compute a given function or a relation of their inputs. Since its inception in the seminal paper by Yao [22], communication complexity has been an important and widely studied research area. This is the case both because of the interesting and intriguing mathematics involved in its study, and also because of the fundamental connections it bears with many other areas in theoretical computer science, such as data structures, streaming algorithms, circuit lower bounds, decision tree complexity, VLSI designs, etc.

Different models of communication have been proposed and studied. In the basic and standard *two-party interactive* model, two parties Alice and Bob, each receive an input say $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, respectively. They interact with each other possibly communicating several messages in order to jointly compute, say a given function $f(x, y)$ of their inputs. If only one message is allowed, say from Alice to Bob, and Bob outputs $f(x, y)$ without any further interaction with Alice, then the model is called *one-way*. Though seemingly simple, this model has numerous nontrivial questions as well as applications to other areas such as lower bounds for streaming algorithms, see for example [15]. Other models such as the *Simultaneous message passing* (SMP) model, and *multi-party* models are also studied. We refer readers to the textbook [11] for a comprehensive introduction to the field of classical communication complexity. In 1993, Yao [23] introduced *quantum* communication complexity and since then it has also become a very active and vibrant area of research. In the quantum communication models, the parties are allowed to use quantum computers to process their inputs and to use quantum channels to send messages.

In this paper we are primarily concerned with the one-way model and we assume that the single message is always, say from Alice to Bob. Let us first briefly discuss a few classical models. In the deterministic one-way model, the parties act in a deterministic fashion, and compute $f$ correctly on all input pairs $(x, y)$. The minimum communication required for accomplishing this is called the *deterministic complexity* of $f$ and is denoted by $\mathsf{D}^1(f)$. Allowing the parties to use randomness and to err on their inputs with a small non-zero probability, often results in considerable savings in communication. The communication of the best public-coin one-way protocol that has error at most $\epsilon$ on all inputs, is referred to as the *one-way public-coin randomized* communication complexity of $f$ and is denoted by $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(f)$. Similarly we can define the *one-way private-coin randomized* communication complexity of $f$, denoted by $\mathsf{R}_\epsilon^1(f)$ and in the quantum model, the *one-way quantum* communication complexity of $f$, denoted by $\mathsf{Q}_\epsilon^1(f)$. Please refer to Section 2.2 for explicit definitions. When the subscript is omitted, $\epsilon$ is assumed to be $1/3$.

Sometimes the requirement on communication protocols is less stringent and it is only required that the average error, under a given distribution $\mu$ on the inputs, is small. The communication of the best one-way classical protocol that has average error at most $\epsilon$ under $\mu$, is referred to as the *one-way distributional* communication complexity of $f$ and is denoted by $\mathsf{D}_\epsilon^{1,\mu}(f)$. We can define the one-way distributional quantum communication complexity $\mathsf{Q}_\epsilon^{1,\mu}(f)$ in a similar way. A useful connection between the public-coin randomized and distributional communication complexities via the Yao's Principle [21] states that for a given $\epsilon \in (0, 1/2)$, $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(f) = \max_\mu \mathsf{D}_\epsilon^{1,\mu}(f)$. A distribution $\mu$, that achieves the maximum in Yao's Principle, that is for which $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(f) = \mathsf{D}_\epsilon^{1,\mu}(f)$, is referred to as a *hard* distribution for $f$. This principle also holds in many other models and allows for a good handle on the public-coin randomized complexity in scenarios where the distributional complexity is much easier to understand. Often, the distributional complexity when the inputs of Alice and Bob are drawn independently from a product distribution, is easier to understand. Nonetheless, often as is the case with several important functions such as Set Disjointness (DISJ) and Inner Product (IP), the maximum in Yao's Principle, in the one-way model, occurs for a product distribution, and hence it paves the way for understanding the public-coin randomized complexity.

A fundamental question about one-way quantum communication complexity is its relation to the corresponding randomized version for a total function. To be more specific, what is the largest gap between $\mathsf{R}^{1,\mathsf{pub}}(f)$ and $\mathsf{Q}^{1,\mathsf{pub}}(f)$ for a total function $f$? Though some researchers conjecture that they are actually equal to each other up to a multiplicative constant, no subexponential upper bound of $\mathsf{R}^{1,\mathsf{pub}}(f)$ is known in terms of $\mathsf{Q}^{1,\mathsf{pub}}(f)$. To decrease the gap, one may need to prove strong quantum lower bounds and strong classical upper bounds. For instance, if we can find a bound $B(f)$ such that $\mathsf{Q}^{1,\mathsf{pub}}(f) = 1/\mathrm{poly}(B(f))$ and $\mathsf{R}^{1,\mathsf{pub}}(f) = \mathrm{poly}(B(f))$, then we get $\mathsf{R}^{1,\mathsf{pub}}(f) = \mathrm{poly}(\mathsf{Q}^{1,\mathsf{pub}}(f))$. In this paper, we try to prove both classical upper bounds and quantum lower bounds. Detailed discussions of our results with comparison to previous ones follows.

Let us now discuss our first main result which is in the classical one-way model. We ask the reader to refer to Section 2 for the definitions of various quantities involved in the discussion below.

### 1.1. Classical upper bound

For a boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, its *Vapnik–Chervonenkis (VC) dimension*, denoted by $\mathsf{VC}(f)$, is an important complexity measure, widely studied specially in the context of *computational learning theory*. Kremer, Nisan and Ron [12, Thm. 3.2] found a beautiful connection between the distributional complexity of $f$ under product distributions on $\mathcal{X} \times \mathcal{Y}$, and $\mathsf{VC}(f)$, as follows.

**Theorem 1** ([12]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean function and let $\epsilon \in (0, 1/2)$ be a constant. Let $\mu$ be a* product *distribution on $\mathcal{X} \times \mathcal{Y}$. There is a universal constant $\kappa$ such that,*

$$\mathsf{D}_\epsilon^{1,\mu}(f) \le \kappa \cdot \frac{1}{\epsilon} \log \frac{1}{\epsilon} \cdot \mathsf{VC}(f). \qquad (1)$$

Note that such a relation cannot hold for non-product distributions $\mu$ since otherwise it would translate, via the Yao's Principle, into $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(f) = O(\mathsf{VC}(f))$, for all boolean $f$. This is not true as is exhibited by several functions for example the Greater Than ($\mathsf{GT}_n$) function, in which Alice and Bob need to determine which of their $n$-bit inputs is bigger. For this function, $\mathsf{R}_\epsilon^{1,\mathsf{pub}}(\mathsf{GT}_n) = \Theta(n)$ but $\mathsf{VC}(\mathsf{GT}_n) = 1$. Nonetheless for these functions, any hard distribution $\mu$, is highly correlated between $\mathcal{X}$ and $\mathcal{Y}$. Therefore it is conceivable that such a relationship, as in Eq. (1), could still hold, possibly after taking into account the amount of correlation in a given non-product distribution. This question, although probably never explicitly asked in any previous work, appears to be quite fundamental. We answer it in the positive by the following.

**Theorem 2.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean function and let $\epsilon \in (0, 1/2)$ be a constant. Let $\mu$ be a distribution (possibly non-product) on $\mathcal{X} \times \mathcal{Y}$. Let $XY$ be joint random variables distributed according to $\mu$. There is a universal constant $\kappa$ such that,*

$$\mathsf{D}_\epsilon^{1,\mu}(f) \le \kappa \cdot \frac{1}{\epsilon} \log \frac{1}{\epsilon} \cdot \left( \frac{1}{\epsilon} \cdot I(X : Y) + 1 \right) \cdot \mathsf{VC}(f).$$

*In particular, for constant $\epsilon$,*

$$\mathsf{D}_\epsilon^{1,\mu}(f) = O\left( (I(X : Y) + 1) \cdot \mathsf{VC}(f) \right).$$

*Above $I(X : Y)$ represents the mutual information between correlated random variables $X$ and $Y$, distributed according to $\mu$.*

Let us discuss below a few aspects of this result and its relationship with what is previously known. Note that in combination with Yao's Principle, Theorem 2 gives us the following (where the mutual information is now considered under a hard distribution for $f$).

$$\mathsf{R}^{1,\mathsf{pub}}(f) = O\left( (I(X : Y) + 1) \cdot \mathsf{VC}(f) \right). \tag{2}$$

1. It is easily observed using Sauer's Lemma (Lemma 2, Section 2.) that the deterministic complexity of $f$ has
$$\mathsf{D}^1(f) = O(\mathsf{VC}(f) \cdot \log |\mathcal{Y}|). \tag{3}$$
This is because Alice can simply send the name of $f_x$ in $O(\mathsf{VC}(f) \cdot \log |\mathcal{Y}|)$ bits since $|\mathcal{F}| \le |\mathcal{Y}|^{\mathsf{VC}(f)}$. Now our result (2) is on one hand stronger than (3) in the sense $I(X : Y) \le \log |\mathcal{Y}|$ always, and $I(X : Y)$ could be much smaller than $\log |\mathcal{Y}|$ depending on $\mu$. An example of such a case is the Inner Product ($\mathsf{IP}_n$) function in which Alice and Bob need to determine the inner product (mod 2) of their $n$-bit input strings. For $\mathsf{IP}_n$, a hard distribution is the uniform distribution which is product, and hence $I(X : Y) = 0$, whereas $\log |\mathcal{Y}| = n$. However on the other hand (2) is also weaker than (3) in the sense it only upper bounds the public-coin randomized complexity, whereas (2) upper bounds the deterministic complexity of $f$.
2. Aaronson [1] shows that for a total or partial boolean function $f$,
$$\mathsf{R}^1(f) = O(\mathsf{Q}^1(f) \cdot \log |\mathcal{Y}|). \tag{4}$$
Again (2) is stronger than (4) in the sense that $I(X : Y)$ could be much smaller than $\log |\mathcal{Y}|$ depending on $\mu$. Also it is known that, $\mathsf{Q}^1(f) = \Omega(\mathsf{VC}(f))$ always, following from Nayak [16], and $\mathsf{Q}^1(f)$ could be much larger than $\mathsf{VC}(f)$. An example is the Greater Than ($\mathsf{GT}_n$) function for which $\mathsf{Q}^1(\mathsf{GT}_n) = \Omega(n)$, whereas $\mathsf{VC}(\mathsf{GT}_n) = O(1)$. On the other hand (2) only holds for total boolean functions whereas (4) also holds for partial boolean functions.
3. As mentioned before, for all total boolean functions $f$, $\mathsf{R}^{1,\mathsf{pub}}(f) = \Omega(\mathsf{VC}(f))$, and $\mathsf{R}^{1,\mathsf{pub}}(f)$ could be much larger than $\mathsf{VC}(f)$ (as in function $\mathsf{GT_n}$). Now Eq. (2) says that in the latter case, the mutual information $I(X : Y)$ under any hard distribution $\mu$ must be large. That is, a hard distribution $\mu$ must be highly correlated.
4. It is known that for total boolean functions $f$, for which a hard distribution is the product, there is no separation between the one-way public-coin randomized and quantum communication complexities. Now our theorem gives a *smooth* extension of this fact to the functions whose hard distributions are not product ones. Note that for most, if not all, specific functions of interest such as EQ, IP, DISJ, etc., the mutual information of a hard distribution is very easy to calculate.

A generalization of the VC-dimension for non-boolean functions, is referred to as the *pseudo-dimension* (Definition 2, Section 2). For a non-boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{1, \dots, k\}$ ($k \ge 2$ an integer), we show a similar upper bound on $\mathsf{D}_\epsilon^{1,\mu}(f)$ in terms of $k$, $I(X : Y)$ and the pseudo-dimension of $f' \stackrel{\mathsf{def}}{=} \frac{f}{k}$.

**Theorem 3.** *Let $k \ge 2$ be an integer. Let $f : \mathcal{X} \times \mathcal{Y} \to \{1, \dots, k\}$ and $\epsilon \in (0, 1/6)$ be a constant. Let $f' : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ be such that $f'(x, y) = f(x, y)/k$. Let $\mu$ be a distribution (possibly non-product) on $\mathcal{X} \times \mathcal{Y}$, and $XY$ be joint random variables distributed according to $\mu$. Then there is a universal constant $\kappa$ such that,*

$$\mathsf{D}_{3\epsilon}^{1,\mu}(f) \le \kappa \cdot \frac{k^4}{\epsilon^5} \cdot \left( \log \frac{1}{\epsilon} + d \log^2 \frac{dk}{\epsilon} \right) \cdot (I(X : Y) + \log k)$$

*where $d \stackrel{\mathsf{def}}{=} \mathcal{P}_{\frac{\epsilon^2}{576k^2}}(f')$ is the $\frac{\epsilon^2}{576k^2}$-pseudo-dimension of $f'$.*

Let us now discuss our other main result which we show in the quantum one-way model.

*1.2. Quantum lower bound*

For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, a measure of its complexity that is often very useful in understanding its classical randomized communication complexity, is the *rectangle bound* (denoted by $\mathsf{rec}(f)$), also often known as the *corruption bound*. The rectangle bound $\mathsf{rec}(f)$ is actually defined first via a distributional version $\mathsf{rec}^{\mu}(f)$. It is a well studied measure and $\mathsf{rec}^{\mu}(f)$ is well known to form a lower bound on $\mathsf{D}^{\mu}(f)$ both in the one-way and two-way models. In fact, in a celebrated result, Razborov [19] provided optimal lower bound on the randomized communication complexity of the Set Disjointness function, by arguing a lower bound on its rectangle bound.

It is natural to ask if this measure also forms a lower bound on the quantum communication complexity. We answer this question in the positive for this question in the one-way model. We show that, for a total or partial function, the quantum distributional one-way communication complexity under a given product distribution $\mu$ is lower bounded by the corresponding one-way rectangle bound. Our precise result is as follows:

**Theorem 4.** *Let $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ be a total function and let $\epsilon \in (0, 1/2)$ be a constant. Let $\mu$ be a product distribution on $\mathcal{X} \times \mathcal{Y}$ and let $\mathsf{rec}_{\epsilon}^{1,\mu}(f) > 2 \cdot \log(1/\epsilon)$. Then,*

$$\mathsf{Q}^{1,\mu}_{\epsilon^3/8}(f) \geq \frac{1}{2} \cdot (1 - 2\epsilon) \cdot (S(\epsilon/2) - S(\epsilon/4)) \cdot (\lfloor \mathsf{rec}_{\epsilon}^{1,\mu}(f) \rfloor - 1) = \Omega(\mathsf{rec}_{\epsilon}^{1,\mu}(f)), \tag{5}$$

*where for $p \in (0, 1)$, $S(p)$ is the binary entropy function $S(p) \stackrel{\mathsf{def}}{=} -p \log p - (1 - p) \log(1 - p)$.*

*If $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z} \cup \{*\}$ is a partial function then,*

$$\mathsf{Q}^{1,\mu}_{\epsilon^6/(2\cdot 15^4)}(f) \geq \frac{1}{2} \cdot (1 - 2\epsilon) \cdot \frac{\epsilon^2}{300} \cdot (\lfloor \mathsf{rec}_{\epsilon}^{1,\mu}(f) \rfloor - 1) = \Omega(\mathsf{rec}_{\epsilon}^{1,\mu}(f)).$$

Let us make a few important remarks here related to this result.

1. Recently, Jain, Klauck and Nayak [9] showed that for any relation $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, the rectangle bound of $f$ tightly characterizes the randomized one-way classical communication complexity of $f$.

   **Theorem 5** (*[9]*). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\epsilon \in (0, 1/2)$. Then,*

   $$\mathsf{R}^{1,\mathsf{pub}}_{\epsilon}(f) = \Theta(\mathsf{rec}^{1}_{\epsilon}(f)).$$

   While showing Theorem 5, Jain, Klauck and Nayak [9] have shown that for all relations $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ and for all distributions $\mu$ (product and non-product) on $\mathcal{X} \times \mathcal{Y}$; $\mathsf{D}^{1,\mu}_{\epsilon}(f) = \Omega(\mathsf{rec}^{1,\mu}_{4\epsilon}(f))$. However in the quantum setting we are making a similar statement only for (total or partial) functions $f$ and only for product distributions $\mu$ on $\mathcal{X} \times \mathcal{Y}$. In fact it **does NOT** hold if we let $\mu$ to be non-product. It can be shown that there is a total function $f$ and a non-product distribution $\mu$ such that $\mathsf{Q}^{1,\mu}_{\epsilon}(f)$ is exponentially smaller than $\mathsf{rec}^{1,\mu}_{\epsilon}(f)$. This fact is implicit in the work of Gavinsky et al. [5]. We make an explicit statement of this in Appendix and skip its proof for brevity.
2. Let $\epsilon \in (0, 1/4)$. Jain, Klauck and Nayak [9] have shown that for all relations $g \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$,

   $$\mathsf{R}^{1,[]}_{2\epsilon}(g) = O(\mathsf{rec}^{1,[]}_{\epsilon}(g)).$$

   Here the superscript [] represents maximization over all product distributions. From Theorem 4 for a (total or partial) function $f$ we get,

   $$\mathsf{Q}^{1,[]}_{\epsilon^6/(2\cdot 15^4)}(f) = \Omega(\mathsf{rec}^{1,[]}_{\epsilon}(f)).$$

   Since $\mathsf{R}^{1,[]}_{\epsilon}(f) \geq \mathsf{Q}^{1,[]}_{\epsilon}(f)$, combining everything we get,

   **Theorem 6.** *Let $\epsilon \in (0, 1/4)$. Let $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z} \cup \{*\}$ be a (possibly partial and non-boolean) function. Then*

   $$\mathsf{R}^{1,[]}_{\epsilon^6/(2\cdot 15^4)}(f) \geq \mathsf{Q}^{1,[]}_{\epsilon^6/(2\cdot 15^4)}(f) = \Omega(\mathsf{R}^{1,[]}_{2\epsilon}(f)).$$

   It was known earlier that for total boolean functions, $\mathsf{Q}^{1,[]}(f)$ is tightly bounded by $\mathsf{R}^{1,[]}(f)$. We extend such a relationship here to apply for non-boolean (partial) functions as well. We remark that the earlier proofs for total boolean functions used the VC-dimension result, Theorem 1, of Kremer, Nisan and Ron [12]. We get the same result here without requiring it.

We finally present an application of our result Theorem 4 in the context of studying security of extractors against *quantum adversaries*. An extractor is a function that is used to extract almost uniform randomness from a source of imperfect randomness. Extractors are well studied objects and have found several uses in many cryptographic applications and also in complexity theory. Recently, security of various extractors has been increasingly studied in the presence of quantum adversaries; since such secure extractors are then useful in several applications such as privacy amplification in quantum key distribution and key-expansion in quantum bounded storage models [10,13,14]. In particular, König and Terhal [14] have shown that any boolean extractor that can extract a uniform bit from sources of *min-entropy k* is also secure against quantum adversaries with their memory bounded by a function of *k*.

We get a similar statement for boolean extractors, as a corollary of our result Theorem 4. We obtain this corollary by observing a key connection between the minimum min-entropy that an extractor function $f$ needs to extract a uniform bit and its rectangle bound. The precise statement of our result, its relationship with the result of [14], and a detailed discussion is deferred to Section 5.

*1.3. Organization*

In the following Section 2 we discuss various information theoretic preliminaries and the model of one-way communication. In Section 3 we present the upper bounds in the classical setting and in Section 4 we present the lower bounds in the quantum setting. The application concerning extractors is discussed in Section 5. We finally conclude with some open questions in Section 6.

## 2. Preliminaries

*2.1. Information theory*

In this section we present some information theoretic notation, definitions and facts that we use in the rest of the paper. For an introduction to classical and quantum information theory, we refer the reader to the texts by Cover and Thomas [4] and Nielsen and Chuang [17] respectively. Most of the facts stated in this section without proofs may be found in these books.

All logarithms in this paper are taken with base 2, unless otherwise specified. 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 $A$, $\|A\|_1 \stackrel{\text{def}}{=} \mathsf{Tr}(\sqrt{A^\dagger A})$ denotes its $\ell_1$ norm. For $p \in (0, 1)$, let $S(p) \stackrel{\text{def}}{=} -p \log p - (1-p)\log(1-p)$, denote the binary entropy function. We have the following fact.

**Fact 1.** *For* $\delta \in [0, 1/2]$, $\quad S(\frac{1}{2} + \delta) \leq 1 - 2\delta^2$ *and* $S(\delta) \leq 2\sqrt{\delta}$.

A quantum state, usually represented by letters $\rho, \sigma$ etc., is a positive semi-definite trace one operator in a given Hilbert space. 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 $P(\mathcal{E}) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{E}} P(x)$ denotes the probability of the event $\mathcal{E} \subseteq \mathcal{X}$. A distribution $P$ on $\mathcal{X} \times \mathcal{Y}$ is said to be *product* across $\mathcal{X}$ and $\mathcal{Y}$, if it can be written as $P = P_{\mathcal{X}} \otimes P_{\mathcal{Y}}$, where $P_{\mathcal{X}}, P_{\mathcal{Y}}$ are distributions on $\mathcal{X}, \mathcal{Y}$ respectively and $\otimes$ is the tensor operation. Often for product distributions we do not mention the sets across which it is product if it is clear from the context.

Let $X$ be a classical random variable (or simply random variable) taking values in $\mathcal{X}$. For a random variable $X$, we also let $X$ represent its probability distribution. The *entropy* of $X$ denoted $S(X)$ is defined to be $S(X) \stackrel{\text{def}}{=} -\mathsf{Tr} X \log X$. Since $X$ is classical an equivalent definition would be $S(X) \stackrel{\text{def}}{=} -\sum_{x \in \mathcal{X}} \Pr[X = x] \log \Pr[X = x]$ . Let $X, Y$ be a correlated random variables taking values in $\mathcal{X}, \mathcal{Y}$ respectively. $XY$ are said to be *independent* if their joint distribution is product. The *mutual information* between them, denoted $I(X : Y)$ is defined to be $I(X : Y) \stackrel{\text{def}}{=} S(X) + S(Y) - S(XY)$ and *conditional entropy* denoted $S(X|Y)$ is defined to be $S(X|Y) \stackrel{\text{def}}{=} S(XY) - S(Y)$. It is easily seen that $S(X|Y) = \mathbf{E}_{y \leftarrow Y}[S(X|(Y = y))]$.

We have the following facts.

**Fact 2.** *For all random variables* $X, Y$; $I(X : Y) \geq 0$*; in other words* $S(X) + S(Y) \geq S(XY)$*. If* $X, Y$ *are independent then we have* $I(X : Y) = 0$*; in other words* $S(XY) = S(X) + S(Y)$.

The definitions and facts stated in the above paragraph for classical random variables also hold *mutatis mutandis* for quantum states as well. For example for a quantum state $\rho$, its entropy is defined as $S(\rho) \stackrel{\text{def}}{=} -\mathsf{Tr}\rho \log \rho$. For brevity, we avoid making all the corresponding statements explicitly. As is the case with classical random variables, for a quantum system say $Q$, we also often let $Q$ represent its quantum state. We have the following fact.

**Fact 3.** *Any quantum state* $\rho$ *in m-qubits has* $S(\rho) \leq m$*. Also let* $XQ$ *be a joint classical-quantum system with* $X$ *being a classical random variable, then* $I(X : Q) \leq \min\{S(X), S(Q)\}$.

For a system $XYM$, let us define $I(X : M|Y) \stackrel{\text{def}}{=} S(X|Y) + S(M|Y) - S(XM|Y)$. If $Y$ is a classical system then it is easily seen that $I(X : M|Y) = \mathbf{E}_{y \leftarrow Y}[I(X : M|(Y = y))]$.

For random variables $X_1, \ldots, X_n$ and a correlated (possibly quantum) system $M$, we have the following *chain rule of mutual information*, which will be crucially used in our proofs.

$$I(X_1 \ldots X_n : M) = \sum_{i=1}^{n} I(X_i : M | X_1 \ldots X_{i-1}). \tag{6}$$

By convention, conditioning on $X_1 \ldots X_{i-1}$ for $i = 1$ means conditioning on the true event. The following is an important information theoretic fact known as Fano's inequality, which relates the probability of disagreement for correlated random variables to their mutual information.

**Lemma 1** (*Fano's Inequality*). *Let* $X$ *be a random variable taking values in* $\mathcal{X}$*. Let* $Y$ *be a correlated random variable and let* $P_e \stackrel{\text{def}}{=} \Pr(X \neq Y)$*. Then,*

$$S(P_e) + P_e \log(|\mathcal{X}| - 1) \geq S(X|Y).$$

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 1** (*Vapnik–Chervonenkis (VC) Dimension*). A set $S \subseteq \mathcal{Y}$ is said to be *shattered* by a set $\mathcal{G}$ of boolean functions from $\mathcal{Y}$ to $\{0, 1\}$, if $\forall R \subseteq S$, $\exists g_R \in \mathcal{G}$ such that $\forall s \in S$, $(s \in R) \Leftrightarrow (g_R(s) = 1)$. The largest value $d$ for which there is a set $S$ of size $d$ that is shattered by $\mathcal{G}$ is the *Vapnik–Chervonenkis dimension* of $\mathcal{G}$ and is denoted by $\mathsf{VC}(\mathcal{G})$.

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 $f_x(y) \stackrel{\text{def}}{=} f(x, y)$, $\forall y \in \mathcal{Y}$. Let $\mathcal{F} \stackrel{\text{def}}{=} \{f_x : x \in \mathcal{X}\}$. Then the Vapnik–Chervonenkis dimension of $f$, denoted by $\mathsf{VC}(f)$, is defined to be $\mathsf{VC}(\mathcal{F})$.

Let $f$ and $\mathcal{F}$ be as defined in the above definition. We call a function $f$ *trivial* iff $|\mathcal{F}| = 1$, in other words iff the value of the function, for all $x$, is determined only by $y$. We call $f$ *non-trivial* iff it is not trivial. Note that a boolean $f$ is non-trivial if and only if $\mathsf{VC}(f) \geq 1$. Throughout this paper we assume all our functions to be non-trivial. Following is a useful fact, with several applications, relating the VC-dimension of $f$ to the size of $\mathcal{F}$. It is usually attributed to Sauer [20], however it has been independently discovered by several different people as well.

**Lemma 2** (*Sauer's Lemma [20]*). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean function. Let $d \stackrel{\text{def}}{=} \mathsf{VC}(f)$. Let $m \stackrel{\text{def}}{=} |\mathcal{Y}|$, then*

$$|\mathcal{F}| \leq \sum_{i=0}^{d} \binom{m}{i} \leq m^d.$$

The following result from Blumer, Ehrenfeucht, Haussler, and Warmuth [2] is one of the most fundamental results from computational learning theory and in fact an important application of Sauer's Lemma.

**Lemma 3.** *Let $H$ be class of boolean functions over a finite domain $\mathcal{Y}$ with VC-dimension $d$, let $\pi$ be an arbitrary probability distribution over $\mathcal{Y}$, and let $0 < \epsilon, \delta < 1$. Let $L$ be any algorithm that takes as input a set $S \in \mathcal{Y}^m$ of $m$ examples labeled according to an unknown function $h \in H$, and outputs a hypothesis function $h' \in H$ that is consistent with $h$ on the sample $S$. If $L$ receives a random sample of size $m \geq m_0(d, \epsilon, \delta)$ distributed according to $\pi^m$, where*

$$m_0(d, \epsilon, \delta) = c_0 \left( \frac{1}{\epsilon} \log \frac{1}{\delta} + \frac{d}{\epsilon} \log \frac{1}{\epsilon} \right)$$

*for some constant $c_0 > 0$, then with probability at least $1 - \delta$ over the random samples, $\Pr_\pi[h'(y) \neq h(y)] \leq \epsilon$.*

A similar learning result also holds for non-boolean functions. For this let us first define the following generalization of the VC-dimension, known as the *pseudo-dimension*.

**Definition 2** (*Pseudo-Dimension*). A set $S \subseteq \mathcal{Y}$ is said to be $\gamma$*-shattered* by a set $\mathcal{G}$ of functions from $\mathcal{Y}$ to $\mathcal{Z} \subseteq \mathbb{R}$, if there exists a vector $\mathbf{w} = (w_1, \ldots, w_k) \in \mathcal{Z}^k$ of dimension $k = |S|$ for which the following holds. For all $R \subseteq S$, $\exists g_R \in \mathcal{G}$ such that $\forall s \in S$, $(s \in R) \Rightarrow (g_R(s) > w_i + \gamma)$ and $(s \notin R) \Rightarrow (g_R(s) < w_i - \gamma)$. The largest value $d$ for which there is a set $S$ of size $d$ that is $\gamma$-shattered by $\mathcal{G}$ is the $\gamma$*-pseudo-dimension* of $\mathcal{G}$ and is denoted by $\mathcal{P}_\gamma(\mathcal{G})$.

Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. For all $x \in \mathcal{X}$ let $f_x : \mathcal{Y} \to \mathcal{Z}$ be defined as $f_x(y) \stackrel{\text{def}}{=} f(x, y)$, $\forall y \in \mathcal{Y}$. Let $\mathcal{F} \stackrel{\text{def}}{=} \{f_x : x \in \mathcal{X}\}$. Then the $\gamma$-pseudo-dimension of $f$, denoted by $\mathcal{P}_\gamma(f)$, is defined to be $\mathcal{P}_\gamma(\mathcal{F})$.

The following result of Bartlett, Long and Williamson [3] is similar to the learning lemma of Blumer et al. [2] and concerns non-boolean functions.

**Theorem 7.** *Let $\mathcal{G}$ be a class of functions over a finite domain $\mathcal{Y}$ into the range $[0, 1]$. Let $\pi$ be an arbitrary probability distribution over $\mathcal{Y}$ and let $\epsilon \in (0, 1/2)$ and $\delta \in (0, 1)$. Let $d \stackrel{\text{def}}{=} \mathcal{P}_{\epsilon^2/576}(\mathcal{G})$. Then there exists a deterministic learning algorithm $L$ which has the following property. Given as input a set $S \in \mathcal{Y}^m$ of $m$ examples chosen according to $\pi^m$ and labeled according to an unknown function $g \in \mathcal{G}$, $L$ outputs a hypothesis $g' \in \mathcal{G}$ such that if $m \geq m_0(d, \epsilon, \delta)$ where*

$$m_0(d, \epsilon, \delta) = c_0 \left( \frac{1}{\epsilon^4} \log \frac{1}{\delta} + \frac{d}{\epsilon^4} \log^2 \frac{d}{\epsilon} \right)$$

*for some constant $c_0 > 0$, then with probability at least $1 - \delta$ over the random samples,*

$$\sum_{y \in \mathcal{Y}} \pi(y) \cdot |h'(y) - h(y)| \leq \epsilon.$$

Following is a fundamental quantum information theoretic fact shown by Holevo [8].

**Theorem 8** (*The Holevo Bound [8]*). *Let $X$ be classical random variable taking values in $\mathcal{X}$. Let $M$ be a quantum system and let $Y$ be a random variable obtained by performing a quantum measurement on $M$. Then,*

$$I(X : Y) \leq I(X : M). \tag{7}$$

The following is an interesting and useful information theoretic fact first shown by Helstrom [6].

**Theorem 9** (*[6]*). *Let $XQ$ be a joint classical-quantum system where $X$ is a classical boolean random variable. For $a \in \{0, 1\}$, let the quantum state of $Q$ when $X = a$ be $\rho_a$. The optimal success probability of predicting $X$ with a measurement on $Q$ is given by*

$$\frac{1}{2} + \frac{1}{2} \cdot \| \Pr[X = 0]\rho_0 - \Pr[X = 1]\rho_1 \|_1.$$

*2.2. One-way communication*

In this article we only consider the two-party one-way model of communication. Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. The relations we consider are always total in the sense that for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$, there is at least one $z \in \mathcal{Z}$, such that $(x, y, z) \in f$. In a one-way protocol $\mathcal{P}$ for computing $f$, Alice and Bob get inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. Alice sends a single message to Bob, and their intention is to determine an answer $z \in \mathcal{Z}$ such that $(x, y, z) \in f$. In the one-way protocols we consider, the single message is always from Alice to Bob. A total function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, can be viewed as a type of relation in which for every $(x, y)$ there is a unique $z$, such that $(x, y, z) \in f$. A partial function is a special type of relations such that for some inputs $(x, y)$, there is a unique $z$, such that $(x, y, z) \in f$ and for all other inputs $(x, y)$, $(x, y, z) \in f$, $\forall z \in \mathcal{Z}$. We view a partial function $f$ as a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \cup \{*\}$, such that the inputs $(x, y)$ for which $f(x, y) = *$ are exactly the ones for which $(x, y, z) \in f$, $\forall z \in \mathcal{Z}$.

Let us first consider classical communication protocols. We let $\mathsf{D}^1(f)$ represent the deterministic one-way communication complexity, that is the communication of the best deterministic protocol computing $f$ correctly on all inputs. For $\epsilon \in (0, 1/2)$, let $\mu$ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. We let $\mathsf{D}^{1,\mu}_\epsilon(f)$ represent the distributional one-way communication complexity of $f$ under $\mu$ with expected error $\epsilon$, 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 $\epsilon$ under $\mu$. It is easily noted that $\mathsf{D}^{1,\mu}_\epsilon(f)$ is always achieved by a deterministic one-way protocol, and will henceforth restrict ourselves to deterministic protocols in the context of distributional communication complexity. We let $\mathsf{R}^{1,\mathsf{pub}}_\epsilon(f)$ represent the public-coin randomized one-way communication complexity of $f$ with worst case error $\epsilon$, 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 $\epsilon$. The analogous quantity for private coin randomized protocols is denoted by $\mathsf{R}^1_\epsilon(f)$. The public- and private-coin randomized communication complexities are not much different, as shown in Newman's result [18] that

$$R^1(f) = O(R^{1,pub}(f) + \log\log|\mathcal{X}| + \log\log|\mathcal{Y}|). \tag{8}$$

The following result due to Yao [21] is a very useful fact connecting worst-case and distributional communication complexities. It is a consequence of the *min–max* theorem in game theory [11, Thm. 3.20, page 36].

**Lemma 4** (*Yao's Principle [21]*). $\mathsf{R}^{1,\mathsf{pub}}_\epsilon(f) = \max_\mu \mathsf{D}^{1,\mu}_\epsilon(f)$.

We define $\mathsf{R}^{1,[]}_\epsilon(f) \stackrel{\text{def}}{=} \max_{\mu \text{ product}} \mathsf{D}^{1,\mu}_\epsilon(f)$. Note that $\mathsf{R}^{1,[]}_\epsilon(f)$ could be significantly smaller than $\mathsf{R}^{1,\mathsf{pub}}_\epsilon(f)$ as is exhibited by the Greater Than ($\mathsf{GT}_n$) function for which $\mathsf{R}^{1,\mathsf{pub}}(\mathsf{GT}_n) = \Omega(n)$, whereas $\mathsf{R}^{1,[]}_\epsilon(f) = O(1)$.

In a one-way quantum communication protocol, Alice and Bob are allowed to do quantum operations and Alice can send a quantum message (qubits) to Bob. Given $\epsilon \in (0, 1/2)$, the one-way quantum communication complexity $\mathsf{Q}^1_\epsilon(f)$ is defined to be the communication of the best one-way quantum protocol with error at most $\epsilon$ on all inputs. Given a distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$, we can similarly define the quantum distributional one-way communication complexity of $f$, denoted $\mathsf{Q}^{1,\mu}_\epsilon(f)$, to be the communication of the best one-way quantum protocol $\mathcal{P}$ for $f$ such that the average error of $\mathcal{P}$ over the inputs drawn from the distribution $\mu$ is at most $\epsilon$. We define $\mathsf{Q}^{1,[]}_\epsilon(f) \stackrel{\text{def}}{=} \max_{\mu \text{ product}} \mathsf{Q}^{1,\mu}_\epsilon(f)$.

## 3. A new upper bound on classical one-way distributional communication complexity

In this section we present the upper bounds on the distributional communication complexity, $\mathsf{D}^{1,\mu}_\epsilon(f)$ for any distribution $\mu$ (possibly non-product) on $\mathcal{X} \times \mathcal{Y}$. We begin by restating the precise result for boolean functions.

**Theorem 10.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean function and let $\epsilon \in (0, 1/2)$ be a constant. Let $\mu$ be a distribution (possibly non-product) on $\mathcal{X} \times \mathcal{Y}$. Let $XY$ be joint random variables distributed according to $\mu$. There is a universal constant $\kappa$ such that,*

$$\mathsf{D}^{1,\mu}_\epsilon(f) \le \kappa \cdot \frac{1}{\epsilon}\log\frac{1}{\epsilon} \cdot \left(\frac{1}{\epsilon} \cdot I(X:Y) + 1\right) \cdot \mathsf{VC}(f).$$

*In other words,*

$$\mathsf{D}^{1,\mu}_\epsilon(f) = O\left((I(X:Y) + 1) \cdot \mathsf{VC}(f)\right).$$

For showing this result we will crucially use the following fact shown by Harsha, Jain, McAllester and Radhakrishnan [7] concerning communication required for generating correlations. We begin with the following definition.

**Definition 3** (*Correlation Protocol*). Let $(X, Y)$ be a pair of correlated random variables taking values in $\mathcal{X} \times \mathcal{Y}$. Let Alice be given $x \in \mathcal{X}$, sampled according to the distribution $X$. Alice should transmit a message to Bob, such that Alice and Bob can together generate a value $y \in \mathcal{Y}$ distributed according to the conditional distribution $Y|_{X=x}$; that is the pair $(x, y)$ should have joint distribution $(X, Y)$. Alice and Bob are allowed to use public randomness. Note that the generated value $y$ should be known to both Alice and Bob.

Harsha et al. [7] showed that the minimal expected number of bits that Alice needs to send (in the presence of shared randomness), denoted $T^R(X:Y)$, is characterized by the mutual information $I(X:Y)$ as follows.

**Theorem 11** ([7]). *There exists a universal positive constant $l$ such that,*

$$I(X:Y) \leq T^R(X:Y) \leq 4I(X:Y) + l.$$

We will also need the following lemma.

**Lemma 5.** *Let $m \geq 1$ be an integer. Let $XY$ be correlated random variables. Let $\mu_x$ be the distribution of $Y|X = x$. Let $Y'$ represent another random variable correlated with $X$ such that the distribution of $Y'|(X = x)$ is $\mu_x^{\otimes m}$ ($m$ independent copies of $\mu_x$). Then,*

$$I(X:Y') \leq m \cdot I(X:Y).$$

**Proof.** Consider,

$$\begin{aligned} I(X:Y') &= S(Y') - \mathbf{E}_{x\leftarrow X}[S(Y'|X = x)] \\ &= S(Y') - m \cdot \mathbf{E}_{x\leftarrow X}[S(Y|X = x)] \\ &\leq m \cdot S(Y) - m \cdot \mathbf{E}_{x\leftarrow X}[S(Y|X = x)] \\ &= m \cdot I(X:Y). \end{aligned}$$

The first inequality above follows from Fact 2 by noting that $Y'$ is $m$-copies of $Y$. □

We are now ready for the proof of Theorem 10.

**Proof of Theorem 10.** Let $m \stackrel{\text{def}}{=} m_0(\mathsf{VC}(f), \epsilon/4, \epsilon/4) = c_0 \cdot \left(\frac{1}{\epsilon/4}\log\frac{1}{\epsilon/4}\right) \cdot (\mathsf{VC}(f) + 1)$ as in Lemma 3. Let $l$ be the constant as in Theorem 11. Let $c \stackrel{\text{def}}{=} 4m \cdot I(X:Y) + l$. We exhibit a public coin protocol $\mathcal{P}$ with inputs drawn from $\mu$, in which Alice sends two messages $M_1$ and $M_2$ to Bob. The expected length of $M_1$ is at most $c$ and the length of $M_2$ is always at most $m$. The average error (over inputs and coins) of $\mathcal{P}$ is at most $\epsilon/2$. Let $\mathcal{P}'$ be the protocol that simulates $\mathcal{P}$ but aborts and outputs 0, whenever the length of $M_1$ in $\mathcal{P}$ exceeds $2c/\epsilon$. From Markov's inequality this happens with probability at most $\epsilon/2$. Hence the expected error of $\mathcal{P}'$ is at most $\epsilon/2 + \epsilon/2 = \epsilon$. Since the expected error (over coins and inputs) of $\mathcal{P}'$ is at most $\epsilon$, there exists a deterministic protocol (by fixing coins suitably) with communication bounded by $2c/\epsilon + m$ and distributional error at most $\epsilon$. This implies our result from definition of $\mathsf{D}^{1,\mu}_{\epsilon}(f)$ and by setting $\kappa$ appropriately.

For $x \in \mathcal{X}$, let $\mu_x$ be the distribution of $Y|X = x$. In $\mathcal{P}$, on receiving the input $x \in \mathcal{X}$, Alice first sends a message $M_1$ to Bob, according to the corresponding correlation protocol as in Definition 3, and they together sample from the distribution of $\mu_x^{\otimes m}$. Let $y_1, \ldots, y_m$ be the samples generated. Note that from the properties of correlation protocol both Alice and Bob know the values of $y_1, \ldots, y_m$. Alice then sends to Bob the second message $M_2$ which is the values of $f(x, y_1), \ldots, f(x, y_m)$. Bob then considers the first $x'$ (according to the lexicographically increasing order) such that $\forall i \in [m], f(x', y_i) = f(x, y_i)$ and outputs $f(x', y)$, where $y$ is his actual input. Using Lemma 3, it is easy to verify that for every $x \in \mathcal{X}$, the average error (over randomness in the protocol and inputs of Bob) in this protocol $\mathcal{P}$ will be at most $\epsilon/2$. Hence also the overall average error of $\mathcal{P}$ is at most $\epsilon/2$. Also from Theorem 11 and Lemma 5, we can verify that the expected length of $M_1$ in $\mathcal{P}$ will be at most $4m \cdot I(X:Y) + l$. □

Following similar arguments and using Theorems 7 and 11, we obtain a similar result for non-boolean functions as follows.

**Theorem 12.** *Let $k \geq 2$ be an integer. Let $f : \mathcal{X} \times \mathcal{Y} \to [k]$ be a non-boolean function and let $\epsilon \in (0, 1/6)$ be a constant. Let $f' : \mathcal{X} \times \mathcal{Y} \to [0, 1]$ be such that $f'(x, y) = f(x, y)/k$. Let $\mu$ be a distribution (possibly non-product) on $\mathcal{X} \times \mathcal{Y}$. Let $XY$ be joint random variables distributed according to $\mu$. There is a universal constant $\kappa$ such that,*

$$\mathsf{D}^{1,\mu}_{3\epsilon}(f) \leq \kappa \cdot \frac{k^4}{\epsilon^5} \cdot \left(\log\frac{1}{\epsilon} + d\log^2\frac{dk}{\epsilon}\right) \cdot (I(X:Y) + \log k)$$

*where $d \stackrel{\text{def}}{=} \mathcal{P}_{\frac{\epsilon^2}{576k^2}}(f')$ is the $\frac{\epsilon^2}{576k^2}$-pseudo-dimension of $f'$.*

**Proof.** Let $m \stackrel{\text{def}}{=} m_0(d, \epsilon/k, \epsilon) = c_0\left(\frac{k^4}{\epsilon^4}\log\frac{1}{\epsilon} + \frac{dk^4}{\epsilon^4}\log^2\frac{dk}{\epsilon}\right)$ as in Theorem 7. Let $l$ be the constant as in Theorem 11. Let $c \stackrel{\text{def}}{=} 4m \cdot I(X:Y) + l$. We exhibit a public coin protocol $\mathcal{P}$ for $f$, with inputs drawn from $\mu$, in which Alice sends two messages $M_1$ and $M_2$ to Bob. The expected length of $M_1$ is at most $c$ and the length of $M_2$ is always at most $O(m\log k)$. The average error (over inputs and coins) of $\mathcal{P}$ is at most $2\epsilon$. Let $\mathcal{P}'$ be the protocol that simulates $\mathcal{P}$ but aborts and outputs 0, whenever the length of $M_1$ in $\mathcal{P}$ exceeds $c/\epsilon$. From Markov's inequality this happens with probability at most $\epsilon$. Hence the expected error (over coins and inputs) of $\mathcal{P}'$ is at most $2\epsilon + \epsilon = 3\epsilon$. From $\mathcal{P}'$, by fixing coins suitably, we finally get a deterministic protocol with communication bounded by $c/\epsilon + O(m\log k)$ and distributional error at most $3\epsilon$. This implies our result from definition of $\mathsf{D}^{1,\mu}_{3\epsilon}(f)$ and by setting $\kappa$ appropriately.

In $\mathcal{P}$, Alice and Bob intend to first determine $f'(x, y)$ and then output $kf'(x, y)$. For $x \in \mathcal{X}$, let $\mu_x$ be the distribution of $Y|X = x$. On receiving the input $x \in \mathcal{X}$, Alice first sends a message $M_1$ to Bob, according to the corresponding correlation protocol as in Definition 3, and they together sample from the distribution of $\mu_x^{\otimes m}$. Let $y_1, \ldots, y_m$ be the samples generated.

Alice then sends to Bob the second message $M_2$ which is the values of $f'(x, y_1), \ldots, f'(x, y_m)$ . Bob then considers $x'$ as obtained from the learning algorithm $L$ (as in Theorem 7) and then outputs $kf'(x', y)$, where $y$ is his actual input. Therefore from Theorem 7, with probability $1 - \epsilon$ over the samples $y_1, \ldots, y_m$,

$$\sum_{y \in \mathcal{Y}} \pi(y) \cdot |f'(x', y) - f'(x, y)| \leq \epsilon/k. \tag{9}$$

Note that, $(f'(x', y) \neq f'(x, y)) \Rightarrow |f'(x', y) - f'(x, y)| \geq 1/k$. Hence for samples $y_1, \ldots, y_m$, for which (9) holds, using Markov's inequality, we have $\Pr_{y \leftarrow \mu_x}[f'(x', y) \neq f'(x, y)] \leq \epsilon$. Therefore, for any fixed $x$, the error of $\mathcal{P}$ is at most $2\epsilon$ and hence also the overall error of $\mathcal{P}$ is at most $2\epsilon$.

From Theorem 11 and Lemma 5, we can verify that the expected length of $M_1$ in $\mathcal{P}$ will be at most $4m \cdot I(X : Y) + l$. The length of $M_2$ is at most $O(m \log k)$, since using a prefix free encoding[1] each $f'(x, y_i)$ can be specified in $O(\log k)$ bits. This completes the proof. □

## 4. A new lower bound on quantum one-way distributional communication complexity

In this section we present our lower bound on the quantum one-way distributional communication complexity of a function $f$, in terms of the one-way rectangle bound of $f$. We begin with a few definitions leading to the definition of the one-way rectangle bound.

**Definition 4** (*Rectangle*). A *one-way rectangle* $R$ is a set $S \times \mathcal{Y}$, where $S \subseteq \mathcal{X}$. For a distribution $\mu$ over $\mathcal{X} \times \mathcal{Y}$, let $\mu_R$ represent the distribution arising from $\mu$ conditioned on the event $R$ and let $\mu(R)$ represent the probability (under $\mu$) of the event $R$.

**Definition 5** (*One-way $\epsilon$-Monochromatic*). Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. We call a distribution $\lambda$ on $\mathcal{X} \times \mathcal{Y}$, *one-way $\epsilon$-monochromatic* for $f$ if there is a function $g : \mathcal{Y} \to \mathcal{Z}$ such that $\Pr_{XY \sim \lambda}[(X, Y, g(Y)) \in f] \geq 1 - \epsilon$.

Note that in the case that $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ is a total boolean function, and $\lambda = \lambda_{\mathcal{X}} \otimes \lambda_{\mathcal{Y}}$ is a product distribution, the requirement for $\lambda$ to be one-way $\epsilon$-monochromatic becomes

$$\mathbf{E}_{Y \sim \lambda_{\mathcal{Y}}}[\max\{\Pr_{X \sim \mu_{\mathcal{X}}}[f(X, Y) = 0], \Pr_{X \sim \mu_{\mathcal{X}}}[f(X, Y) = 1]\}] \geq 1 - \epsilon.$$

**Definition 6** (*Rectangle Bound*). Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. For distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$, the *one-way rectangle bound* is defined as:

$$\mathsf{rec}_{\epsilon}^{1,\mu}(f) \stackrel{\text{def}}{=} \min\left\{\log_2 \frac{1}{\mu(R)} : R \text{ is one-way rectangle and } \mu_R \text{ is one-way } \epsilon\text{-monochromatic}\right\}.$$

The *one-way rectangle bound* for $f$ is defined as:

$$\mathsf{rec}_{\epsilon}^{1}(f) \stackrel{\text{def}}{=} \max_{\mu} \mathsf{rec}_{\epsilon}^{1,\mu}(f).$$

We also define,

$$\mathsf{rec}_{\epsilon}^{1,[]}(f) \stackrel{\text{def}}{=} \max_{\mu:\text{product}} \mathsf{rec}_{\epsilon}^{1,\mu}(f).$$

We restate our precise result here followed by its proof.

**Theorem 13.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a total function and let $\epsilon \in (0, 1/2)$ be a constant. Let $\mu$ be a product distribution on $\mathcal{X} \times \mathcal{Y}$ and let $\mathsf{rec}_{\epsilon}^{1,\mu}(f) > 2(\log(1/\epsilon))$. Then,*

$$\mathsf{Q}_{\epsilon^3/8}^{1,\mu}(f) \geq \frac{1}{2} \cdot (1 - 2\epsilon) \cdot (S(\epsilon/2) - S(\epsilon/4)) \cdot (\lfloor \mathsf{rec}_{\epsilon}^{1,\mu}(f) \rfloor - 1).$$

*If $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \cup \{*\}$ is a partial function then,*

$$\mathsf{Q}_{\epsilon^6/(2\cdot 15^4)}^{1,\mu}(f) \geq \frac{1}{2} \cdot (1 - 2\epsilon) \cdot \frac{\epsilon^2}{300} \cdot (\lfloor \mathsf{rec}_{\epsilon}^{1,\mu}(f) \rfloor - 1).$$

We begin with the following information theoretic fact.

**Lemma 6.** *Let $0 \leq d < c \leq 1/2$. Let $Z$ be a binary random variable with $\min\{\Pr(Z = 0), \Pr(Z = 1)\} \geq c$. Let $M$ be a correlated quantum system. Let $Z'$ be a classical boolean random variable obtained by performing a measurement on $M$ such that, $\Pr(Z \neq Z') \leq d$, then*

$$I(Z : M) \geq I(Z : Z') \geq S(c) - S(d).$$

**Proof.** The first inequality follows from the Holevo bound, Theorem 8. For the second inequality we note that $S(Z) \geq S(c)$ (since the binary entropy function is monotonically increasing in $(0, 1/2]$) and from Fano's inequality, Lemma 1, we have $S(Z|Z') \leq S(d)$. Therefore,

$$I(Z : Z') = S(Z) - S(Z|Z') \geq S(c) - S(d). \quad \square$$

[1] Prefix free encoding is needed to avoid ambiguity of messages and to know when a particular message has terminated.

We are now ready for the proof of Theorem 13.

**Proof of Theorem 13. For total boolean functions:** For simplicity of the explanation, we first present the proof assuming $f$ to be a total boolean function. Let $r \stackrel{\text{def}}{=} \lfloor \mathsf{rec}_\epsilon^{1,\mu}(f) \rfloor$ or $\lfloor \mathsf{rec}_\epsilon^{1,\mu}(f) \rfloor - 1$ so as to make $r$ even. Let $\mathcal{P}$ be the optimal one-way quantum protocol for $f$ with distributional error under $\mu$ at most $\epsilon^3/4$. (Although we have made a stronger assumption regarding the error in the statement of the Theorem, we do not need it here and will only need it later while handling non-boolean functions.) Let $M$ represent the $m \stackrel{\text{def}}{=} \mathsf{Q}_{\epsilon^3/4}^{1,\mu}(f)$ qubit quantum message of Alice in $\mathcal{P}$. Let $XY$ be the random variables corresponding to Alice and Bob's inputs, jointly distributed according to $\mu$. Our intention is to define binary random variables $T_1, \ldots, T_{r/2}$ such that they are determined by $X$ (and hence a specific value for $T_1, \ldots, T_{r/2}$ would correspond to a subset of $\mathcal{X}$) and $\forall i \in \{0, \ldots, \frac{r}{2} - 1\}$,

$$I(M : T_{i+1} | T_1 \ldots T_i) \geq (1 - 2\epsilon) \cdot (S(\epsilon/2) - S(\epsilon/4)).$$

Therefore from Fact 3 and the chain rule of mutual information, Eq. (6), we have,

$$\begin{aligned} m \geq S(M) &\geq I(M : T_1 \ldots T_{r/2}) \\ &= \sum_{i=0}^{r/2-1} I(M : T_{i+1} | T_1 \ldots T_i) \\ &\geq (1 - 2\epsilon) \cdot (S(\epsilon/2) - S(\epsilon/4)) \cdot \frac{r}{2}. \end{aligned}$$

This completes our proof.

We define $T_1, \ldots, T_{r/2}$ in an inductive fashion. The following construction of $T_{i+1}$ also works for $i = 0$; we will give more details afterwards.

For $i \in \{0, \ldots, \frac{r}{2} - 1\}$, assume that we have defined $T_1, \ldots, T_i$ and we intend to define $T_{i+1}$. Let $\mathsf{GOOD}_1$ be the set of "heavy bands", *i.e.* those strings $t \in \{0, 1\}^i$ such that $\Pr(T_1, \ldots, T_i = t) > 2^{-r}$. Then,

$$\Pr(T_1, \ldots, T_i \in \mathsf{GOOD}_1) \geq 1 - 2^{-r+i} \geq 1 - 2^{-r/2-1}.$$

Let $\epsilon_t$ be the error of the protocol $\mathcal{P}$ conditioned on $T_1, \ldots, T_i = t$. Note that $\mathbf{E}[\epsilon_t]$ is the same as the overall expected error of $\mathcal{P}$; hence $\mathbf{E}[\epsilon_t] \leq \epsilon^3/4$. Now using Markov's inequality we get a set $\mathsf{GOOD}_2 \in \{0, 1\}^i$ of "small error bands" such that $\Pr(T_1 \ldots T_i \in \mathsf{GOOD}_2) \geq 1 - \epsilon$ and $\forall t \in \mathsf{GOOD}_2,\ \epsilon_t \leq \epsilon^2/4$. Let $\mathsf{GOOD} \stackrel{\text{def}}{=} \mathsf{GOOD}_1 \cap \mathsf{GOOD}_2$ contains those heavy bands with small error. Therefore (since $r/2 > \log(1/\epsilon)$, from the hypothesis of the theorem),

$$\Pr(T_1 \ldots T_i \in \mathsf{GOOD}) \geq 1 - 2^{-r/2-1} - \epsilon \geq 1 - 2\epsilon. \tag{10}$$

For $t \in \{0, 1\}^i$ and $y \in \mathcal{Y}$, let

$$\delta_{t,y} \stackrel{\text{def}}{=} \min\{\Pr[f(X, y) = 0 | (T_1 \ldots T_i = t)],\ \Pr[f(X, y) = 1 | (T_1 \ldots T_i = t)]\}$$

where the probabilities are over the protocol's randomness and $X$, the $\mathcal{X}$-part of the product distribution $\mu$. Also let, $\epsilon_{t,y}$ be the expected error of $\mathcal{P}$ conditioned on $Y = y$ and $T_1 \ldots T_i = t$.

For $t \notin \mathsf{GOOD}$, we define $T_{i+1} | (T_1 \ldots T_i = t) = 0$. Let $t \in \mathsf{GOOD}$ from now on. Our intention is to identify for every $t$ a $y_t \in \mathcal{Y}$, such that $\epsilon_{t,y_t} \leq \epsilon/4$ and $\delta_{t,y_t} \geq \epsilon/2$. We will then let $T_{i+1} | (T_1 \ldots T_i = t)$ to be $f(X, y_t) | (T_1 \ldots T_i = t)$. Lemma 6 will now imply, $I(M : T_{i+1} | (T_1 \ldots T_i = t)) \geq S(\epsilon/2) - S(\epsilon/4)$. Therefore,

$$\begin{aligned} I(M : T_{i+1} | T_1 \ldots T_i) &\geq \sum_{t \in \mathsf{GOOD}} \Pr(T_1 \ldots T_i = t) \cdot I(M : T_{i+1} | (T_1 \ldots T_i = t)) \\ &\geq (1 - 2\epsilon) \cdot (S(\epsilon/2) - S(\epsilon/4)) \quad \text{(using Eq. (10))} \end{aligned}$$

and we would be done.

Now in order to identify a desired $y_t$, we proceed as follows. Since $r \leq \mathsf{rec}_\epsilon^{1,\mu}(f)$; from the definition of rectangle bound and given that $\mu$ is a product distribution we have the following. For all $S \subseteq \mathcal{X}$ with $\mu(S \times \mathcal{Y}) > 2^{-r}$ or in other words with $\Pr[X \in S] > 2^{-r}$,

$$\mathbf{E}_{y \leftarrow Y}\Big[\min\{\Pr[f(X, y) = 0 | X \in S],\ \Pr[f(X, y) = 1 | X \in S]\}\Big] > \epsilon. \tag{11}$$

Note that since $t \in \mathsf{GOOD}$, $\Pr[T_1 \ldots T_i = t] > 2^{-r}$. Recall that conditioning on $t$ implies a subset of $\mathcal{X}$. Hence (11) implies that $\mathbf{E}_{y \leftarrow Y}[\delta_{t,y}] > \epsilon$. Now using Markov's inequality and the fact that, $\forall (t, y),\ \delta_{t,y} \leq 1/2$, we get a set $\mathsf{GOOD}_t \subseteq \mathcal{Y}$ such that $\Pr[Y \in \mathsf{GOOD}_t] \geq \epsilon$ and $\forall y \in \mathsf{GOOD}_t, \delta_{t,y} \geq \epsilon/2$.

Since $t \in \mathsf{GOOD}$, we have $\epsilon_t \leq \epsilon^2/4$. Note that $\epsilon_t = \mathbf{E}_{y \leftarrow Y}[\epsilon_{t,y}]$. Using a Markov argument again we finally get a $y_t \in \mathsf{GOOD}_t$, such that $\epsilon_{t,y_t} \leq \epsilon/4$. Note that since $y_t \in \mathsf{GOOD}_t$, we have $\delta_{t,y_t} \geq \epsilon/2$ and we are done.

Now we finish the total boolean functions part by adding a few remarks for construction of $T_1$. The above process works with minor adjustments which basically delete all appearances of $t$ and $T_1, \ldots, T_i$. Let us start from the definition of $\delta_y = \min\{\Pr[f(X, y) = 0], \Pr[f(X, y) = 1]\}$ and $\epsilon_y =$ the expected error condition on $Y = y$. Using similar argument we can find a particular $y$ *s.t.* $\epsilon_y \leq \epsilon/4$ and $\delta_y \geq \epsilon/2$, then we let $T_1 = f(X, y)$ and we have $I(M : T_{i+1}) \geq S(\epsilon/2) - S(\epsilon/4)$.

**For total non-boolean functions:** Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a total non-boolean function and let $r$ be as before. We follow the same inductive argument as before to define $T_1 \dots T_{r/2}$. For $i \in \{0, \dots, \frac{r}{2} - 1\}$, assume that we have defined $T_1 \dots T_i$. As before we identify a set $\mathsf{GOOD} \subseteq \{0, 1\}^i$ with $\Pr[T_1 \dots T_i \in \mathsf{GOOD}] \geq 1 - 2\epsilon$, such that $\forall t \in \mathsf{GOOD}$, $\Pr[T_1 \dots T_i = t] > 2^{-r}$ and $\epsilon_t \leq \epsilon^2/8$. Since $r \leq \mathsf{rec}_\epsilon^{1,\mu}(f)$, from the definition of rectangle bound and the fact that $\mu$ is product, we have , $\forall S \subseteq \mathcal{X}$ with $\mu(S \times \mathcal{Y}) > 2^{-r}$,

$$\mathbf{E}_{y \leftarrow Y}\Big[\max_{z \in \mathcal{Z}} \{\Pr[f(X, y) = z | X \in S]\}\Big] < 1 - \epsilon. \tag{12}$$

For $t \in \{0, 1\}^i$ and $y \in \mathcal{Y}$, let $\epsilon_{t,y}$ be as before and let,

$$\delta_{t,y} \stackrel{\text{def}}{=} \max_{z \in \mathcal{Z}} \{\Pr[f(X, y) = z | (T_1 \dots T_i = t)]\}.$$

For $t \notin \mathsf{GOOD}$, let us define $T_{i+1} | (T_1 \dots T_i = t)$ to be 0. Let $t \in \mathsf{GOOD}$ from now on. Note that (12) implies $\mathbf{E}_{y \leftarrow Y}[\delta_{t,y}] < 1 - \epsilon$. Using Markov's inequality we get a set $\mathsf{GOOD}_t \subseteq \mathcal{Y}$ with $\Pr[Y \in \mathsf{GOOD}_t] \geq \epsilon/2$ and $\forall y \in \mathsf{GOOD}_t$, $\delta_{t,y} \leq 1 - \epsilon/2$. Since $\mathbf{E}_{y \leftarrow Y}[\epsilon_{t,y}] = \epsilon_t \leq \epsilon^2/8$, again using a Markov argument we get a $y_t \in \mathsf{GOOD}_t$, such that $\epsilon_{t,y_t} \leq \epsilon/4$. Since $\delta_{t,y_t} \leq 1 - \epsilon/2$ (and $\epsilon \in (0, 1/2)$), observe that there would exist a set $S_{t,y_t} \subseteq \mathcal{Z}$ such that,

$$\min\{\Pr[f(X, y_t) \in S_{t,y_t} | (T_1 \dots T_i = t)], \Pr[f(X, y_t) \in \mathcal{Z} - S_{t,y_t} | (T_1 \dots T_i = t)]\} \geq \epsilon/2.$$

Let us now define $T_{i+1} | (T_1 \dots T_i = t)$ to be 1 if and only if $f(X, y_t) \in S_{t,y_t} | (T_1 \dots T_i = t)$ and 0 otherwise. Note that since $\epsilon_{t,y_t} \leq \epsilon/4$, conditioned on $T_1 \dots T_i = t$, there exists a measurement on $M$, that can predict the value of $T_{i+1}$ with success probability at least $1 - \epsilon/4$. The rest of the proof follows as before.

**For partial non-boolean functions:** Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} \cup \{*\}$ be a partial function and let $r$ be as before. Let $i \in \{0, \dots, \frac{r}{2} - 1\}$. We follow a similar inductive argument as in the case of total non-boolean functions, except for the definition of $T_{i+1} | (T_1 \dots T_i = t)$. As before we identify a set $\mathsf{GOOD} \subseteq \{0, 1\}^i$ with $\Pr[T_1 \dots T_i \in \mathsf{GOOD}] \geq 1 - 2\epsilon$, such that $\forall t \in \mathsf{GOOD}$, $\Pr[T_1 \dots T_i = t] > 2^{-r}$ and $\epsilon_t \leq \epsilon^5/(2 \cdot 15^4)$. Since $r \leq \mathsf{rec}_\epsilon^{1,\mu}(f)$, from the definition of rectangle bound and the fact that $\mu$ is product, we have the following. For all $S \subseteq \mathcal{X}$ with $\mu(S \times \mathcal{Y}) > 2^{-r}$,

$$\mathbf{E}_{y \leftarrow Y}\Big[\max_{z \in \mathcal{Z}} \{\Pr[f(X, y) = (z \text{ or } *) | X \in S]\}\Big] < 1 - \epsilon. \tag{13}$$

For $t \in \{0, 1\}^i$ and $y \in \mathcal{Y}$, let $\epsilon_{t,y}$ be as before and let

$$\delta_{t,y} \stackrel{\text{def}}{=} \max_{z \in \mathcal{Z}} \{\Pr[f(X, y) = (z \text{ or } *) | (T_1 \dots T_i = t)]\}.$$

Recall that conditioning on $t$ implies a subset of $\mathcal{X}$. For $t \notin \mathsf{GOOD}$, let us define $T_{i+1} | (T_1 \dots T_i = t)$ to be 0. Let us assume $t \in \mathsf{GOOD}$ from now on. Let $\mathsf{GOOD}_t \subseteq \mathcal{Y}$ be such that $\forall y \in \mathsf{GOOD}_t$, $\delta_{t,y} \leq 1 - \epsilon/2$. Using Markov arguments as before we get a $y_t \in \mathsf{GOOD}_t$, such that $\delta_{t,y_t} \leq 1 - \epsilon/2$ and $\epsilon_{t,y_t} \leq (\epsilon/15)^4 \stackrel{\text{def}}{=} \epsilon'$. Since $\delta_{t,y_t} \leq 1 - \epsilon/2$ it implies $\Pr[f(X, y_t) = *] \leq 1 - \epsilon/2$. Observe now that can we get a set $S_{t,y_t} \subseteq \mathcal{Z}$ such that,

$$\min\{\Pr[f(X, y_t) \in S_{t,y_t} | (T_1 \dots T_i = t)], \Pr[f(X, y_t) \in \mathcal{Z} - S_{t,y_t} | (T_1 \dots T_i = t)]\} \geq \epsilon/6. \tag{14}$$

Let $O$ be the output of $\mathsf{Bob}$ when $Y = y_t$. All along the arguments below we condition on $T_1 \dots T_i = t$. Note that since $\mathsf{Bob}$ outputs some $z \in \mathcal{Z}$ even if $f(x, y) = *$, let us assume without loss of generality that $q \stackrel{\text{def}}{=} \Pr[O \in S_{t,y_t}] \geq 1/2$ (otherwise similar arguments would hold by switching the roles of $S_{t,y_t}$ and $\mathcal{Z} - S_{t,y_t}$). Let us define $T_{i+1}$ to be 1 if $(f(X, y_t) \in S_{t,y_t} \cup \{*\})$ and 0 otherwise. Note that Eq. (14) implies $\Pr[T_{i+1} = 1] \leq 1 - \epsilon/6$. Now,

$$\begin{aligned} q &= \Pr[O \in S_{t,y_t} | (T_{i+1} = 1)] \cdot \Pr[T_{i+1} = 1] + \Pr[O \in S_{t,y_t} \text{ and } T_{i+1} = 0] \\ &\leq \Pr[O \in S_{t,y_t} | (T_{i+1} = 1)] \cdot \Pr[T_{i+1} = 1] + \epsilon' \\ &\leq \Pr[O \in S_{t,y_t} | (T_{i+1} = 1)] \cdot (1 - \epsilon/6) + \epsilon'. \end{aligned}$$

This implies,

$$\begin{aligned} \Pr[O \in S_{t,y_t} | (T_{i+1} = 1)] &\geq \frac{q - \epsilon'}{1 - \epsilon/6} \\ &\geq (q - \epsilon')(1 + \epsilon/6) \\ &= q + q\epsilon/6 - \epsilon'(1 + \epsilon/6) \\ &\geq q + \epsilon/12 - \epsilon(1 + 1/12)/(2^3 \cdot 15^4) \qquad (\text{since } q \geq 1/2 \text{ and } \epsilon \leq 1/2) \\ &\geq q + 0.08\epsilon. \end{aligned} \tag{15}$$

Let us define $O' = 1$ iff $O \in S_{t,y_t}$ and $O' = 0$ otherwise. Then,

$$\begin{aligned} I(M : T_{i+1}) &\geq I(O' : T_{i+1}) \\ &= S(O') - \Pr[T_{i+1} = 1] \cdot S(O'|(T_{i+1} = 1)) - \Pr[T_{i+1} = 0] \cdot S(O'|(T_{i+1} = 0)) \\ &\geq S(q) - S(q + 0.08\epsilon) - S(\epsilon') \\ &\geq 1 - S(0.5 + 0.08\epsilon) - S(\epsilon') \\ &\geq 1 - (1 - 2(0.08\epsilon)^2) - 2(\epsilon/15)^2 \\ &\geq \epsilon^2/300. \end{aligned}$$

The second inequality above follows using Eq. (15); the fact that $\Pr[O' = 1|(T_{i+1} = 0)] \leq \epsilon_{t,y_t} \leq \epsilon' \leq 0.5$ (since $(O' = 1|(T_{i+1} = 0))$ is an error event); and the fact that the function $S(p)$ is monotonically decreasing in $[\frac{1}{2}, 1]$ and monotonically increasing in $[0, \frac{1}{2}]$. The third inequality again follows since the function $S(p)$ is concave and monotonically decreasing in $[\frac{1}{2}, 1]$. The fourth inequality follows from Fact 1. The rest of the proof follows as before. □

## 5. Application: Security of boolean extractors against quantum adversaries

In this section we present a consequence of our lower bound result Theorem 13 to prove security of extractors against quantum adversaries. In this section we are only concerned with boolean extractors. We begin with following definitions.

**Definition 7** (*Min-Entropy*). Let $P$ be a distribution on $[N]$. The *min-entropy* of $P$ denoted $S_\infty(P)$ is defined to be $-\log \max_{i \in [N]} P(i)$.

**Definition 8** (*Strong Extractor*). Let $\epsilon \in (0, 1/2)$. Let $Y$ be uniformly distributed on $\mathcal{Y}$. A *strong* $(k, \epsilon)$*-extractor* is a function $h : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ such that for any random variable $X$ distributed on $\mathcal{X}$, independent of $Y$ and with $S_\infty(X) \geq k$ we have,

$$\|h(X, Y)Y - U \otimes Y\|_1 < 2\epsilon,$$

where $U$ is the uniform distribution on $\{0, 1\}$.

In other words, even given $Y$ (and not $X$); $h(X, Y)$ is still close (in $\ell_1$ distance) to being a uniform bit.

Let $X, Y, h$ be as in the definition above. Let us consider a random variable $M$, taking values in some set $\mathcal{M}$, correlated with $X$ and independent of $Y$. Let us now limit the correlation that $M$ has with $X$, in the sense that $\forall m \in \mathcal{M}$, $S_\infty(X|M = m) \geq k$. Since $h$ is a strong $(k, \epsilon)$-extractor, it is easy to verify that in such a case,

$$\begin{aligned} &\forall m \in \mathcal{M}, \quad \|h(X, Y)Y|(M = m) - U \otimes Y|(M = m)\|_1 < 2\epsilon \\ &\Rightarrow \quad \|h(X, Y)YM - U \otimes YM\|_1 < 2\epsilon. \end{aligned}$$

In other words, still close (in $\ell_1$ distance) to being a uniform bit.

Now let us ask what happens if the system $M$ is a quantum system. In that case, is it still true that given $M$ and $Y$, $h(X, Y)$ is close to being a uniform bit? This question has been increasingly studied in recent times specially for its applications for example in privacy amplification in Quantum key distribution protocols and in the Quantum bounded storage models [10,13,14].

However when $M$ is a quantum system, the min-entropy of $X$, conditioned on $M$, is not easily captured since conditioning on a quantum system needs to be carefully defined. An alternate way to capture the correlation between $X$ and $M$ is via the *guessing probability*. Let us consider the following definition.

**Definition 9** (*Guessing-Entropy*). Let $X$ be a classical random variable taking values in $\mathcal{X}$. Let $M$ be a correlated quantum system with the joint classical-quantum state being $\rho_{XM} = \sum_x \Pr[X = x]|x\rangle\langle x| \otimes \rho_x$. Then the *guessing-entropy* of $X$ given $M$, denoted $S_g(X \leftarrow M)$ is defined to be:

$$S_g(X \leftarrow M) \stackrel{\text{def}}{=} -\log \max_{\mathcal{E}} \sum_x \Pr(X = x)\mathrm{Tr}(E_x \rho_x)$$

where the maximum is taken over all POVMs $\mathcal{E} \stackrel{\text{def}}{=} \{E_x : x \in \mathcal{X}\}$. (Please refer to [17] for a definition of POVMs).

The guessing-entropy turns out to be a useful notion in the quantum contexts. Let $h, X, Y, M$ be as before, where $M$ is a quantum system. König and Terhal [14] have in a high level shown that if the guessing entropy $S_g(X \leftarrow M)$, is at least $k$, then given $M$ and $Y$ (and not $X$), $h(X, Y)$ is still close to a uniform bit. We state their precise result here.

**Theorem 14.** *Let $\epsilon \in (0, 1/2)$. Let $h : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a strong $(k, \epsilon)$-extractor. Let $U$ be the uniform distribution on $\{0, 1\}$. Let $YXM$ be a classical-quantum system with $YX$ being classical and $M$ quantum. Let $Y$ be uniformly distributed and independent of $XM$ and,*

$$S_g(X \leftarrow M) > k + \log 1/\epsilon.$$

*Then,*

$$\|h(X, Y)YM - U \otimes YM\|_1 < 6\sqrt{\epsilon}.$$

We show a similar result as follows.

**Theorem 15.** *Let $\epsilon \in (0, 1/2)$. Let $h : \{0, 1\}^n \times \{0, 1\}^m \rightarrow \{0, 1\}$ be a strong $(k, \epsilon)$-extractor. Let $U$ be the uniform distribution on $\{0, 1\}$. Let $YXM$ be a classical-quantum system with $YX$ being classical and $M$ quantum. Let $X$ be uniformly distributed on $\{0, 1\}^n$. Let $Y$ be uniformly distributed on $\{0, 1\}^m$ and independent of $XM$ and,*

$$I(X : M) < b(\epsilon) \cdot (n - k). \tag{16}$$

*Then,*

$$\|h(X, Y)YM - U \otimes YM\|_1 < 1 - a(\epsilon) \tag{17}$$

*where $a(\epsilon) \stackrel{\text{def}}{=} \frac{1}{4} \cdot (\frac{1}{2} - \epsilon)^3$ and $b(\epsilon) \stackrel{\text{def}}{=} \epsilon \cdot (S(\frac{1}{4} - \frac{\epsilon}{2}) - S(\frac{1}{8} - \frac{\epsilon}{4}))$.*

Before proving Theorem 15, we will make a few points comparing it with Theorem 14.

1. Let's observe that if $M$ is a classical system, then

$$\begin{aligned} S_g(X \leftarrow M) &= -\log \mathbf{E}_{m \leftarrow M}[2^{-S_\infty(X|M=m)}] \\ &\leq \mathbf{E}_{m \leftarrow M}[S_\infty(X|M = m) \cdot \log_e 2] \\ &\leq \mathbf{E}_{m \leftarrow M}[S_\infty(X|M = m)] \\ &\leq S(X|M). \end{aligned}$$

   The first inequality follows from the convexity of the exponential function. The last inequality follows easily from definitions. This implies,

$$I(X : M) = S(X) - S(X|M) \leq S(X) - S_g(X \leftarrow M). \tag{18}$$

   So if $M$ is classical, then the implication of Theorem 15 appears stronger than the implication in Theorem 14 (although being weak in terms of the dependence on $\epsilon$.) We cannot show the inequality (18) when $M$ is a quantum system but conjecture it to be true. If the conjecture is true, Theorem 15 would have stronger implication than Theorem 14 in the quantum case as well.
2. The proof of Theorem 14 in [14] crucially uses some properties of the so called *pretty good measurements* (PGMs). Our result follows here without using PGMs and via completely different arguments.
3. Often in applications concerning the Quantum bounded storage model, an upper bound on the number of qubits of $M$ is available. This implies the same upper bound on $I(X : M)$. If this bound is sufficiently small such that it suffices the assumption of Theorem 15, then $h$ could be used to extract a private bit successfully, in the presence of a quantum adversary.
4. Since Theorem 15 concerns mutual information between the systems $X$ and $M$, $X$ is required to be uniformly distributed in the statement of it. However since Theorem 14 concerns the guessing entropy of $X$ given $M$, the requirement that $X$ needs to be uniformly distributed does not figure in and just its guessing entropy given $M$ is required to be large.

Let us return to the proof of Theorem 15. We begin with the following key observation. It essentially states that a boolean function which can extract a bit from sources of low min-entropy has high one-way rectangle bound under the uniform distribution.

**Lemma 7.** *Let $\epsilon \in (0, 1/2)$. Let $h : \{0, 1\}^n \times \{0, 1\}^m \rightarrow \{0, 1\}$ be a strong $(k, \epsilon)$-extractor. Let $\mu \stackrel{\text{def}}{=} U_n \otimes U_m$, where $U_n, U_m$ are uniform distributions on $\{0, 1\}^n$ and $\{0, 1\}^m$ respectively. Then*

$$\mathsf{rec}^{1,\mu}_{1/2-\epsilon}(h) > n - k.$$

**Proof.** Let $R \stackrel{\text{def}}{=} S \times \{0, 1\}^m$ be any one-way rectangle where $S \subseteq \{0, 1\}^n$ with $\mu(R) \geq 2^{-n+k}$ which essentially means that $|S| \geq 2^k$. Let $X$ be uniformly distributed on $S$. This implies that $S_\infty(X) \geq k$. Let $Y$ be uniformly distributed on $\{0, 1\}^m$ and independent of $X$. Since $h$ is a strong extractor, from Definition 8 we have (where $U$ is the uniform distribution on $\{0, 1\}$):

$$\begin{aligned} &\|h(X, Y)Y - U \otimes Y\|_1 < 2\epsilon \\ &\Leftrightarrow \mathbf{E}_{y \leftarrow Y}[\|h(X, y) - U\|_1] < 2\epsilon. \end{aligned} \tag{19}$$

Let $g : \{0, 1\}^m \rightarrow \{0, 1\}$ be any function. Then,

$$\begin{aligned} \Pr[h(X, Y) = g(Y)] &= \frac{1}{2} \cdot \|h(X, Y) - g(Y)\|_1 \\ &= \frac{1}{2} \cdot \mathbf{E}_{y \leftarrow Y}[\|h(X, y) - g(y)\|_1] \\ &\leq \frac{1}{2} \cdot \mathbf{E}_{y \leftarrow Y}[\|h(X, y) - U\|_1 + \|U - g(y)\|_1] \\ &= \frac{1}{2} + \frac{1}{2} \cdot \mathbf{E}_{y \leftarrow Y}[\|h(X, y) - U\|_1] \quad (\text{since } \forall y, \|U - g(y)\|_1 = 1) \\ &< \frac{1}{2} + \epsilon \quad (\text{from Eq. (19)}). \end{aligned}$$

Now from Definition 5, above implies that $\mu_R$ (uniform distribution on $R$) is not $1/2-\epsilon$ monochromatic. Hence from the definition of the rectangle bound (Definition 6) we have $\mathsf{rec}^{1,\mu}_{1/2-\epsilon}(h) > n-k$. □

We will also need the following information theoretic fact.

**Lemma 8.** *Let $RQ$ be a joint classical-quantum system where $R$ is a classical boolean random variable. Let $U$ be the uniform distribution on $\{0,1\}$. There is a measurement that can be done on $Q$ to guess value of $R$ with probability*

$$\frac{1}{2}+\frac{1}{2}\cdot\|RQ-U\otimes Q\|_1.$$

**Proof.** For $a\in\{0,1\}$, let the quantum state of $Q$ when $R=a$ be $\rho_a$. Let us note that

$$\|RQ-U\otimes Q\|_1=\|\Pr[R=0]\rho_0-\Pr[R=1]\rho_1\|_1.$$

Now Helstrom's Theorem (Theorem 9) immediately helps us conclude the desired. □

We are now ready for the proof of Theorem 15.

**Proof of Theorem 15.** We prove our result in the contrapositive manner. Let,

$$\|h(X,Y)MY-U\otimes MY\|_1>1-a(\epsilon).$$

Note that this is equivalent to:

$$\mathbf{E}_{y\leftarrow Y}[\|h(X,y)M-U\otimes M\|_1]>1-a(\epsilon). \tag{20}$$

Let us consider a one-way communication protocol $\mathcal{P}$ for $h$ where the inputs $X$ and $Y$ of Alice and Bob respectively are drawn independently from the uniform distributions on $\{0,1\}^n$ and $\{0,1\}^m$ respectively. Let $\mu$ be the distribution of $XY$. Now let $M$ be sent as the message of Alice in $\mathcal{P}$. Note that Lemma 8 implies that for a given input $y$, Bob will be able to output the correct answer with probability $\frac{1}{2}+\frac{1}{2}\cdot\|h(X,y)M-U\otimes M\|_1$. Hence we get that the distributional error of $\mathcal{P}$ will be at most

$$\begin{aligned}\mathbf{E}_{y\rightarrow Y}\left[1-\frac{1}{2}-\frac{1}{2}\cdot\|h(X,y)M-U\otimes M\|_1\right]&=\frac{1}{2}-\frac{1}{2}\cdot\mathbf{E}_{y\rightarrow Y}\left[\frac{1}{2}\cdot\|h(X,y)M-U\otimes M\|_1\right]\\&<\frac{1}{2}-\frac{1}{2}(1-a(\epsilon))\quad\text{(from Eq. (20))}\\&=\frac{a(\epsilon)}{2}=\frac{1}{8}\cdot\left(\frac{1}{2}-\epsilon\right)^3.\end{aligned}$$

Let $\epsilon'\stackrel{\text{def}}{=}1/2-\epsilon$. Therefore $\mathcal{P}$ has distributional error $<\epsilon'^3/8$. Arguing as in the proof of Theorem 13 we get that,

$$\begin{aligned}I(X:M)&\geq\frac{1}{2}\cdot(1-2\epsilon')(S(\epsilon'/2)-S(\epsilon'/4))\cdot\mathsf{rec}^{1,\mu}_{\epsilon'}(h)\\&=\epsilon\cdot\left(S\left(\frac{1}{4}-\frac{\epsilon}{2}\right)-S\left(\frac{1}{8}-\frac{\epsilon}{4}\right)\right)\cdot\mathsf{rec}^{1,\mu}_{1/2-\epsilon}(h)\\&=b(\epsilon)\cdot\mathsf{rec}^{1,\mu}_{1/2-\epsilon}(h)\\&>b(\epsilon)\cdot(n-k).\end{aligned}$$

The last inequality follows from Lemma 7 since $h$ is a strong $(k,\epsilon)$-extractor. □

## 6. Conclusion

The main goal of this work is to show bounds for general total functions instead of specific ones, with the motivation of approaching the conjecture of $\mathsf{R}^{1,\mathrm{pub}}(f)=O(\mathsf{Q}^{1,\mathrm{pub}}(f))$ mentioned in Section 1. In the wake of our quantum lower bound result, it is natural to ask whether in the two-way model also, there is a similar relationship between quantum distributional communication complexity of a function $f$, under product distributions, and the corresponding rectangle bound.

Further explorations along this approach are expected. For example, concerning the classical upper bound, a natural question to ask is whether the bound could be tightened, especially in terms of its dependence on the mutual information $I(X:Y)$ between the inputs, under a given non-product distribution. Is it actually true that $\mathsf{D}^{1,\mu}_{\epsilon}(f)=O(I(X:Y)+\mathsf{VC}(f))$? Also, can we say more on the quantum lower bound result for non-product distributions?

## Acknowledgment

We thank Ashwin Nayak for many helpful discussions.

Work was done while the first author was at Institute for Quantum Computing, University of Waterloo where it was supported in part by grant from ARO/NSA USA. Most work was done when the second author was in California Institute of Technology (Caltech), supported by the National Science Foundation under grant PHY-0456720 and the Army Research Office under grant W911NF-05-1-0294 through Institute for Quantum Information at Caltech.

## Appendix

Let $n \geq 1$ be a sufficiently large integer. Let the Noisy Partial Matching ($\mathsf{NPM_n}$) function be as follows.

**Input:**
Alice: A string $x \in \{0, 1\}^n$.
Bob: A string $w \in \{0, 1\}^n$ and a Matching $M$ on $[2n]$ comprising of $n$ disjoint edges.

**Output:**
For a matching $M$ and a string $x$, let $Mx$ represent the $n$ bit string corresponding to the $n$ edges of $M$ obtained as follows. For an edge $e \stackrel{\text{def}}{=} (i, j)$ in $M$ the bit included in $Mx$ is $x_i \oplus x_j$, where $x_i, x_j$ represent the $i, j$-th bit of $x$.

Output bit $b \in \{0, 1\}$ if and only if the Hamming distance between strings $(Mx) \oplus b^n$ and $w$ is at most n/3. If there is no such bit $b$ then output 0.

Now let the non-product distribution $\mu$ on inputs of Alice and Bob be as follows. Let Alice be given $x$ drawn uniformly from $\{0, 1\}^n$. Let Bob be given matching $M$ drawn uniformly from the set of all matchings on $[2n]$. With probability $1/2$, Bob is given $w$ uniformly from the set of all strings with Hamming distance at most $n/3$ from $Mx$ and with probability $1/2$, he is given $w$ uniformly from the set of all strings with Hamming distance at most $n/3$ from $(Mx) \oplus 1^n$. Note that in $\mu$ there is correlation between the inputs of Alice and Bob and hence $\mu$ is non-product. Now we have the following.

**Theorem 16** *([5], Implicit). Let $n \geq 1$ be a sufficiently large integer and let $\epsilon \in (0, 1/2)$. Let $\mathsf{NPM}_n$ and $\mu$ be as described above. Then, $\mathsf{rec}_\epsilon^{1,\mu}(\mathsf{NPM}_n) = \Omega(\sqrt{n})$ whereas $\mathsf{Q}_\epsilon^{1,\mu}(\mathsf{NPM}_n) = O(\log n)$.*

## References


[1] S. Aaronson, The learnability of quantum states, Proceedings of the Royal Society A463 (2088) (2007).
[2] A. Blumer, A. Ehrenfeucht, D. Haussler, M.K. Warmuth, Learnability and the Vapnik–Chervonenkis dimension, Journal of the Association for Computing Machinery 36 (4) (1989) 929–965.
[3] P.L. Bartlett, P.M. Long, R.C. Williamson, Fat-shattering and the learnability of real-valued functions, Journal of Computer and System Sciences 52 (3) (1996) 434–452.
[4] T.M. Cover, J.A. Thomas, Elements of Information Theory, in: Wiley Series in Telecommunications, John Wiley & Sons, New York, NY, USA, 1991.
[5] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, R. de Wolf, Exponential separations for one-way quantum communication complexity, with applications to cryptography, in: Proceedings of The 39th Annual ACM Symposium on Theory of Computing, STOC, 2007, pp. 516–525.
[6] C.W. Helstrom, Quantum Detection and Estimation Theory, Academic Press Inc., New York, NY, USA, 1976.
[7] P. Harsha, R. Jain, D. McAllester, J. Radhakrishnan, The communication complexity of correlation, in: Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity, CCC, 2007.
[8] A.S. Holevo, Some estimates of the information transmitted by quantum communication channels, Problemy Peredachi Informatsii 9 (1973) 311. English translation in Problems of InformationTransmission 9 (1973) 177–183.
[9] R. Jain, H. Klauck, A. Nayak, Direct product theorems for classical communication complexity via subdistribution bounds, in: Proceedings of the 40th ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, SESSION: 13A, ISBN: 978-1-60558-047-0, 2008, pp. 599–608.
[10] R. König, U. Maurer, R. Renner, On the power of quantum memory, IEEE Transactions on Information Theory 51 (2391) (2005).
[11] E. Kushilevitz, N. Nisan, Communication Complexity, Cambridge University Press, Cambridge, UK, 1997.
[12] I. Kremer, N. Nisan, D. Ron, On randomized one-round communication complexity, in: Proceedings of The 27th ACM Symposium on Theory of Computing, STOC, 1995, pp. 596–605.
[13] R. König, R. Renner, Universally composable privacy amplification against quantum adversaries, in: Second Theory of Cryptography Conference, TCC, 2005.
[14] R.T. König, B.M. Terhal, The bounded-storage model in the presence of a quantum adversary, IEEE Transactions on Information Theory 54 (2) (2008) 749–762.
[15] S.M. Muthukrishnan, Data streams: Algorithms and applications, Foundations and Trends in Theoretical Computer Science 1 (2) (2005).
[16] A. Nayak, Optimal lower bounds for quantum automata and random access codes, in: Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science, FOCS, 1999, pp. 124–133.
[17] M.A. Nielsen, I.L. Chuang, Quantum Computation and Quantum Information, Cambridge University Press, Cambridge, UK, 2000.
[18] I. Newman, Private vs. common random bits in communication complexity, Information Processing Letters 39 (2) (1991) 67–71.
[19] A. Razborov, On the distributional complexity of disjointness, Theoretical Computer Science 106 (1992) 385–390.
[20] N. Sauer, On the density of families of sets, Journal of Combinatorial Theory 13 (1972) 145–147.
[21] A.C.-C. Yao, Probabilistic computations: Toward a unified measure of complexity, in: Proceedings of the 18th IEEE Conference on Foundations of Computer Science, 1977, pp. 222–227.
[22] A.C.-C. Yao, Some complexity questions related to distributive computing, in: Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing, STOC, 1979, pp. 209–213.
[23] A.C.-C. Yao, Quantum circuit complexity, in: Proceedings of the 34th Annual Symposium on Foundations of Computer Science, FOCS, 1993, pp. 352–361.