Three-input Majority Function is the Unique Optimal for the Bias Amplification using Nonlocal Boxes

Ryuhei Mori
School of Computing, Tokyo Institute of Technology, Tokyo 152-8552, Japan
(Dated: May 2, 2016)

Brassard et al. [Phys. Rev. Lett. 96, 250401 (2006)] showed that shared nonlocal boxes with the CHSH probability greater than $\frac{3+\sqrt{2}}{4}$ yields trivial communication complexity. There still exists the gap with the maximum CHSH probability $\frac{3+\sqrt{2}}{4}$ achievable by quantum mechanics. It is an interesting open question to determine the exact threshold for the trivial communication complexity. Brassard et al.’s idea is based on the recursive bias amplification by the 3-input majority function. It was not obvious if other choice of function exhibits stronger bias amplification. We show that the 3-input majority function is the unique optimal, so that one cannot improve the threshold $\frac{3+\sqrt{2}}{4}$ by Brassard et al.’s bias amplification.

PACS numbers: 03.65.Ud, 03.65.Ta, 03.67.Mn

I. INTRODUCTION

Bell showed that quantum mechanics allows correlations that cannot be simulated by classical physics [1]. Clauser, Horne, Shimony, and Holt (CHSH) found simpler nonclassical correlations [2]. Apart from the concrete mathematical description of quantum mechanics, we can only consider abstract statistical behavior realized by quantum mechanics. A nonlocal box is an abstract device which represents statistical behavior of separable measurements on a possibly entangled state on quantum mechanics and superquantum theory as well. A nonlocal box is assumed to be shared by two parties, Alice and Bob. A nonlocal box has input ports and output ports on the both sides. A nonlocal box is specified by the conditional probability distribution $p(a,b | x,y)$ representing the probability of outputting $a$ to Alice and $b$ to Bob when Alice and Bob input $x$ and $y$ into the nonlocal box, respectively. Here, all of $x$, $y$, $a$, and $b$ are assumed to be either of 0 or 1. They cannot communicate by using the nonlocal box since it satisfies the no-signaling condition

$$\sum_{b \in \{0,1\}} p(a,b | x,0) = \sum_{b \in \{0,1\}} p(a,b | x,1)$$
$$\sum_{a \in \{0,1\}} p(a,b | 0,y) = \sum_{a \in \{0,1\}} p(a,b | 1,y).$$

The CHSH probability $P_{\text{CHSH}}$ is a measure of the nonlocality of the nonlocal box defined by

$$P_{\text{CHSH}} := \frac{1}{4} \sum_{x,y} \sum_{a,b} p(a,b | x,y).$$

While the maximum CHSH probability given by classical physics is $P_{\text{CHSH}} = 3/4$, that for quantum mechanics is $P_{\text{CHSH}} = \frac{3+\sqrt{2}}{4}$ [3]. On the other hand, Popescu and Rohrlich showed that there exists the nonlocal box, called the PR box, with $P_{\text{CHSH}} = 1$ [4]. Hence, it is a natural question why quantum mechanics cannot achieve the CHSH probability greater than $\frac{3+\sqrt{2}}{4}$. Van Dam showed that if Alice and Bob share unlimited number of PR boxes, they can compute arbitrary function $f(x,y)$ only by sending 1 bit to each other where $x$ and $y$ are $n$ bits owned by Alice and Bob, respectively [5]. It gives the explanation why Nature does not allow $P_{\text{CHSH}} = 1$ since we strongly believe that the trivial communication complexity must not be allowed by Nature. Furthermore, Brassard et al. showed that the nonlocal box with $P_{\text{CHSH}} > \frac{3+\sqrt{2}}{4}$ yields the trivial communication complexity on the probabilistic setting [6]. It has not been known whether or not the communication complexity is trivial when the CHSH probability is between $\frac{3+\sqrt{2}}{4}$ and $\frac{3+\sqrt{2}}{4}$. Later, Pawlowski et al. completely characterized the quantum CHSH probability $\frac{3+\sqrt{2}}{4}$ by using new principle called information causality [7]. However, it is still interesting to determine the exact threshold of $P_{\text{CHSH}}$ for the trivial communication complexity.

In this paper, we show that the trivial communication complexity below $\frac{3+\sqrt{2}}{4}$ cannot be proved by Brassard et al.’s technique. Their technique is based on the recursive bias amplification from exponentially small bias to constant bias by using the 3-input majority function $\text{Maj}_3$. It was not obvious that $\text{Maj}_3$ is the best choice for the bias amplification. It seems to be curious that $\text{Maj}_3$ is the optimal function for the bias amplification if it is true. In this paper, we show that $\text{Maj}_3$ is the unique optimal function for the bias amplification.

**Theorem 1.** The 3-input majority function is the unique optimal for Brassard et al.’s technique of the bias amplification using the nonlocal boxes. Hence, one cannot obtain the threshold for the trivial communication complexity smaller than $\frac{3+\sqrt{2}}{4}$ by Brassard et al.’s technique.

For the proof of Theorem 1, we use the Fourier analysis of boolean functions developed in theoretical computer science.
II. PRELIMINARIES

We introduce some notions and notations.

Definition 2. For a boolean function $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$, XOR protocol with bias $\epsilon$ is a process of computations by Alice and Bob in which Alice and Bob compute $a$ and $b$, respectively, by using nonlocal boxes and shared random bits and without any communication such that $a \oplus b = f(x, y)$ with probability $(1 + \epsilon)/2$.

There is a simple XOR protocol with bias $2^{-n}$ for arbitrary function $g$.

Lemma 3. There is XOR protocol with bias $2^{-n}$ for arbitrary function $f(x, y)$ without using nonlocal boxes.

Proof. Let $r \in \{0, 1\}^n$ be shared uniform random bits. Let $a = f(x, r)$. Let $b = 0$ if $r = y$ and $b = r'$ otherwise where $r' \in \{0, 1\}$ is Bob’s private uniform random bit. Then, $a \oplus b = f(x, y)$ with probability $\frac{1}{2} + \frac{1}{2^{n+1}}$.

Definition 4. The nonlocal box is said to be isotropic if

$$\sum_{a, b : a \oplus b = x \land y} p(a, b | x, y)$$

does not depend on $x$ and $y$ and if the marginal distributions for $a$ and $b$ are uniform for any $x$ and $y$.

It was shown in [2][10] that the isotropic nonlocal box can be simulated by arbitrary nonlocal box with the same CHSH probability.

Lemma 5. Using arbitrary given nonlocal box, the isotropic nonlocal box with the same CHSH probability can be simulated.

From Lemma [5] in this paper, we assume that all nonlocal boxes are isotropic. Forster et al. showed that non-isotropic nonlocal boxes can be used for the nonlocality distillation, which is the amplification of the CHSH probability [11]. Brunner and Skrzypczyk showed that there exists non-isotropic nonlocal box with $P_{\text{CHSH}} = 3/4 + \epsilon$ for arbitrary small $\epsilon > 0$ which allows the simulation of nonlocal box arbitrarily close to the PR box [12]. Of course, such nonlocal box cannot be simulated on quantum mechanics even if the CHSH probability of the nonlocal box is achievable by quantum mechanics. In this paper, we do not consider the nonlocality distillation, but consider the XOR protocol using isotropic nonlocal boxes.

The Fourier analysis is the main mathematical tool of this work.

Definition 6. Any boolean function $f : \{+1, -1\}^n \rightarrow \{+1, -1\}$ can be represented by a polynomial on $\mathbb{R}$ uniquely

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \prod_{i \in S} x_i$$

where $[n] := \{1, 2, \ldots, n\}$. Here, $(\hat{f}(S))_{S \subseteq [n]}$ are called the Fourier coefficients of $f$. When we consider the Fourier coefficients of boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, we regard $f$ as the function from $\{+1, -1\}^n$ to $\{+1, -1\}$. From Parseval’s identity, the sum of squares of the Fourier coefficients is 1.

Let $\text{supp}(\hat{f}) := \{ S \subseteq [n] | \hat{f}(S) \neq 0 \}$. For $S \subseteq [n]$, let $1_S$ be a vector on $\mathbb{F}_2$ of length $n$ such that $i$-th element of $1_S$ is 1 iff $i \in S$. Let $\text{dim}(\hat{f})$ be the Fourier dimension of $f$ which is the dimension of linear space on $\mathbb{F}_2$ spanned by $\{1_S | S \in \text{supp}(\hat{f})\}$.

For odd $n$, let $\text{Maj}_n : \{0, 1\}^n \rightarrow \{0, 1\}$ be the majority function on $n$ variables. For even $n$, let $\text{Maj}_n$ be the set of majority functions on $n$ variables where the definitions for the tie cases are arbitrary. Since there are $\left(\binom{n}{2}\right)$ tie cases, $|\text{Maj}_n| = 2^\left\lfloor \frac{n}{2}\right\rfloor$ for even $n$. Note that a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ which ignores one of the $2k$ input variables, and outputs the majority of the other $2k - 1$ variables is a member of $\text{Maj}_{2k}$. Finally, let $\delta := 2P_{\text{CHSH}} - 1$, i.e., $P_{\text{CHSH}} = \frac{3}{4} - \delta$. Here, we call $\delta$ the bias of the CHSH probability.

III. BRASSARD ET AL.’S PROTOCOL

Brassard et al.’s basic idea is bias amplification by $\text{Maj}_3$. They showed that $\text{Maj}_3$ can be computed by using two PR boxes. Here, we give a simple argument showing that two PR boxes are sufficient to compute $\text{Maj}_3(x \oplus y)$. The $\mathbb{F}_2$-polynomial representation of the 3-input majority function is $\text{Maj}_3(z_1, z_2, z_3) = z_1z_2 + z_2z_3 + z_3z_1$. Hence, it holds

$$\text{Maj}_3(x_1 \oplus y_1, x_2 \oplus y_2, x_3 \oplus y_3) = (x_1 \oplus x_2)(y_2 \oplus y_3) + (x_2 \oplus x_3)(y_1 \oplus y_2) + (x_1 \oplus x_2 \oplus x_3 \oplus x_1)(y_1 y_2 y_3)$$

(1)

The following is the protocol for computing $a$ and $b$. Alice and Bob can compute their local terms $A(x) := x_1x_2x_3x_1$ and $B(y) := y_1 y_2 y_3 y_3 y_1$ without communication, respectively. For each of first two terms, they use the PR box. For the first PR box, Alice and Bob input $x_1 \oplus x_2$ and $y_1 \oplus y_2$ and gain $a_1$ and $b_1$, respectively. Similarly, for the second PR box, Alice and Bob input $x_2 \oplus x_3$ and $y_1 \oplus y_2$ and obtain $a_2$ and $b_2$, respectively. The Alice and Bob compute $a := A(x) \oplus a_1 \oplus a_2$ and $b := B(y) \oplus b_1 \oplus b_2$, respectively. This is the XOR protocol without error using two PR boxes. Von Neumann showed that the probability of correctness is amplified by noisy $\text{Maj}_3$ iff the computation of $\text{Maj}_3$ succeeds with probability greater than 5/6 [13]. Hence, the
threshold of the above protocol is given by the condition $P^2_{\text{CHSH}} + (1 - P_{\text{CHSH}})^2 > 5/6 \iff P_{\text{CHSH}} > \frac{\sqrt{5}}{4}$. On this condition, the iterative applications of $\text{Maj}_3$ to independent samples obtained by the protocol in Lemma 3 give a constant bias.

Brassard et al. invented the above elegant protocol, and showed that if $P_{\text{CHSH}} > \frac{\sqrt{5}}{4}$, there exists XOR protocol with constant bias for arbitrary function $f$. However, there is no reason why $\text{Maj}_3$ should be used for the bias amplification. We can use arbitrary functions, e.g., the majority function on 5 variables, in place of $\text{Maj}_3$. Of course the majority functions $\text{Maj}_n$ minimize the threshold value corresponding to 5/6 in the case of $\text{Maj}_3$ on given number $n$ of input variables. However, non-majority function may require smaller number of nonlocal boxes than the majority functions. Hence, non-majority functions are also candidates for the generalization of Brassard et al’s protocol. We have to generalize two quantities “2” and “5/6” in the case of $\text{Maj}_3$. We have to consider the protocol according to $\text{Maj}_3$. In this case, we show that for arbitrary given function $f$, we restrict $f$ to be XOR function, i.e., $f(x,y) = g(x \oplus y)$ for some $g: \{0,1\}^n \rightarrow \{0,1\}$. It seems to be a natural restriction since the inputs $x$ and $y$ have meaning only when their XOR is taken. Linden et al. showed that quantum mechanics has no advantage on XOR protocol for computation of XOR function.

IV. NON-ADAPTIVE PR-CORRECT PROTOCOLS

Brassard et al. consider the protocol according to the $\mathbb{F}_2$-polynomial representation for computing $\text{Maj}_3$. In this section, we show that for arbitrary given function $f: \{0,1\}^n \rightarrow \{0,1\}$, the above protocol is the best protocol for computing $f(x,y)$ among all protocols satisfying the non-adaptivity and the PR-correctness.

Definition 7. A protocol is said to be non-adaptive if inputs for nonlocal boxes does not depend on outputs of other nonlocal boxes. A protocol is said to be PR-correct if the protocol computes the target function $f(x,y)$ without error when the nonlocal boxes are PR boxes. A protocol is said to be non-redundant if the inputs $(l_i(x), r_i(x)) = 1, \ldots, t$ for the nonlocal box satisfy

$$A(x) \oplus B(y) \oplus \bigoplus_{i=1}^t (C_i \land l_i(x) \land r_i(x)) = 0 \iff (C_i)_{i=1}^t = 0, A(x) = B(y).$$

The following lemma was shown by Kaplan et al. \[15\]. Here, we give a short proof using Fourier analysis.

**Lemma 8.** The outputs of both players in non-adaptive PR-correct non-redundant protocol must be parity of the outputs of nonlocal boxes and a function of local inputs.

**Proof.** Let $l_i(x), \ldots, l_t(x)$ and $r_1(y), \ldots, r_t(y)$ be the inputs of nonlocal boxes from Alice and Bob, respectively. Let $a_1, \ldots, a_t$ and $b_1, \ldots, b_t$ be the outputs of the nonlocal boxes for Alice and Bob, respectively. From any protocol, one can obtain a modified protocol using $(l'_i(x) := l_i(x) \lor l_i(0), r'_i(x) := r_i(y) \lor r_i(0)) = 1, \ldots, t$ as the inputs for nonlocal boxes since replacements of $a_i$ and $b_i$ by $a'_i \lor b'_i$ of $r_i(0) \land l_i(0)$ and $l'_i(0) \land r_i(0)$ for $i = 1, \ldots, t$, respectively, simulate the original protocol where $(a'_i, b'_i)_{i=1}^t$ is the outputs of the nonlocal boxes in the modified protocol. This transformation preserves non-adaptivity, PR-correctness and non-redundancy. This transformation also preserves whether or not the outputs of both players are parity of the outputs of nonlocal boxes and a function of local inputs. Hence, without loss of generality, we can assume that $l_1(0) = \cdots = l_t(0) = l_1(0) = \cdots = r_t(0) = 0$. Assume that $a = u_0(a_1, \ldots, a_t)$ and $b = v_0(b_1, \ldots, b_t)$. Since the protocol is PR-correct, $a \lor b = u_0(a_1, \ldots, a_t) \lor v_0(a_1 \land z_i(x, y), \ldots, a_t \land z_t(x, y))$ must be constant for all $(a_1, \ldots, a_t) \in \{0,1\}^t$ where $z_i(x, y) := l_i(x) \lor r_i(y)$. By letting $z = 0$ ($y = 0$), we obtain that $v_y(a) = u_0(v_0)$ or its negation for any $y$ (x), respectively. Hence, there exists boolean functions $F: \{0,1\}^t \rightarrow \{0,1\}$, $\psi: \{0,1\}^n \rightarrow \{0,1\}$ such that $u_x(a_1, \ldots, a_t) = \varphi(x) \lor F(a_1, \ldots, a_t)$ and $v_y(b_1, \ldots, b_t) = \psi(y) \lor F(b_1, \ldots, b_t)$. On the other hand, it holds on the $\{+1, -1\}$ domain that

$$ab = \left(\sum_{S \subseteq [t]} \hat{u}_x(S) \prod_{i \in S} a_i \right) \left(\sum_{S \subseteq [t]} \hat{v}_y(S) \prod_{i \in S} a_i \right) = \sum_{S_1,S_2 \subseteq [t]} \hat{a}_{S_1}(S_1) \hat{b}_{S_2}(S_2) \prod_{i \in S_1 \cup S_2} z_i(x, y) \prod_{i \in (S_1 \cap S_2) - (S_1 \cap S_2) = S} a_i \cdot \prod_{i \in S} a_i$$

(3)

This is the Fourier expansion of $u_x(a_1, \ldots, a_t) \lor v_y(a_1 \land z_i(x, y), \ldots, a_t \land z_t(x, y))$ as a function of $a_1, \ldots, a_t$. Since the function must be constant, the Fourier coefficients for the empty set must be $\pm 1$, i.e.,

$$\sum_{S \subseteq [t]} \varphi(x) \psi(y) \hat{F}(S)^2 \prod_{i \in S} z_i(x, y) \in \{+1, -1\}$$

for any $x, y \in \{0,1\}^n$. Hence, for any $x, y \in \{0,1\}^n$, $\prod_{i \in S} z_i(x, y)$ must be common for all $S \in \text{supp}(\hat{F})$. The equality $\prod_{i \in S} z_i(x, y) = \prod_{i \in S} z_i(x, y)$ for $S_1 \neq S_2$ implies $\prod_{i \in (S_1 \cup S_2) - (S_1 \cap S_2)} z_i(x, y) = 1$ that means the existence of a redundant nonlocal box. Hence, $\hat{F}(S_1) \neq 0$
for unique $S_1 \subseteq [t]$. It implies that $u_x$ and $v_y$ are the parity or the negation of parity on $S_1$. □

Naturally, we can ask whether or not the non-redundancy is restriction, i.e., whether or not we can reduce the error probability of the protocol by using the redundancy when the nonlocal boxes are not the PR boxes. The following lemma says that redundancy does not help to reduce the error probability of non-adaptive PR-correct protocol.

**Lemma 9.** For arbitrary given non-adaptive PR-correct protocol, there exists non-adaptive PR-correct non-redundant protocol whose error probability is at most that of the original protocol for any bias $\delta$ of the CHSH probability.

**Proof.** As in the proof of Lemma 8 we can assume without loss of generality that $l_1(0) = \cdots = l_t(0) = r_1(0) = \cdots = r_t(0) = 0$. Similarly to (8), when the nonlocal boxes are not necessarily the PR boxes, $ab$ is equal to
\[
\sum_{S \subseteq [t]} \left( \sum_{S_1, S_2} \prod_{i \in S_1} (S_{1,i} \cup S_{2,i}) = S \right) \prod_{i \in S_2} a_i \sum_{\{e_i \mid i \in S_1\}} \prod_{i \in S_1} (e_i z_i(x, y))
\]
where $e_i$ represents the error of the output of $i$-th nonlocal box, i.e., $e_i = +1$ if the $i$-th nonlocal box computes correctly and $e_i = -1$ otherwise. Recall that the bias of the CHSH probability is $\delta$, i.e., the expectation of $e_i$ is $\delta$. Since the nonlocal boxes are isotropic, $e_i$ is independent of any other variables $x, y, (a_i)_{i \in [t]}$ and $(r_i)_{i \in [t]}$ for $i \in [t]$. Since the nonlocal boxes are isotropic, $a_i$ is uniformly distributed for all $i \in [t]$. Hence, the expectation of $ab$ (the bias of $a \oplus b$) is
\[
\sigma(x, y) \varphi(x) \psi(y) \sum_{S_1 \subseteq [t]} \widehat{F}(S_1) \hat{F}(S_2) \Pi_{i \in S_2} e_i z_i(x, y)
\]
where $\sigma(x, y)$ denotes the common sign of $\prod_{i \in S_1} z_i(x, y) \in \{+1, -1\}$ for all $S_1 \in \text{supp}(\hat{F})$. Since the protocol is PR-correct, $\sigma(x, y) \varphi(x) \psi(y) \in \{+1, -1\}$ must be equal to $f(x, y)$. Hence, the output of the protocol is correct with probability $(1 + \text{Stab}(F))/2$. On the other hand, since $\prod_{i \in S} z_i(x, y) \in \{+1, -1\}$ is common for all $S \in \text{supp}(\hat{F})$, we can obtain new non-adaptive PR-correct protocol by replacing $u_x(a_1, \ldots, a_t)$ and $v_y(b_1, \ldots, b_t)$ by $\varphi(x) \oplus \bigoplus_{i \in S_1} a_i$ and $\psi(y) \oplus \bigoplus_{i \in S_2} b_i$ for $S^* := \text{argmin}_{S \in \text{supp}(\hat{F})} |S|$, respectively. In order to obtain non-adaptive PR-correct non-redundant protocol, we shrink the set $S^*$ to $T \subseteq S^*$ if $S^*$ includes the redundancy (The local terms $\varphi(x)$ and $\psi(y)$ should also be modified according to the shrinkage). The bias of the probability of correctness of the protocol is $\delta^{[T]} \geq \delta^{[S^*]} \geq \text{Stab}_b(F)$. □

Lemma 9 implies that if we are interested in the minimization of the error probability among all non-adaptive PR-correct non-redundant protocols, we only have to consider non-adaptive PR-correct non-redundant protocols.

**V. THE NUMBER OF NONLOCAL BOXES**

Lemma 9 implies that arbitrary non-adaptive PR-correct non-redundant protocol corresponds to $\mathbb{F}_2$-polynomial representation of $f(x, y)$
\[
f(x, y) = A(x) \oplus B(y) \oplus \sum_{i=1}^t l_i(x) r_i(y).
\]
Since the bias of the correctness of the corresponding protocol is $\delta^*$, we define the following measure of the complexity.

**Definition 10.** The nonlocal box complexity $\text{NLBC}(f)$ is minimum $t$ such that there exists a representation (4).

The nonlocal box complexity can be characterized by the rank of some matrix on $\mathbb{F}_2$. The following theorem slightly generalizes a theorem in [13].

**Theorem 11.** For any $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$,
\[
\text{NLBC}(f) = \text{rank}_{\mathbb{F}_2}(M_f)
\]
where $f'(x, y) = f(x, y) \oplus f(x, 0) \oplus f(0, y) \oplus f(0, 0)$, and where $M_f$ is a $2^n \times 2^n$ matrix on $\mathbb{F}_2$ such that its $(x, y)$ element is equal to $f'(x, y)$.

**Proof.** First, we show $\text{NLBC}(f) \leq \text{rank}_{\mathbb{F}_2}(M_f)$. If $\text{rank}_{\mathbb{F}_2}(M_f) = r$, there is a matrix factorization $M_f = UV$ for some $2^n \times r$ matrix $U$ and $r \times 2^n$ matrix $V$. It implies that $f'(x, y) = \bigoplus_{i=1}^r a_i(x) b_i(y)$ where $a_i(x)$ denotes $(x, i)$ element of $U$ and where $b_i(y)$ denotes $(y, i)$ element of $V$. Hence, it holds $f(x, y) = f(x, 0) \oplus f(0, 0) \oplus f(0, y) \oplus \bigoplus_{i=1}^r a_i(x) b_i(y)$, and hence $\text{NLBC}(f) \leq r$.

Conversely, if $\text{NLBC}(f) = t$, there is a representation $f(x, y) = A(x) \oplus B(y) \oplus \bigoplus_{i=1}^t l_i(x) r_i(y)$. There also exists a representation $f'(x, y) = A'(x) \oplus B'(y) \oplus \bigoplus_{i=1}^t l_i(x) r_i(y)$. Since $f'(0, 0) = f'(x, 0) = 0$ for all $x$ and $y$, by expanding constant terms in $l_i(x)$ and $r_i(x)$, we obtain a representation $f'(x, y) = \bigoplus_{i=1}^t l_i'(x) r_i'(y)$. It implies that there is a matrix factorization $M_f = UV$ for $2^n \times r$ matrix $U$ and $r \times 2^n$ matrix $V$ where $(x, t)$ element of $U$ is $l_i'(x)$ and $(i, t)$ element of $V$ is $r_i'(y)$. Hence, $\text{rank}_{\mathbb{F}_2}(M_f) \leq t$.

Note that if we restrict the decomposition to be symmetric, i.e., $l_i = r_i$ for all $i = 1, \ldots, t$, extra 1 dimension is required [10].

**Lemma 12.** For any $g : \{0, 1\}^n \to \{0, 1\}$, $\text{NLBC}(g^{\oplus 2}) = 0$ only when $g$ is a parity of some variables or its negation. Furthermore, it always holds $\text{NLBC}(g^{\oplus 2}) \neq 1$. □
This is a contradiction.

On the other hand, the diagonal elements of $M_g$ are zero. That implies $M_g^{\oplus} = 0$, and hence $\text{NLBC}(g)$ is 0.

From Theorem 11, it holds $\text{rank}_2(M_g^{\oplus}) = 1$. Since $M_g^{\oplus}$ is a symmetric matrix, there is a decomposition $M_g^{\oplus} = vv^t$ where $v$ denotes a $F_2$-vector of length $2^n$.

On the other hand, the diagonal elements of $M_g$ must be zero. That implies $v = 0$, and hence $\text{NLBC}(g^\oplus) = 0$. This is a contradiction.

**Example 13.** The following table shows the nonlocal box complexity of $\text{Maj}_n$ computed numerically by a computer.

| $n$ | 3 | 5 | 7 | 9 | 11 | 13 | 15 | 17 |
|-----|---|---|---|---|----|----|----|----|
| $\text{NLBC}(\text{Maj}_n)$ | 2 | 14 | 26 | 254 | 494 | 1090 | 1818 | 65534 |

In Example 13 it is not easy to find any rule between $n$ and the nonlocal box complexity although $\text{NLBC}(\text{Maj}_n) = 2^{n-1} - 2$ may happen frequently, e.g., $n = 3, 5, 9, 17$. Generally, it is considered to be difficult to express $\text{rank}_2(M_f)$ in a simple form for arbitrary given $f$. Note that the rank on $\mathbb{R}$ is always at least the rank on $F_2$. Since $\text{rank}_2(M_f)$ is equal to the number of nonzero Fourier coefficients of $g$ [14], $2^{n-1} + 1$ is an upper bound of $\text{NLBC}(\text{Maj}_n)$ for odd $n$ (an inequality $\text{rank}_2(M_f) \leq 2 \cdot \text{NLBC}(f) \leq \text{rank}_2(M_f)$ can be obtained in a similar way as Theorem 11). Here, we introduce a lower bound of the nonlocal box complexity using one-way deterministic communication complexity.

**Theorem 14.** For any $g: \{0, 1\}^n \rightarrow \{0, 1\}$,

$$\text{NLBC}(g^\oplus) \geq \dim(g) - 1.$$  

*Proof.* It holds $\text{NLBC}(f) + 1 \geq D_1(f)$ for arbitrary boolean function $f: \{0, 1\}^n \times \{0, 1\} \rightarrow \{0, 1\}$ where $D_1(f)$ denotes the one-way deterministic communication complexity of $f$ since Bob can compute $f$ from $A(x)$ and $(A_i(x))_{i=1, \ldots, \text{NLBC}(f)}$. In [18], it is shown that $D_1(g^\oplus) = \dim(g)$.

**Example 15.** The Fourier coefficient $\widehat{\text{Maj}_n}(S)$ for the majority function for odd $n$ is nonzero iff $|S|$ is odd. Hence, it holds $\dim(M_n) = n$ and hence, $\text{NLBC}(\text{Maj}_n) \geq n - 1$. From Example 13 this lower bound is tight for $n = 3$, but becomes looser as $n$ increases. This lower bound seems to not be asymptotically tight. However, this lower bound is sufficient to show Theorem 11.

**VI. BIAS AMPLIFICATION**

We now consider the bias amplification by $g^\oplus$ similarly to Brassard et al.’s protocol. If $z$ is a random variable taking +1 with probability $\frac{1}{2}$ and -1 with probability $\frac{1}{2}$, its expectation is $\varepsilon$. The expectation $\varepsilon$ is called the bias of random variable $z$. If the inputs for $g$ is independently and identically distributed and have bias $\varepsilon$, the bias of output of $g$ is given in the following formula.

**Definition 16.** For any $g: \{0, 1\}^n \rightarrow \{0, 1\}$, we define

$$\text{Bias}_n(g) := \sum_{S \subseteq [n]} \hat{g}(S)\varepsilon^{|S|}.$$  

**Example 17.** Since $\text{Maj}_3(z_1, z_2, z_3) = (1/2)(z_1 + z_2 + z_3 - z_1z_2z_3)$, it holds $\text{Bias}_n(\text{Maj}_3) = (3/2)\varepsilon - (1/2)\varepsilon^3$. Roughly speaking, the input bias $\varepsilon$ is amplified to $(3/2)\varepsilon$ for small $\varepsilon$.

When a boolean function $g$ is computed correctly with probability $\frac{1}{2} + \frac{1}{2}$, the output bias of $g$ is $\rho\text{Bias}_n(g)$. We say that the bias is amplified by $g$ if the absolute value of bias of output of $g$ is larger than that of input and if the sign of bias is preserved. The bias is amplified by the noisy $g$ for sufficiently small input bias iff $\text{Bias}_n(\hat{g}(z)) = 0$ and $\rho > \rho_n(g)$ where

$$\rho_n(g) := \frac{1}{\max\{1, \sum_{i=1}^n \hat{g}(i)\}}.$$  

The majority functions minimize $\rho_n(g)$.

**Lemma 19.** For $g: \{0, 1\}^n \rightarrow \{0, 1\}$,

$$\rho_n(g) \geq \frac{2^{n-1}}{n\left(\frac{n+1}{2}\right)}, \quad \text{if } n \text{ is odd},$$

$$\rho_n(g) \geq \frac{2^n}{n\left(\frac{n+1}{2}\right)}, \quad \text{if } n \text{ is even}.$$  

The equality is achieved by and only by the majority functions on $n$ variables. Asymptotically, it holds $\rho_n(g) \geq \sqrt{\pi/(2n)}(1 + O(n^{-1/2}))$.

*Proof.* It holds $\sum_{S \subseteq [n]} \hat{g}(S) = \mathbb{E}[g(x_1 + \cdots + x_n)] \leq \mathbb{E}[|x_1 + \cdots + x_n|]$ where the equality holds only when $g$ is $\text{Maj}_n$. Hence, only the majority functions $\text{Maj}_n$ maximize $\sum_{S \subseteq [n]} \hat{g}(S)$. It is easy to complete the rest of the proof.

Note that the lower bound for even $n$ is equal to the lower bound for $n - 1$. The condition on $\delta$ for the bias amplification by Brassard et al.’s protocol is $\delta^{\text{NLBC}(g^\oplus)} > \rho_n(g)$.

**Definition 20.** For any $g: \{0, 1\}^n \rightarrow \{0, 1\}$,

$$\delta_n(g) := \begin{cases} \rho_n(g)^{\text{NLBC}(g^\oplus)}, & \text{if } \hat{g}(z) = 0 \text{ and } \rho_n(g) < 1 \\ 1, & \text{otherwise} \end{cases}.$$
If $\delta > \delta_B(g)$ for some $g: \{0,1\}^n \to \{0,1\}$, there exists a XOR protocol with constant bias.

**Example 21.** It holds $\delta_B(\text{Maj}_3) = \sqrt{2/3}$ that means that the threshold for the CHSH probability is $\frac{1+\sqrt{2/3}}{2} = \frac{\sqrt{2}}{2}$.

We can now rephrase Theorem 1 in the following form.

**Theorem 22.**

$$\inf_{g: \{0,1\}^n \to \{0,1\}, n \in \mathbb{N}} \delta_B(g) = \frac{\sqrt{2}}{3}.$$  

Furthermore, $\delta_B(g) = \sqrt{2/3}$ iff $g$ is essentially equivalent to $\text{Maj}_3$.

Here, we say that $g$ is essentially equivalent to $\text{Maj}_3$ if $g$ is the majority of some fixed three input variables and ignores the other $n-3$ input variables. First, we show that Theorem 22 holds for $n \leq 4$.

**Lemma 23.** It holds $\delta_B(g) \geq \sqrt{2/3}$ for all boolean functions $g$ on at most 4 variables. Furthermore, for $n \leq 4$, only functions essentially equivalent to $\text{Maj}_3$ satisfy $\delta_B(g) = \sqrt{2/3}$.

**Proof.** From Lemma 12, we only have to consider the case $\text{NLBC}(g^{\oplus}) \geq 2$. From Lemma 19, $\rho_B(g) \geq 2/3$, and hence, $\delta_B(g) \geq \sqrt{2/3}$. From Example 21, it is achieved by $\text{Maj}_3$.

Next, we show the uniqueness. From the above argument, it holds $\delta_B(g) = \sqrt{2/3}$ only when $\rho_B(g) = 2/3$ and $\text{NLBC}(g^{\oplus}) = 2$. From Lemma 19, $\rho_B(g) = 2/3$ only when $g$ is one of the 64 majority functions on 4 variables. In the following, we show that for $g \in \text{Maj}_3$, $\text{NLBC}(g^{\oplus}) = 2$ only when $g$ is essentially equivalent to $\text{Maj}_3$. From Theorem 14, $\{i \in [n] \mid \hat{g}(\{i\}) \neq 0\} \leq \dim(\hat{g}) \leq 3$. If $\{|i \in [n] \mid \hat{g}(\{i\}) \neq 0\| \leq 2$, it holds $\sum_{i \in [n]} \hat{g}(\{i\}) \leq \sqrt{2} < 3/2$ from the Cauchy-Schwarz inequality. If $\{|i \in [n] \mid \hat{g}(\{i\}) \neq 0\| = 3$, $g$ depends only on three variables since $\dim(\hat{g}) \leq 3$. Hence, $g$ is essentially equivalent to $\text{Maj}_3$.

From the following lemma, only boolean functions with small Fourier dimension may outperform $\text{Maj}_3$.

**Lemma 24.** For any $g: \{0,1\}^n \to \{0,1\}$,

$$\delta_B(g) \geq \left(\frac{1}{\dim(\hat{g})}\right)^{\frac{3}{2\dim(\hat{g})-1}}.$$  

In particular, if $\dim(\hat{g}) \geq 5$, it holds $\delta_B(g) > \sqrt{2/3}$.

**Proof.** It holds

$$\dim(\hat{g}) \geq \left|\{i \in [n] \mid \hat{g}(\{i\}) \neq 0\}\right| \geq \left(\frac{\sum_{i \in [n]} \hat{g}(\{i\})^2}{\sum_{i \in [n]} \hat{g}(\{i\})}\right)^2 \geq \left(\sum_{i \in [n]} \hat{g}(\{i\})^2\right)^2.$$  

In the above, the first inequality is trivial. The second inequality is the Cauchy-Schwarz inequality. The third inequality holds since sum of squares of all of the Fourier coefficients is 1. Hence, $\rho_B(g) \geq \dim(\hat{g})^{-1/2}$. From Theorem 14, we obtain this theorem.

Lemmas 23 and 21 give the complete proof of Theorem 22.

**Proof of Theorem 22.** From Lemma 21, we only have to show that if $\{|i \in [n] \mid \hat{g}(\{i\}) \neq 0\| \leq \dim(\hat{g}) \leq 4$, $\delta_B(g) \leq \sqrt{2/3}$ only for $g$ essentially equivalent to $\text{Maj}_3$. Assume $\{|i \in [n] \mid \hat{g}(\{i\}) \neq 0\| = 4$. Then, the boolean function $g$ depends only on 4 input variables since it holds $\dim(\hat{g}) = 4$. From Lemma 23, there is no function on 4 variables satisfying $\delta_B(g) \leq \sqrt{2/3}$ except for functions essentially equivalent to $\text{Maj}_3$. Next, we assume $\{|i \in [n] \mid \hat{g}(\{i\}) \neq 0\| = 3$. In this case, $\sum_{i \in [n]} \hat{g}(\{i\}) \leq \sqrt{3}$. Since $(1/\sqrt{3})^{1/2} > \sqrt{2/3}$, we can assume $\text{NLBC}(g^{\oplus}) \leq 2$. Then, the boolean function $g$ depends only on 3 input variables since $\dim(\hat{g}) \leq \text{NLBC}(g^{\oplus}) + 1 \leq 3$. From Lemma 23, there is no function on 3 variables satisfying $\delta_B(g) \leq \sqrt{2/3}$ except for $\text{Maj}_3$. Next, we assume $\{|i \in [n] \mid \hat{g}(\{i\}) \neq 0\| = 2$. In this case, $\sum_{i \in [n]} \hat{g}(\{i\}) \leq \sqrt{2}$. Since $(1/\sqrt{2})^{1/2} > \sqrt{2/3}$, we can assume $\text{NLBC}(g^{\oplus}) \leq 1$. From Lemma 12, it holds $\delta_B(g) = 1$. We conclude that there is no function satisfying $\delta_B(g) \leq \sqrt{2/3}$ except for functions essentially equivalent to $\text{Maj}_3$.

**VII. CONCLUSION**

In this paper, we show that the 3-input majority function is the unique optimal function for Brassard et al.’s bias amplification. This paper also develops mathematical framework for the problem using Fourier analysis. On the other hand, in this paper, the function $g^{\oplus}$ for the bias amplification is restricted to be XOR function although it seems to be a natural restriction. Furthermore, the computation of functions $g^{\oplus}$, in this paper, is restricted to be non-adaptive and PR-correct. Adaptive protocols may allow more reliable computation than non-adaptive protocols 12. Hence, the result of this paper does not show the limitation of the idea of the bias amplification, but show only the limitation of the idea of the bias amplification by non-adaptive and PR-correct computation of XOR function. Note that the whole protocol is adaptive since an output of function $g^{\oplus}$ will be an input of $g^{\oplus}$. The bias amplification by adaptive computation of non-XOR function would be an interesting direction of research.

**ACKNOWLEDGMENTS**

This work was supported by MEXT KAKENHI Grant Number 24106008.
[1] J. S. Bell, Physics 1, 195 (1964).
[2] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. 23, 880 (1969).
[3] B. S. Cirel’son, Lett. Math. Phys. 4, 93 (1980).
[4] S. Popescu and D. Rohrlich, Found. Phys. 24, 379 (1994).
[5] W. van Dam, “Implausible consequences of superstrong nonlocality,” (2005), arXiv:quant-ph/0501159.
[6] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Phys. Rev. Lett. 96, 250401 (2006).
[7] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Nature 461, 1101 (2009).
[8] R. O’Donnell, Analysis of Boolean Functions (Cambridge University Press, 2014).
[9] L. Masanes, A. Acín, and N. Gisin, Phys. Rev. A 73, 012112 (2006).
[10] A. J. Short, Phys. Rev. Lett. 102, 180502 (2009).
[11] M. Forster, S. Winkler, and S. Wolf, Phys. Rev. Lett. 102, 120401 (2009).
[12] N. Brunner and P. Skrzypczyk, Phys. Rev. Lett. 102, 160403 (2009).
[13] J. von Neumann, Automata Studies, 43 (1956).
[14] N. Linden, S. Popescu, A. J. Short, and A. Winter, Phys. Rev. Lett. 99, 180502 (2007).
[15] M. Kaplan, S. Laplante, I. Kerenidis, and J. Roland, Quantum Inf. Comput. 11, 40 (2011).
[16] A. Lempel, SIAM J. Comput. 4, 175 (1975).
[17] A. Bernasconi and B. Codenotti, IEEE Trans. Comput. 48, 345 (1999).
[18] A. Montanaro and T. Osborne, “On the communication complexity of XOR functions,” (2010), arXiv:0909.3392v2.