

## BETTER PROTOCOL FOR XOR GAME USING COMMUNICATION PROTOCOL AND NONLOCAL BOXES

RYUHEI MORI

*School of Computing, Tokyo Institute of Technology, Ookayama  
Meguro-ku, Tokyo, 152-8552, Japan*

Received June 4, 2017

Revised September 27, 2017

Buhrman showed that an efficient communication protocol implies a reliable XOR game protocol. This idea rederives Linial and Shraibman’s lower bound of randomized and quantum communication complexities, which was derived by using factorization norms, with worse constant factor in much more intuitive way. In this work, we improve and generalize Buhrman’s idea, and obtain a class of lower bounds for randomized communication complexity including an exact Linial and Shraibman’s lower bound as a special case. In the proof, we explicitly construct a protocol for XOR game from a randomized communication protocol by using a concept of nonlocal boxes and Pawłowski et al.’s elegant protocol, which was used for showing the violation of information causality in superquantum theories.

*Keywords:* Communication complexity, XOR game, nonlocality, CHSH inequality, Fourier analysis

*Communicated by:* R Cleve & R de Wolf

### 1 Introduction

Communication complexity is one of the central tool in theoretical computer science. In this work, we investigate an extremely simple technique for lower bounding communication complexity and XOR-amortized communication complexity. Let  $C_\rho(f, \mu)$  be a deterministic communication complexity for computing  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  with an error probability at most  $(1 - \rho)/2$  on an input distribution  $\mu$ . Let  $\beta(f, \mu)$  be the largest bias of an XOR game for  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  on an input distribution  $\mu$ , i.e., the largest winning probability of the XOR game is  $(1 + \beta(f, \mu))/2$ . Buhrman gave a clear argument on a relationship of  $C_\rho(f, \mu)$  and  $\beta(f, \mu)$  [1, 2].

**Fact 1** (Buhrman’s argument a.k.a. discrepancy bound). For any  $\rho \in [0, 1]$ ,

$$\beta(f, \mu) \geq \rho 2^{-C_\rho(f, \mu)}.$$

*Proof.* We will construct a protocol for the XOR game of  $f$  on input distribution  $\mu$  with the bias  $\rho 2^{-C_\rho(f, \mu)}$  using the communication protocol  $P$  corresponding to  $C_\rho(f, \mu)$ . Alice and Bob use  $C_\rho(f, \mu)$  shared random bits, and regard them as a transcript of the communication protocol  $P$ . Alice and Bob check the consistency at each side, and output a uniform random bit if they are inconsistent. If the random bits are consistent at each side, Alice and Bob output the output of the communication protocol  $P$  and 0, respectively.

If the shared random bits are inconsistent transcript at least one side, the XOR of their output is a uniform random bit. If the shared random bits are consistent transcript at the both sides, the XOR of their output is the output of the communication protocol  $P$ , which is  $f(x, y)$  with probability at least  $(1 + \rho)/2$ . Hence, the winning probability of this protocol for the XOR game is at least  $2^{-C_\rho(f, \mu)}(1 + \rho)/2 + (1 - 2^{-C_\rho(f, \mu)})(1/2) = (1 + \rho 2^{-C_\rho(f, \mu)})/2$ .  $\square$

Buhrman’s argument gives a lower bound  $C_\rho(f, \mu) \geq \log \frac{\rho}{\beta(f, \mu)}$  of the communication complexity, which is equivalent to the well-known discrepancy bound up to an additive constant. Buhrman’s argument gave an operational meaning to the discrepancy bound. Note that a similar argument was independently shown in [3] for lower bounding communication complexity for simulating no-signaling distributions. Furthermore, the above argument with small modification gives Linial and Shraibman’s lower bound, which was obtained by using factorization norms, with a worse constant factor [4, 1]. This argument can be straightforwardly generalized to a quantum setting [1, 2]. Let  $C_\rho^*(f, \mu)$  be a quantum communication complexity, which is the number of bits transmitted (not qubits), for computing  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  using shared quantum states with an error probability at most  $(1 - \rho)/2$  on an input distribution  $\mu$ . Let  $\beta^*(f, \mu)$  be the largest bias of an XOR game using shared quantum states for  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  on an input distribution  $\mu$ . Then, a straightforward generalization of Fact 1 shows  $C_\rho^*(f, \mu) \geq \log \frac{\rho}{\beta^*(f, \mu)}$ . In the following, we demonstrate how to show  $\max_\mu C_\rho^*(\text{IP}_n, \mu) = \Theta(n)$  for any constant  $\rho \in (0, 1]$  where  $\text{IP}_n(x, y) := \langle x, y \rangle := \bigoplus_{i=1}^n x_i \wedge y_i$ . Let  $f^{\oplus \ell}(x_1, \dots, x_{\ell n}) := f(x_1, \dots, x_n) \oplus \dots \oplus f(x_{(\ell-1)n+1}, \dots, x_{\ell n})$ , and  $\mu^{\otimes \ell}(x_1, \dots, x_{\ell n}) := \mu(x_1, \dots, x_n) \times \dots \times \mu(x_{(\ell-1)n+1}, \dots, x_{\ell n})$ . Cleve et al. showed that the XOR game in the quantum physics satisfies the perfect parallel repetition theorem, i.e.,  $\beta^*(f^{\oplus \ell}, \mu^{\otimes \ell}) = \beta^*(f, \mu)^\ell$  [5]. From the perfect parallel repetition theorem, we obtain

$$\min_\mu \beta^*(\text{IP}_n, \mu) \leq \min_\nu \beta^*(\text{IP}_n, \nu^{\otimes n}) = \min_\nu \beta^*(\text{AND}, \nu)^n = 2^{-\frac{n}{2}}.$$

The last equality is obtained by the Tsirelson bound [6] (In fact, the above inequality is equality. It is easy to see that the worst input distribution for  $\text{IP}_n$  is the uniform distribution). Hence, we obtain

$$\max_\mu C_\rho^*(\text{IP}_n, \mu) \geq \log \frac{\rho}{\min_\mu \beta^*(\text{IP}_n, \mu)} \geq \frac{n}{2} + \log \rho.$$

This bound was obtained by Kremer [7] and Linial and Shraibman [4]. However, the above derivation is extremely simple and intuitive, and only needs Buhrman’s argument, Cleve et al.’s perfect parallel repetition theorem and the Tsirelson bound.

In this work, we improve Buhrman’s argument, Fact 1, for deterministic and randomized communication complexities by using quantum theory. First, we obtain

$$\beta^*(f, \mu) \geq \rho 2^{-\frac{1}{2}C_\rho(f, \mu)}$$

or equivalently

$$C_\rho(f, \mu) \geq 2 \log \frac{\rho}{\beta^*(f, \mu)}.$$

Note that this lower bound is worse than Linial and Shraibman’s lower bound [4], but is easier to evaluate. Since  $\beta^*(f, \mu) \geq \beta(f, \mu)$ , an improvement from the previous lower bound

$\log \frac{\rho}{\beta(f, \mu)}$  is at most a factor 2. Although this improvement is typically not significant, in this work, we obtain non-trivial lower bounds of XOR-amortized communication complexities of the equality function.

**Theorem 1.** For any  $\rho \in (0, 1]$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \lim_{\ell \rightarrow \infty} \max_{\mu} \frac{1}{\ell} C_{\rho^\ell}(\text{EQ}_n^{\oplus \ell}, \mu^{\otimes \ell}) &\geq 2 \log 3 + 2 \log \rho \\ \lim_{n \rightarrow \infty} \lim_{\ell \rightarrow \infty} \max_{\nu} \frac{1}{\ell} C_{\rho^\ell}(\text{EQ}_n^{\oplus \ell}, (\nu^{\otimes n})^{\otimes \ell}) &\geq 2 \log \left( 1 + \frac{2}{\sqrt{3}} \right) + 2 \log \rho \end{aligned}$$

where  $\mu$  and  $\nu$  are probability distributions on  $\{0, 1\}^n \times \{0, 1\}^n$  and  $\{0, 1\} \times \{0, 1\}$ , respectively. Here,  $2 \log 3 \approx 3.1699$  and  $2 \log(1 + 2/\sqrt{3}) \approx 2.2150$ .

Since randomized/distributional communication complexity of the equality function is constant [8], these improvements are meaningful. If the factor 2 is missing, the above two lower bounds for the equality function are smaller than 2 for  $\rho = 1$ . In this sense, this improvement is significant.

By applying an argument for generalized discrepancy theory [2, 1], we can further improve this bound, and obtain a class of lower bounds for randomized communication complexity including an exact Linial and Shraibman’s lower bound as a special case.

**Theorem 2** (Linial and Shraibman’s lower bound [4, 1]). For any  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  and  $\epsilon \in [0, 1/2]$ ,

$$R_\epsilon(f) \geq 2 \log \max_{h, \mu} \frac{(1 - \epsilon) \mathbb{E}_{(x, y) \sim \mu} [f(x, y)h(x, y)] - \epsilon}{\beta^*(h, \mu)} \tag{1}$$

where  $R_\epsilon(f)$  is the randomized communication complexity of  $f$  with an error probability at most  $\epsilon$ , where  $h: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ , and where  $\mu$  is a probability distribution on  $\{0, 1\}^n \times \{0, 1\}^n$ .

While the original proof of Theorem 2 uses factorization norms of matrices, the proof in this paper is based on Buhrman’s argument, which derived Theorem 2 in an intuitive way without the constant factor 2 [1, 2]. For obtaining the constant factor 2, we use a concept of nonlocal boxes and Pawłowski et al.’s idea, which showed the violation of information causality in superquantum theories [9]. If we employ superquantum theory for deriving lower bounds of randomized communication complexity, the constant factor 2 is replaced by larger constant while  $\beta^*(h, \mu)$  is also replaced by larger quantity. Since the evaluation of the generalized version of  $\beta^*(h, \mu)$  is not easy at this point, any concrete lower bound by the generalized lower bound is not obtained in this paper. If the generalized  $\beta^*(h, \mu)$  is not much larger than the original  $\beta^*(h, \mu)$ , the generalized lower bound improves the constant factor of Linial and Shraibman’s lower bound.

## 2 Preliminaries

### 2.1 Nonlocal box

The nonlocal box is an abstract device with two input ports and two output ports. When  $x \in \{0, 1\}$  and  $y \in \{0, 1\}$  are given to a nonlocal box, a nonlocal box randomly outputs  $a \in \{0, 1\}$  and  $b \in \{0, 1\}$ . A nonlocal box is specified by a conditional probability distribution  $p(a, b \mid x, y)$ . Here, a nonlocal box is an abstract device representing an “entangled state”

where  $x$  and  $y$  correspond to a choice of “measurements”, and where  $a$  and  $b$  correspond to “outcomes” of the measurements  $x$  and  $y$ , respectively. Hence, the conditional probability distribution must satisfy the no-signaling condition

$$\begin{aligned} \sum_{b \in \{0,1\}} p(a, b | x, 0) &= \sum_{b \in \{0,1\}} p(a, b | x, 1) \\ \sum_{a \in \{0,1\}} p(a, b | 0, y) &= \sum_{a \in \{0,1\}} p(a, b | 1, y) \end{aligned}$$

since if the no-signaling condition is violated, two distant parties can communicate only by measuring a shared state, which is a communication faster than light, and must be forbidden. Let the CHSH probability be

$$P_{\text{CHSH}} := \frac{1}{4} \sum_{\substack{a \in \{0,1\}, b \in \{0,1\}, \\ x \in \{0,1\}, y \in \{0,1\} \\ a \oplus b = x \wedge y}} P(a, b | x, y).$$

When

$$p(a, b | x, y) = \begin{cases} P_{\text{CHSH}}/2, & \text{if } a \oplus b = x \wedge y \\ (1 - P_{\text{CHSH}})/2, & \text{otherwise} \end{cases}$$

the nonlocal box is said to be isotropic. The CHSH bias  $2P_{\text{CHSH}} - 1$  is denoted by  $\delta$ , i.e.,  $P_{\text{CHSH}} = (1 + \delta)/2$ . Classical physics and quantum physics can simulate isotropic nonlocal boxes with CHSH bias up to  $1/2$  [10] and  $1/\sqrt{2}$  [6], respectively.

## 2.2 Communication complexities

Let  $C_\rho(f, \mu)$  be a deterministic communication complexity for computing  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  with an error probability at most  $(1 - \rho)/2$  on an input distribution  $\mu$ . Let  $C_\rho^*(f, \mu)$  be a quantum communication complexity, which is the number of bits transmitted (not qubits), for computing  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$  using shared quantum states with an error probability at most  $(1 - \rho)/2$  on an input distribution  $\mu$ . Let  $C_{\text{NL}(\delta), \rho}(f, \mu)$  be a communication complexity with isotropic nonlocal boxes with CHSH bias  $\delta \geq 1/2$ . Let  $R_\epsilon(f)$  be a randomized communication complexity with an error probability at most  $\epsilon \in [0, 1/2]$ .

*Remark 2.* In this paper, all  $C_\rho(f, \mu)$  (and its variants) can be replaced by  $C_\rho^\oplus(f, \mu)$  which is the communication complexity for computing  $a$  and  $b$  by Alice and Bob, respectively such that  $a \oplus b = f(x, y)$ . Since the difference between  $C_\rho(f, \mu)$  and  $C_\rho^\oplus(f, \mu)$  is at most two, and is negligible for the amortized case, in this paper, we use  $C_\rho(f, \mu)$  for the simplicity.

## 2.3 XOR game

In a two-player XOR game  $(f, \mu)$ , Alice and Bob are given  $x \in \{0, 1\}^n$  and  $y \in \{0, 1\}^n$  according to the input distribution  $\mu$ , and output  $a \in \{0, 1\}$  and  $b \in \{0, 1\}$ , respectively without communication for computing  $f: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ . Alice and Bob win if and only if  $a \oplus b = f(x, y)$ . Let  $\beta(f, \mu)$  be the largest bias (of the winning probability) of an XOR game for  $f$  on an input distribution  $\mu$ , i.e., the largest winning probability of the XOR game  $(f, \mu)$  is  $(1 + \beta(f, \mu))/2$ . Let  $\beta^*(f, \mu)$  be the largest bias of an XOR game using shared quantum states. Let  $\beta_{\text{NL}(\delta)}(f, \mu)$  be the largest bias of an XOR game using isotropic nonlocal

boxes with CHSH bias  $\delta \geq 1/2$ . If  $\mu$  is omitted, we assume the worst input distribution, e.g.,  $\beta(f) := \min_{\mu} \beta(f, \mu)$ . It is straightforward to generalize Fact 1 to quantum and nonlocal box settings [2].

**Fact 3.** For any  $\rho \in [0, 1]$  and  $\delta \in [1/2, 1]$ ,

$$\begin{aligned} \beta_{\text{NL}(\delta)}(f, \mu) &\geq \rho 2^{-C_{\text{NL}(\delta), \rho}(f, \mu)} \\ \beta^*(f, \mu) &\geq \rho 2^{-C_{\rho}^*(f, \mu)}. \end{aligned}$$

### 2.4 Fourier analysis

Let  $A: \{+1, -1\}^n \rightarrow \mathbb{R}$ . Let  $\mathbb{E}[A(z)] := \frac{1}{2^n} \sum_{z \in \{+1, -1\}^n} A(z)$ . Let  $\widehat{A}(S) := \mathbb{E}[A(z) \prod_{i \in S} z_i]$  for any  $S \subseteq [n] := \{1, \dots, n\}$ . Here,  $\widehat{A}(S)$  is called a Fourier coefficient. When we consider Fourier coefficients of boolean function  $\{0, 1\}^n \rightarrow \{0, 1\}$ , 0 and 1 are replaced by +1 and -1, respectively [11]. Let  $\|\widehat{A}\|_1 := \sum_{S \subseteq [n]} |\widehat{A}(S)|$ ,  $\|\widehat{A}\|_{\infty} := \max_{S \subseteq [n]} |\widehat{A}(S)|$  and  $\|\widehat{A}\|_0 := |\{S \subseteq [n] \mid \widehat{A}(S) \neq 0\}|$ .

### 3 Main theorems

In [12], it was shown that

$$\beta_{\text{NL}(\delta)}(f) \geq \delta^{C_{\rightarrow}(f)}$$

where  $C_{\rightarrow}(f)$  is a zero-error one-way communication complexity of  $f$ . In this paper, we improve the above inequality by using a two-way communication complexity.

**Theorem 3.** For any  $\rho \in [0, 1]$ ,  $\delta \in [1/2, 1]$ ,

$$\beta_{\text{NL}(\delta)}(f, \mu) \geq \rho \delta^{C_{\rho}(f, \mu)}.$$

Hence, for any  $\rho \in (0, 1]$  and  $\delta \in [1/2, 1)$ ,

$$C_{\rho}(f, \mu) \geq \frac{1}{\log \delta^{-1}} \log \frac{\rho}{\beta_{\text{NL}(\delta)}(f, \mu)}.$$

Theorem 3 seems to be similar to Fact 3. Theorem 3 and Fact 3 are generalizations of different types of Fact 1. In Fact 3, the communication complexities are replaced by those in stronger theories while the bases in the second factor remain 1/2. On the other hand, in Theorem 3, the base 1/2 of the exponent  $C_{\rho}(f, \mu)$  is improved to  $\delta$  while the communication complexity remains deterministic. Especially for  $\delta = 1/\sqrt{2}$ , we obtain

$$C_{\rho}(f, \mu) \geq 2 \log \frac{\rho}{\beta^*(f, \mu)}$$

from  $\beta^*(f, \mu) \geq \beta_{\text{NL}(1/\sqrt{2})}(f, \mu)$ . From the perfect parallel repetition theorem in quantum physics [5, 13], the following corollary is obtained.

**Corollary 4.** For any  $\ell \in \mathbb{N}$  and  $\rho \in (0, 1]$ ,

$$\frac{1}{\ell} C_{\rho^{\ell}}(f^{\oplus \ell}, \mu^{\otimes \ell}) \geq 2 \log \frac{\rho}{\beta^*(f, \mu)}.$$

The following theorem with Corollary 4 gives Theorem 1.

**Theorem 4.**

$$\lim_{n \rightarrow \infty} \min_{\mu} \beta^*(\text{EQ}_n, \mu) = \frac{1}{3}$$

$$\lim_{n \rightarrow \infty} \min_{\nu} \beta^*(\text{EQ}_n, \nu^{\otimes n}) = 2\sqrt{3} - 3$$

where  $\mu$  and  $\nu$  are probability distributions on  $\{0, 1\}^n \times \{0, 1\}^n$  and  $\{0, 1\} \times \{0, 1\}$ , respectively.

Furthermore, by applying generalized discrepancy theory [2, 1], we can obtain lower bounds for randomized communication complexity.

**Theorem 5.** For any  $h: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ ,  $\epsilon \in [0, 1/2]$ , and  $\delta \in [0, 1/2]$ ,

$$\beta_{\text{NL}(\delta)}(h, \mu) \geq \delta^{R_\epsilon(f)} \left( (1 - \epsilon) \mathbb{E}_{(x,y) \sim \mu} [f(x, y)h(x, y)] - \epsilon \right). \tag{2}$$

Hence, if  $(1 - \epsilon) \mathbb{E}_{(x,y) \sim \mu} [f(x, y)h(x, y)] - \epsilon > 0$ ,

$$R_\epsilon(f) \geq \frac{1}{\log \delta^{-1}} \log \max_{h, \mu} \frac{(1 - \epsilon) \mathbb{E}_{(x,y) \sim \mu} [f(x, y)h(x, y)] - \epsilon}{\beta_{\text{NL}(\delta)}(h, \mu)}.$$

For  $\delta = 1/\sqrt{2}$ , we obtain Theorem 2, which is exactly same as Linial and Shraibman’s lower bound [4, 1],

$$R_\epsilon(f) \geq 2 \log \gamma_2^{1/(1-2\epsilon)} - 2 \log \frac{1}{1 - 2\epsilon}$$

where  $\gamma_2^\alpha$  is some approximate norm of a communication matrix of  $f$  [4, 1]. Theorem 5 gives an intuitive proof of Linial and Shraibman’s lower bound and generalizations of Linial and Shraibman’s lower bound by using nonlocal boxes. It is not necessarily easy to upper bound  $\beta_{\text{NL}(\delta)}(f, \mu)$ . However, similarly to the relationship  $\beta^*(f, \mu) \geq \beta_{\text{NL}(1/\sqrt{2})}(f, \mu)$ , some relaxation may give an upper bound of  $\beta_{\text{NL}(\delta)}(f, \mu)$ . Note that Brassard et al. showed that  $\beta_{\text{NL}(\delta)}(f)$  is lower bounded by a positive constant for any  $f$  if  $\delta > \sqrt{2/3}$  [14]. In the area of foundation of quantum physics, it is a well-known open problem of whether  $\beta_{\text{NL}(\delta)}(f)$  is lower bounded by a positive constant for any  $f$  for  $\delta \in (1/\sqrt{2}, \sqrt{2/3}]$  [15, 12].

**4 Proofs of Theorem 3 and 5: Pawłowski et al.’s protocol**

**4.1 Intuition on the proof of Theorem 3**

The proof of Theorem 3 is the most important part in this paper. In the proof of Fact 1, a transcript of communication protocol is “guessed” by uniform random bits, which succeeds with probability at least  $2^{-k}$  where  $k$  denotes the length of the longest transcript. In the proof of Theorem 3, a correct transcript of communication protocol is “selected” by using isotropic nonlocal boxes with CHSH bias  $\delta$ . This “selection” can be implemented by a chain of 1-bit selectors  $\text{Addr}_1(x_0, x_1, y) := x_y$ , which has an XOR game protocol with bias  $\delta$ . We obtain Theorem 3 by showing that the chain of 1-bit selectors of length  $k$  has an XOR game protocol with bias at least  $\delta^k$ . This protocol may be regarded as Pawłowski et al.’s protocol for the pointer jumping function rather than the address function.

**4.2 Proof of Theorem 3**

The main idea comes from Pawłowski et al.’s protocol [9].

**Definition 5.** The address function  $\text{Addr}_n : \{0, 1\}^{2^n} \times \{0, 1\}^n \rightarrow \{0, 1\}$  is defined by

$$\text{Addr}_n(x_0, \dots, x_{2^n-1}, y_1, \dots, y_n) = x_y$$

where  $y := \sum_{i=1}^n y_i 2^{i-1}$ .

**Lemma 6.**

$$\beta_{\text{NL}(\delta)}(\text{Addr}_1) \geq \delta.$$

*Proof.* According to the equation

$$\text{Addr}_1(x_0, x_1, y_0) = x_0 \oplus y_0(x_0 \oplus x_1)$$

Alice and Bob put  $x_0 \oplus x_1$  and  $y_0$  into a nonlocal box and get  $a_0$  and  $b_0$ , respectively. Then, set  $a = x_0 \oplus a_0$  and  $b = b_0$  as the outputs of the XOR game. This protocol has bias  $\delta$ .  $\square$

Pawłowski et al. showed  $\beta_{\text{NL}(\delta)}(\text{Addr}_n) \geq \delta^n$  by the iterative application of Lemma 6 [9, 12]. For the proof of Theorem 3, we first show the following proposition.

**Proposition 7.** *Let  $X$  and  $Y$  be finite sets. For any  $f : X \times Y \rightarrow \{0, 1\}$  and probability distribution  $\mu$  on  $X \times Y$ ,*

$$\beta_{\text{NL}(\delta)}(f, \mu) \geq \delta^{C_1(f, \mu)}.$$

*Proof.* We show the proposition by the induction on the communication complexity. If  $C_1(f, \mu) = 0$ , the proposition trivially holds. Assume that the proposition holds for  $C_1(f, \mu) \leq k$ . Let  $P$  be the communication protocol for  $f$  with the transcript length at most  $k+1$ . Assume that at the first step of  $P$ , Alice sends a bit  $A_1(x)$  to Bob. Let  $f_0$  and  $f_1$  be the restrictions of  $f$  to the rectangles  $A_1^{-1}(0) \times Y$  and  $A_1^{-1}(1) \times Y$ , respectively. Then, the communication complexity of  $f_0$  and  $f_1$  are at most  $k$ . We can extend the domains of  $f_0$  and  $f_1$  to  $X \times Y$  while the values on the original domains and communication complexities are preserved. The extended functions are denoted by  $\bar{f}_0$  and  $\bar{f}_1$ . Then,

$$f(x, y) = \text{Addr}_1(\bar{f}_0(x, y), \bar{f}_1(x, y), A_1(x)).$$

From the hypothesis of the induction, there are XOR game protocols for  $\bar{f}_0$  and  $\bar{f}_1$  with bias at least  $\delta^k$ . By applying the XOR game protocols to  $\bar{f}_0$  and  $\bar{f}_1$ , Alice and Bob gets  $(a_0, a_1)$  and  $(b_0, b_1)$ , respectively, such that  $a_i \oplus b_i = \bar{f}_i(x, y)$  with bias  $\delta^k$  for each  $i \in \{0, 1\}$  independently. Let  $e_i := a_i \oplus b_i \oplus \bar{f}_i(x, y)$  for  $i \in \{0, 1\}$ . Then, we obtain

$$\begin{aligned} f(x, y) &= \text{Addr}_1(a_0 \oplus b_0 \oplus e_0, a_1 \oplus b_1 \oplus e_1, A_1(x)) \\ &= a_{A_1(x)} \oplus \text{Addr}_1(b_0, b_1, A_1(x)) \oplus e_{A_1(x)} \end{aligned}$$

From Lemma 6, Alice and Bob get  $a'$  and  $b'$  such that  $a' \oplus b' = \text{Addr}_1(b_0, b_1, A_1(x))$  with bias  $\delta$ . Let  $e' := a' \oplus b' \oplus \text{Addr}_1(b_0, b_1, A_1(x))$ . Then,

$$f(x, y) = (a_{A_1(x)} \oplus a') \oplus b' \oplus (e' \oplus e_{A_1(x)}).$$

Let  $a := (a_{A_1(x)} \oplus a')$  and  $b := b'$  be Alice and Bob's final output for the XOR game, respectively. This protocol has bias at least  $\delta^{k+1}$ .  $\square$

We can now straightforwardly show Theorem 3.

*The proof of Theorem 3.* A deterministic communication protocol corresponding to  $C_\rho(f, \mu)$  computes some function  $f'$  without error such that  $\mathbb{E}_{(x,y) \sim \mu}[f(x,y)f'(x,y)] \geq \rho$ . Hence, the XOR game protocol in Proposition 7 for  $f'$  has bias at least  $\rho\delta^{C_\rho(f,\mu)}$  for  $f$ .  $\square$

*Remark 8.* The proof of Proposition 7 cannot be generalized for quantum communication complexity nor communication complexity with nonlocal boxes straightforwardly since after Alice uses a nonlocal box, Bob can use the same nonlocal box at most once. We can generalize the above results to restricted protocols in which Alice and Bob must use common nonlocal boxes (quantum states) in common round. However, this seems to be restrictive since the standard quantum teleportation is not allowed in the restricted protocols.

### 4.3 Proof of Theorem 5

Theorem 5 is obtained by Proposition 7 with Buhrman’s idea for generalized discrepancy theory [1, 2]. First, we apply the protocol in Proposition 7 to  $f$  by using a randomized communication protocol corresponding to  $R_\epsilon(f)$ . This protocol computes  $h$  with an XOR of three errors: (a) Error in the computation of the chain of  $\text{Addr}_1$  in Proposition 7. (b) Error of the randomized protocol for  $f$ . (c) Error from the incoincidence of  $f$  and  $h$ . The bias of these errors are  $\delta^{R_\epsilon(f)}$ ,  $1 - 2\epsilon$  and  $\mathbb{E}_{(x,y) \sim \mu}[f(x,y)h(x,y)]$ , respectively. The error (a) is independent of errors (b) and (c). However, errors (b) and (c) are not independent. The XOR of errors (b) and (c) is zero with probability at least  $\Pr_{(x,y) \sim \mu}(f(x,y) = h(x,y))(1 - \epsilon)$ , which corresponds to a bias  $(1 - \epsilon)\mathbb{E}_{(x,y) \sim \mu}[f(x,y)h(x,y)] - \epsilon$ . Hence, we obtain (2).

## 5 Proof of Theorem 4: Bias of XOR game for XOR functions

### 5.1 XOR game for XOR functions

Let  $g^\oplus(x, y) := g(x \oplus y)$  for any  $g: \{0, 1\}^n \rightarrow \{0, 1\}$ . A function in this form is called an XOR function. Let  $q: \{0, 1\}^n \rightarrow \mathbb{R}$  be a non-negative function with  $\sum_z q(z) = 2^n$ . The largest bias of XOR game of XOR function  $g^\oplus$  on XOR input distribution  $2^{-2n}q^\oplus$  can be represented by the largest Fourier amplitude of  $g(z)q(z)$ .

**Lemma 9.**

$$\beta(g^\oplus, 2^{-2n}q^\oplus) = \max_{S \subseteq [n]} |\widehat{gq}(S)| = \|\widehat{gq}\|_\infty$$

where  $(gq)(z) := g(z)q(z)$ .

*Proof.* When an input distribution is fixed, shared random bits do not help to increase the winning probability of an XOR game. Hence, without loss of generality, we can assume that Alice and Bob output  $a = A(x)$  and  $b = B(y)$ , respectively where  $A$  and  $B$  are deterministic boolean functions from  $\{0, 1\}^n$  to  $\{0, 1\}$ . Then, we obtain an upper bound of bias of an XOR game for a XOR function (similarly to the proof on the BLR test in [11])

$$\begin{aligned} \mathbb{E}[g(x \oplus y)q(x \oplus y)A(x)B(y)] &= \sum_{S \subseteq [n]} \widehat{gq}(S)\widehat{A}(S)\widehat{B}(S) \\ &\leq \max_{S \subseteq [n]} |\widehat{gq}(S)| \sum_{S \subseteq [n]} |\widehat{A}(S)\widehat{B}(S)| \\ &\leq \max_{S \subseteq [n]} |\widehat{gq}(S)|. \end{aligned}$$



The last inequality is obtained by the Cauchy–Schwartz inequality and  $\sum_{S \subseteq [n]} \widehat{A}(S)^2 = \sum_{S \subseteq [n]} \widehat{B}(S)^2 = 1$ . Let  $S^* := \operatorname{argmax}_{S \subseteq [n]} |\widehat{gq}(S)|$ . This upper bound can be achieved by  $A(x) = \bigoplus_{i \in S^*} x_i$ ,  $B(y) = \operatorname{sign}(\widehat{gq}(S^*)) \oplus \bigoplus_{i \in S^*} y_i$  where  $\operatorname{sign}(x)$  is 0 if  $x \geq 0$  and 1 otherwise.  $\square$

For general input distribution, Alice and Bob easily make the distribution XOR by using shared random bits since  $g(x \oplus y) = g((x \oplus r) \oplus (y \oplus r))$ . The probability distribution of  $(x', y') := (x \oplus r, y \oplus r)$  is an XOR function. In other word, the worst case input distribution must be an XOR function.

**Lemma 10.**

$$\beta(g^\oplus) = \min_q \|\widehat{gq}\|_\infty.$$

*Remark 11.* If we consider worst boolean function  $g$ , we obtain the lower bound

$$\min_{g,q} \max_{S \subseteq [n]} |\widehat{gq}(S)| \geq \sqrt{\frac{1}{2^n} \mathbb{E}[q(z)^2]} \geq 2^{-\frac{n}{2}} \tag{3}$$

from Parseval’s identity  $\sum_S \widehat{gq}(S)^2 = \mathbb{E}[q(z)^2]$ . This lower bound can be achieved by bent functions and the uniform input distribution [11]. This lower bound  $2^{-n/2}$  is not applicable for general non-XOR function. If  $n = 1$  and  $f(x, y) = x \wedge y$ , i.e., CHSH game, the maximum bias  $\beta(\text{AND}) = 1/2 < 1/\sqrt{2}$ . For general non-XOR functions, Littlewood’s 4/3 inequality gives a lower bound  $2^{-(n+1)/2}$  [16].

Linden et al. showed  $\beta^*(g^\oplus, 2^{-2n}q^\oplus) = \beta(g^\oplus, 2^{-2n}q^\oplus)$  for any  $g$  and  $q$  [17]. We give another proof using Tsirelson’s characterization and Fourier analysis in Appendix A.

### 5.2 Equality function

The negation of the equality function  $\text{EQ}_n$  is an XOR function  $\text{OR}_n^\oplus$ .

**Lemma 12.**

$$\lim_{n \rightarrow \infty} \min_q \|\widehat{\text{OR}_n q}\|_\infty = \frac{1}{3}. \tag{4}$$

*Proof.* First, we show  $\lim_{n \rightarrow \infty} \min_q \|\widehat{\text{OR}_n q}\|_\infty \leq 1/3$ . We consider an optimization of the input distribution  $\mu(z) := 2^{-n}q(z)$  only among a class

$$\mu(z) = \begin{cases} \lambda, & w(z) = 0 \\ \frac{1-\lambda}{\binom{n}{2}}, & w(z) = 2 \\ 0, & \text{otherwise} \end{cases}$$

where  $w(z)$  denotes the Hamming weight of  $z \in \{0, 1\}^n$  and  $\lambda \in [0, 1]$  is a parameter. Then,

$$\begin{aligned} & \min_q \max_{S \subseteq [n]} \left| \mathbb{E} \left[ \text{OR}_n(z) q(z) \prod_{i \in S} z_i \right] \right| \\ & \leq \min_\lambda \max_{S \subseteq [n]} \left| \lambda - (1-\lambda) \left( \frac{(n-|S|)(n-|S|-1)}{n(n-1)} - 2 \frac{(n-|S|)|S|}{n(n-1)} + \frac{|S|(|S|-1)}{n(n-1)} \right) \right| \\ & = \min_\lambda \max_{k=0,1,\dots,n} \left| \lambda - (1-\lambda) \left( \frac{n^2 - 4nk + 4k^2 - n}{n(n-1)} \right) \right| \end{aligned}$$

$$\begin{aligned}
 &= \min_{\lambda} \max_{k=0,1,\dots,n} \left| 2\lambda - 1 + (1 - \lambda) \left( \frac{4k(n-k)}{n(n-1)} \right) \right| \\
 &\leq \min_{\lambda} \max \left\{ 1 - 2\lambda, 2\lambda - 1 + (1 - \lambda) \frac{n}{n-1} \right\} \\
 &= \min_{\lambda} \max \{ 1 - 2\lambda, \lambda + \Theta(n^{-1}) \}
 \end{aligned}$$

By solving  $1 - 2\lambda = \lambda$ , we obtain  $\lambda = 1/3$ .

Next, we show  $\min_q \|\widehat{\text{OR}}_n q\|_{\infty} \geq 1/3$  for any  $n$ . From the minimax principle, it is sufficient to show a randomized protocol for the XOR game for the negation of the equality function with bias at least  $1/3$  for any input  $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$ . We first consider two protocols for XOR game of the negation of the equality function, and then take a probabilistic mixture of them. The first protocol always answer 1, i.e.,  $A(x) = 0, B(y) = 1$ . The second protocol uses shared random bits  $r \in \{0, 1\}^n$  and take inner products with inputs, i.e.,  $A(x) = \langle x, r \rangle, B(y) = \langle y, r \rangle$ . The bias of the first protocol is  $1 - 2\lambda$  where  $\lambda$  denotes the probability of  $x = y$ . The bias of the second protocol is  $\lambda$  [8]. Hence, if we choose the first protocol with probability  $1/3$  and choose the second protocol with probability  $2/3$ , we get the bias  $1/3$ .  $\square$

Next, we show the worst i.i.d. input distribution for the equality function. Similarly to the general case, we can assume that an input distribution for each bit is an XOR function.

**Lemma 13.** For  $n \geq 2$ ,

$$\min_{\nu} \left\| \widehat{\text{OR}}_n (2\nu)^{\otimes n} \right\|_{\infty} = 1 - 2\lambda^{*n}$$

where  $\nu$  denotes a distribution on  $\{0, 1\}$ , and where  $\lambda^*$  denotes the unique root in  $[1/2, 1]$  of

$$4\lambda^n - (2\lambda - 1)^n - 1 = 0. \tag{5}$$

Furthermore,

$$\lim_{n \rightarrow \infty} \min_{\nu} \left\| \widehat{\text{OR}}_n (2\nu)^{\otimes n} \right\|_{\infty} = 2\sqrt{3} - 3.$$

*Proof.*

$$\begin{aligned}
 \min_{\nu} \max_{S \subseteq [n]} |\widehat{\text{OR}}_n (2\nu)^{\otimes n}(S)| &= \min_{\nu} \max_{S \subseteq [n]} \left| \mathbb{E} \left[ \text{OR}(z) (2\nu)^{\otimes n}(z) \prod_{i \in S} z_i \right] \right| \\
 &= \min_{\nu} \max_{S \subseteq [n]} \left| \mathbb{E} \left[ (2\mathbb{I}\{z = +1\} - 1) (2\nu)^{\otimes n}(z) \prod_{i \in S} z_i \right] \right| \\
 &= \min_{\nu} \max_{S \subseteq [n]} \left| 2\nu(0)^n - (\nu(0) - \nu(1))^{|S|} \right| \\
 &= \min_{\nu} \max \{ 1 - 2\nu(0)^n, 2\nu(0)^n - (\nu(0) - \nu(1))^n \} \\
 &= \min_{\nu} \max \{ 1 - 2\nu(0)^n, 2\nu(0)^n - (2\nu(0) - 1)^n \}
 \end{aligned}$$

Let  $u_n(\lambda) := 1 - 2\lambda^n$  and  $v_n(\lambda) := 2\lambda^n - (2\lambda - 1)^n$ . It is easy to see that  $u_n(\lambda)$  is monotonically decreasing for  $\lambda \in [0, 1]$ , and that  $v_n(\lambda)$  is monotonically increasing for  $\lambda \in [1/2, 1]$ . For  $n \geq 2$ ,  $u_n(1/2) \geq v_n(1/2)$ . Hence, for  $n \geq 2$ , we obtain

$$\min_{\nu} \max \{ 1 - 2\nu(0)^n, 2\nu(0)^n - (2\nu(0) - 1)^n \} = 1 - 2\lambda^{*n}$$

where  $\lambda^*$  is the unique root in  $[1/2, 1]$  of (5). Let  $\lambda_n = 1 - c/n$  for some constant  $c > 0$ . Then, we obtain

$$\lim_{n \rightarrow \infty} 4\lambda_n^n - (2\lambda_n - 1)^n - 1 = 4e^{-c} - e^{-2c} - 1$$

Let  $\varphi = e^{-c} < 1$ .

$$4\varphi - \varphi^2 - 1 = 0 \iff \varphi^2 - 4\varphi + 1 = 0 \iff \varphi = 2 - \sqrt{3}$$

Then, we obtain

$$\lim_{n \rightarrow \infty} 1 - 2\lambda^{*n} = 1 - 2e^{-c} = 1 - 2(2 - \sqrt{3}) = 2\sqrt{3} - 3. \quad \square$$

## 6 Remarks

### 6.1 Linial and Shraibman's lower bound for XOR functions

When we apply Linial and Shraibman's lower bound (1) for XOR function  $g^\oplus$ , we obtain

$$R_\epsilon(g^\oplus) \geq 2 \log \max_{h, \mu} \frac{(1 - \epsilon) \mathbb{E}_{(x, y) \sim \mu} [g^\oplus(x, y)h(x, y)] - \epsilon}{\beta^*(h, \mu)}.$$

For  $\epsilon = 0$ , we obtain

$$\begin{aligned} R_0(g^\oplus) &\geq 2 \log \max_{h, q} \frac{\mathbb{E}_{(x, y) \sim 2^{-2n} q^\oplus} [g^\oplus(x, y)h^\oplus(x, y)]}{\beta^*(h^\oplus, 2^{-2n} q^\oplus)} \\ &= 2 \log \max_{h, q} \frac{\sum_S \widehat{g}(S) \widehat{h} q(S)}{\|\widehat{h} q\|_\infty} = 2 \log \|\widehat{g}\|_1. \end{aligned} \tag{6}$$

This fact was shown in [4, 1]. For  $\epsilon \in (0, 1/2)$ , Linial and Shraibman's lower bound can be written as

$$R_\epsilon(g^\oplus) \geq 2 \log \left[ (1 - \epsilon) \|\widehat{g}\|_1^{\frac{\epsilon}{1-\epsilon}} \right] \tag{7}$$

where

$$\|\widehat{g}\|_1^\epsilon := \min \left\{ \|\widehat{h}\|_1 \mid h: \{+1, -1\}^n \rightarrow \mathbb{R}, |h(x) - g(x)| \leq \epsilon \forall x \right\}$$

is called approximate Fourier  $\ell_1$  norm [4]. Obviously,  $\|\widehat{g}\|_1^\epsilon \leq (1 - \epsilon) \|\widehat{g}\|_1$ . However, no lower bound of the approximate Fourier  $\ell_1$  norm by the exact Fourier  $\ell_1$  norm has been known for general  $\epsilon \in (0, 1)$ . Here, we give a necessary and sufficient condition for the equality  $\|\widehat{g}\|_1^\epsilon = (1 - \epsilon) \|\widehat{g}\|_1$ .

**Definition 14.** For  $f: \{+1, -1\}^n \rightarrow \{+1, -1\}$ ,

$$\begin{aligned} f^* &:= \{g: \{+1, -1\}^n \rightarrow [-1, +1] \mid g(x) = \text{sign}(\widehat{f}(S_x)), \forall x, \widehat{f}(S_x) \neq 0\} \\ f^{**} &:= \{h: \{+1, -1\}^n \rightarrow [-1, +1] \mid \exists g \in f^*, h(x) = \text{sign}(\widehat{g}(S_x)), \forall x, \widehat{g}(S_x) \neq 0\}. \end{aligned}$$

where  $S_x := \{i \mid x_i = -1\}$  and  $\text{sign}(r) := r/|r|$  for  $r \neq 0$ .

The complementary slackness condition gives the following theorem.

**Theorem 6.**  $\|\widehat{f}\|_1^\epsilon = (1 - \epsilon)\|\widehat{f}\|_1$  if and only if  $f \in f^{**}$ .

The proof is in Appendix B.

*Remark 15.* Exhaustive search on computer shows that all of 256 boolean functions on 3 variables, 51200 of 65536 boolean functions on 4 variables and at least 2839187456 of 4294967296 boolean functions on 5 variables satisfy  $f \in f^{**}$ .

For  $\text{OR}_n$ ,  $\text{OR}_n^*$  only includes the negation of  $\text{OR}_n$ . Hence,  $\text{OR}_n^{**}$  only includes  $\text{OR}_n$ . Then, Theorem 6 gives  $\|\widehat{\text{OR}_n}\|_1^\epsilon = (1 - \epsilon)\|\widehat{\text{OR}_n}\|_1$  so that

$$R_\epsilon(\text{EQ}_n) \geq 2 \log \left[ (1 - 2\epsilon)\|\widehat{\text{OR}}\|_1 \right] = 2 \log \left[ (1 - 2\epsilon)(3 - 2^{-n+2}) \right]$$

from (7). In general, it is difficult to lower bound the approximate Fourier  $\ell_1$  norm. The following theorem gives a simple lower bound which may be useful for some case.

**Theorem 7.**

$$R_\epsilon(g^\oplus) \geq 2 \log \left( (1 - \epsilon)\|\widehat{g}\|_1 - \epsilon\|\widehat{g}^*\|_1 \right).$$

where

$$\|\widehat{g}^*\|_1 := \min\{\|\widehat{h}\|_1 \mid h \in g^*\}.$$

*Proof.* From Linial and Shraibman's bound for XOR functions,

$$R_\epsilon(g^\oplus) \geq 2 \log \max_{h,q} \frac{(1 - \epsilon) \sum_S \widehat{g}(S) \widehat{h}q(S) - \epsilon}{\|\widehat{h}q\|_\infty}.$$

Here  $hq: \{+1, -1\}^n \rightarrow \mathbb{R}$  is an arbitrary function satisfying  $\sum_{z \in \{+1, -1\}^n} |hq(z)| = 2^n$ . For some  $G \in g^*$ , we choose  $hq$  such that

$$\widehat{h}q(S) = \|\widehat{h}q\|_\infty G(S_x).$$

Then,

$$\begin{aligned} hq(z) &= \sum_{S \subseteq [n]} \|\widehat{h}q\|_\infty G(S_x) \prod_{i \in S} z_i \\ &= \|\widehat{h}q\|_\infty 2^n \widehat{G}(S_z). \end{aligned}$$

Hence,

$$\|\widehat{h}q\|_\infty = \frac{1}{\|\widehat{G}\|_1}.$$

The theorem is obtained. □

From the above proof, we obtain  $\|\widehat{g}^*\|_1 \geq \|\widehat{g}\|_1$ . From the Cauchy–Schwartz inequality, we obtain  $\|\widehat{g}^*\|_1 \leq \sqrt{\|\widehat{g}\|_0}$ , which gives a weaker bound  $R_\epsilon(g^\oplus) \geq 2 \log \left( (1 - \epsilon)\|\widehat{g}\|_1 - \epsilon\sqrt{\|\widehat{g}\|_0} \right)$  shown in [4].

## 6.2 Relationship with information complexity of the equality function

Braverman and Rao showed that information complexity is equal to (direct product-)amortized communication complexity [18]. The information complexity is defined by

$$\text{IC}(f, \mu) := \min_{\pi} I(X; \pi(X, Y) | Y) + I(Y; \pi(X, Y) | X)$$

where  $I$  denotes the mutual information, where  $\pi$  denotes a protocol, and where  $\pi(X, Y)$  denotes the public randomness and a transcript when  $X$  and  $Y$  are given to a protocol  $\pi$  (See [18] and [19] for details). In [19], the information complexity of the equality function is upper bounded by 9. In this section, this upper bound is improved to  $2 \log 5 \approx 4.64$ . We consider a particular protocol introduced in [19]. Alice and Bob use a shared invertible random matrix  $A$ , whose  $i$ -th row is denoted by  $a_i$ . At  $i$ -th step Alice and Bob send  $\langle x, a_i \rangle$  and  $\langle y, a_i \rangle$  to each other, respectively. If  $\langle x, a_i \rangle \neq \langle y, a_i \rangle$ , the protocol terminates and output 0. If  $\langle x, a_i \rangle = \langle y, a_i \rangle$  for all  $i = 1, \dots, n$ , the protocol output 1. Let  $Z$  be a random variable taking a value 1 if  $x = y$  and 0 if  $x \neq y$ . Assume  $\Pr(Z = 1) = \lambda$ . Then, we obtain

$$\begin{aligned} I(X; \pi(X, Y) | Y) &= I(X, Z; \pi(X, Y) | Y) \\ &= I(Z; \pi(X, Y) | Y) + I(X; \pi(X, Y) | Y, Z) \\ &= H(Z | Y) + I(X; \pi(X, Y) | Y, Z) \\ &\leq h(\lambda) + (1 - \lambda) \sum_{i \geq 1} \frac{1}{2^i} i \\ &= h(\lambda) + 2(1 - \lambda) \end{aligned}$$

where  $h(\lambda) := -\lambda \log \lambda - (1 - \lambda) \log(1 - \lambda)$ . Here,  $h(\lambda) + 2(1 - \lambda)$  is maximized at  $\lambda = 1/5$  with the maximum  $\log 5$ . Hence,  $2 \log 5$  is an upper bound of the information complexity.

Here, any protocol obviously requires at least  $h(\lambda)$  bits for each direction in average. If we assume that at least extra 1 bit is required for each direction if  $x \neq y$ , we obtain a lower bound  $h(\lambda) + 1 - \lambda$  for each direction, which is maximized at  $\lambda = 1/3$  with the maximum  $\log 3$ . Hence, the lower bound  $2 \log 3$  of XOR-amortized communication complexity in Theorem 1 intuitively means that “each of the equality problems must be solved by using at least  $h(\lambda) + 1 - \lambda$  bits for each direction”.

### Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP17K17711. The author would like to thank Mark Braverman for the insight that the lower bound 2 for XOR-amortized communication complexity of the equality function seems to be trivial. The author would like to thank Ryan O’Donnell for insightful discussions.

### References

1. T. Lee and A. Shraibman (2009), *Lower bounds in communication complexity*, Foundations and Trends® in Theoretical Computer Science, 3, pp. 263–399.
2. J. Briët, H. Buhrman, T. Lee, and T. Vidick (2013), *Multipartite entanglement in XOR games*, Quantum Information and Computation, 13, pp. 334–360.
3. J. Degorre, M. Kaplan, S. Laplante, and J. Roland (2011), *The communication complexity of non-signaling distributions*, Quantum Inf. Comput., 11, pp. 649–676.

4. N. Linial and A. Shraibman (2009), *Lower bounds in communication complexity based on factorization norms*, Random Structures & Algorithms, 34, pp. 368–394.
5. R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay (2008), *Perfect parallel repetition theorem for quantum XOR proof systems*, computational complexity, 17, pp. 282–299.
6. B. Cirel’son (1980), *Quantum generalizations of Bell’s inequality*, Lett. Math. Phys., 4, pp. 93–100.
7. I. Kremer (1995), *Quantum communication*, Master’s thesis, The Hebrew University of Jerusalem.
8. E. Kushilevitz and N. Nisan (1997), *Communication Complexity*, Cambridge University Press, Cambridge.
9. M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski (2009), *Information causality as a physical principle*, Nature, 461, pp. 1101–1104.
10. J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt (1969), *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett., 23, pp. 880–884.
11. R. O’Donnell (2014), *Analysis of Boolean Functions*, Cambridge University Press.
12. R. Mori (2016), *Three-input majority function as the unique optimal function for the bias amplification using nonlocal boxes*, Phys. Rev. A, 94, p. 052130.
13. T. Lee, A. Shraibman, and R. Špalek (2008), *A direct product theorem for discrepancy*, in The 23rd Annual IEEE Conference on Computational Complexity, 2008, pp. 71–80, IEEE.
14. G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger (2006), *Limit on nonlocality in any world in which communication complexity is not trivial*, Phys. Rev. Lett., 96, p. 250401.
15. H. Buhrman, R. Cleve, S. Massar, and R. de Wolf (2010), *Nonlocality and communication complexity*, Rev. Mod. Phys., 82, pp. 665–698.
16. A. Montanaro (2012), *Some applications of hypercontractive inequalities in quantum information theory*, Journal of Mathematical Physics, 53, p. 122206.
17. N. Linden, S. Popescu, A. J. Short, and A. Winter (2007), *Quantum nonlocality and beyond: Limits from nonlocal computation*, Phys. Rev. Lett., 99, p. 180502.
18. M. Braverman and A. Rao (2014), *Information equals amortized communication*, IEEE Transactions on Information Theory, 60, pp. 6058–6069.
19. M. Braverman (2015), *Interactive information complexity*, SIAM Journal on Computing, 44, pp. 1698–1739.
20. B. Tsirel’son (1987), *Quantum analogues of the Bell inequalities. The case of two spatially separated domains*, Journal of Soviet Mathematics, 36, pp. 557–570.

## Appendix A Quantum XOR game for XOR functions

**Lemma 16** ([17]).  $\beta^*(g^\oplus, 2^{-2n}q^\oplus) = \beta(g^\oplus, 2^{-2n}q^\oplus)$  for any  $g$  and  $q$ .

*Proof.* From Tsirelson’s characterization [20],  $\beta^*(g^\oplus, 2^{-n}q^\oplus)$  is equal to

$$\max_{V,W} \mathbb{E} [g(x \oplus y)q(x \oplus y)\langle V(x), W(y) \rangle]$$

where  $V(x)$  and  $W(y)$  are unit vectors on  $\mathbb{R}$  of dimension  $2n$  for all  $x, y \in \{0, 1\}^n$ . Then, we obtain

$$\begin{aligned} \mathbb{E} [g(x \oplus y)q(x \oplus y)\langle V(x), W(y) \rangle] &= \mathbb{E} \left[ g(x \oplus y)q(x \oplus y) \sum_{i=1}^{2n} V_i(x)W_i(y) \right] \\ &= \sum_{S \subseteq [n]} \widehat{gq}(S) \sum_{i=1}^{2n} \widehat{V}_i(S)\widehat{W}_i(S) \\ &\leq \max_{S \subseteq [n]} |\widehat{gq}(S)| \sum_{S \subseteq [n]} \sum_{i=1}^{2n} \left| \widehat{V}_i(S)\widehat{W}_i(S) \right| \end{aligned}$$

$$\begin{aligned}
 &\leq \max_{S \subseteq [n]} |\widehat{g}q(S)| \sqrt{\left( \sum_{S \subseteq [n]} \sum_{i=1}^{2n} \widehat{V}_i(S)^2 \right) \left( \sum_{S \subseteq [n]} \sum_{i=1}^{2n} \widehat{W}_i(S)^2 \right)} \\
 &= \max_{S \subseteq [n]} |\widehat{g}q(S)| \sqrt{\left( \frac{1}{2^n} \sum_{i=1}^{2n} \sum_{x \in \{0,1\}^n} V_i(x)^2 \right) \left( \frac{1}{2^n} \sum_{i=1}^{2n} \sum_{y \in \{0,1\}^n} W_i(y)^2 \right)} \\
 &= \max_{S \subseteq [n]} |\widehat{g}q(S)| \sqrt{\left( \sum_{x \in \{0,1\}^n} \frac{1}{2^n} \right) \left( \sum_{y \in \{0,1\}^n} \frac{1}{2^n} \right)} \\
 &= \max_{S \subseteq [n]} |\widehat{g}q(S)|. \quad \square
 \end{aligned}$$

**Appendix B Proof of Theorem 6**

Let  $\chi_S(x) := \prod_{i \in S} x_i$ . Then,  $\|\widehat{f}\|_1^\epsilon$  is the solution of the following optimization problem.

$$\begin{aligned}
 &\min : \sum_{S \subseteq [n]} \left| \frac{1}{2^n} \sum_x g(x) \chi_S(x) \right| \\
 &\text{subject to : } |g(x) - f(x)| \leq \epsilon, \quad \forall x \in \{+1, -1\}^n.
 \end{aligned}$$

**Lemma 17.**  $g: \{+1, -1\}^n \rightarrow \mathbb{R}$  is optimal of the above optimization problem if and only if there exists  $h \in g^*$  such that

$$g(x) = \begin{cases} f(x) - \epsilon, & \text{if } \widehat{h}(S_x) > 0 \\ f(x) + \epsilon, & \text{if } \widehat{h}(S_x) < 0 \\ \in [f(x) - \epsilon, f(x) + \epsilon], & \text{otherwise.} \end{cases}$$

*Proof.* The optimization problem is equivalent to the following linear program.

$$\begin{aligned}
 &\min : \sum_{S \subseteq [n]} (\widehat{g}_+(S) + \widehat{g}_-(S)) \\
 &\text{subject to : } \widehat{g}_+(S) - \widehat{g}_-(S) = \frac{1}{2^n} \sum_x g(x) \chi_S(x), \quad \forall S \subseteq [n] \\
 &\quad g(x) - f(x) \leq \epsilon, \quad \forall x \in \{+1, -1\}^n \\
 &\quad f(x) - g(x) \leq \epsilon, \quad \forall x \in \{+1, -1\}^n \\
 &\quad \widehat{g}_+(S) \geq 0, \quad \forall S \subseteq [n] \\
 &\quad \widehat{g}_-(S) \geq 0, \quad \forall S \subseteq [n]
 \end{aligned}$$

where  $\widehat{g}_+(S)$  and  $\widehat{g}_-(S)$  are variables of the above linear program. The Lagrangian of this optimization problem is

$$\begin{aligned}
 \mathcal{L}(g, \widehat{g}_+, \widehat{g}_-, c, \lambda, \rho, \mu_+, \mu_-) &:= \sum_{S \subseteq [n]} (\widehat{g}_+(S) + \widehat{g}_-(S)) \\
 &\quad - \sum_S c_S \left( \widehat{g}_+(S) - \widehat{g}_-(S) - \frac{1}{2^n} \sum_x g(x) \chi_S(x) \right)
 \end{aligned}$$

$$\begin{aligned}
 & + \sum_x \lambda(x) (g(x) - f(x) - \epsilon) + \sum_x \rho(x) (f(x) - g(x) - \epsilon) \\
 & - \sum_S \mu_+(S) \widehat{g}_+(S) - \sum_S \mu_-(S) \widehat{g}_-(S)
 \end{aligned}$$

where the dual feasibility condition is

$$\begin{aligned}
 \lambda(x) &\geq 0, & \rho(x) &\geq 0, & \forall x \in \{+1, -1\}^n \\
 \mu_+(S) &\geq 0, & \mu_-(S) &\geq 0, & \forall S \subseteq [n].
 \end{aligned}$$

First, we consider necessary conditions for optimal primal and dual solutions. The partial derivatives of Lagrangian with respect to primal variables are

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial g(x)} &= \lambda(x) - \rho(x) + \frac{1}{2^n} \sum_S c_S \chi_S(x) \\
 \frac{\partial \mathcal{L}}{\partial \widehat{g}_+(S)} &= 1 - c_S - \mu_+(S) \\
 \frac{\partial \mathcal{L}}{\partial \widehat{g}_-(S)} &= 1 + c_S - \mu_-(S).
 \end{aligned}$$

All of them must be zero at any optimal primal dual pair. Hence,  $c_S \in [-1, +1]$  for any  $S \subseteq [n]$ . A complementary slackness condition says

$$\begin{aligned}
 \widehat{g}_+(S) > 0 &\Rightarrow \mu_+(S) = 0 \iff c_S = +1 \\
 \widehat{g}_-(S) > 0 &\Rightarrow \mu_-(S) = 0 \iff c_S = -1.
 \end{aligned}$$

Since for any primal optimal solution,  $\widehat{g}_+(S) > 0 \iff \widehat{g}(S) > 0$  and  $\widehat{g}_-(S) > 0 \iff \widehat{g}(S) < 0$ , the above condition means that  $h(x) := c_{S_x}$  must be a member of  $g^*$ . Furthermore, from  $\partial \mathcal{L} / \partial g(x) = 0$ ,

$$\begin{aligned}
 \widehat{h}(S_x) > 0 &\Rightarrow \rho(x) > 0 \Rightarrow g(x) = f(x) - \epsilon \\
 \widehat{h}(S_x) < 0 &\Rightarrow \lambda(x) > 0 \Rightarrow g(x) = f(x) + \epsilon
 \end{aligned}$$

This shows the one direction of this Theorem.

Conversely, if there exists  $h \in g^*$  satisfying the condition, we can choose values

$$\begin{aligned}
 \widehat{g}_+(S) &= \max\{0, \widehat{g}(S)\}, & \widehat{g}_-(S) &= \max\{0, -\widehat{g}(S)\} \\
 c_{S_x} &= h(x) \\
 \lambda(x) &= \max\{0, -\widehat{h}(S_x)\}, & \rho(x) &= \max\{0, \widehat{h}(S_x)\} \\
 \mu_+(S) &= 1 - c_S, & \mu_-(S) &= 1 + c_S
 \end{aligned}$$

which satisfy all complementary slackness conditions, and hence are optimal. □

Theorem 6 is immediately obtained from Lemma 17.