Preparing (Supra)Quantum Resources for improved computations using Communication Complexity and Correlation Amplification

Amin Shiraz Gilani and Syed Affan Aslam

April 20, 2018

Abstract

Quantum nonlocality is arguable more general than any other computational mechanism. We present two novel results in this sub-domain of quantum computation. First is a novel communication protocol to simulate a certain type of nonlocal boxes which is optimal over the class of most natural deterministic boxes. Second is a generalization of the proof of optimality of parity, in terms of distillation of nonlocality, over all XOR games.

1 Introduction

A shared non-local no-signalling XOR correlation between n parties of type $a_1 \oplus a_2 \oplus \cdots \oplus a_n = f(x_1, x_2, \cdots x_n)$ can reduce the complexity of certain distributed tasks. It is known that that $n-1$ bits are required to generate an n-partite perfect non-local no-signalling box with any arbitrary correlation. However, the minimal cost for the simulation of noisy correlations is open. Pironio [17] argued that the natural measure of non-locality, Bell Inequality [3] violation, is equivalent to the Communication Complexity of such boxes and hence, the latter also act as the measure of non-locality. In this paper, we answer the question: What is the Communication Complexity of simulating the noisy n-partite natural extension of PR boxes ($a_1 \oplus a_2 \oplus \cdots \oplus a_n = x_1 \cdot x_2 \cdot \cdots \cdot x_n$). We use the linear program setting in the proof where the primal provides the protocol to be followed by the parties and the dual provides the linear expression for Communication Complexity. The dual can also be interpreted as the coefficients of the Bell Inequality that provides the maximal violation. The strategy we introduce here can also be used to compare the non-locality, hence the difficulty, of different n variable Boolean Functions in terms of XOR games.

Finding symmetrical properties of Multiparty No-signalling Polytope has been an important area of study in Quantum Information. This search is also regarded as the search for the "unit" of multiparty correlations — a multiparty counterpart of the bipartite case in which Popescu-Rorhlich correlations happen to be the natural unit. We also know that there are certain "Genuine" Multiparty Correlations. Therefore, for this paper, we will discuss one such property that is the correlation amplification within the Polytope. We show that in the multiparty scenario, any XOR correlation

$$\oplus_{i=1}^{n} a_i = f(\hat{\mathbf{x}})$$

in the non-adaptive space is distilled optimally under parity protocol. The proof can also show that isotropic correlations — correlations with the same error — cannot be amplified for multiparty correlations.

This paper is organized as follows. In section 2, we give the background of all the ideas required to understand the technical contribution of this paper. Section 3 contains detailed problem statement and expands upon the questions we have raised here. Section 4 consists of the work being done primarily related to the questions we tackle while section 5 provides the complete technical results novel to the existing literature. We end with conclusion in section 6. [1]

[1] Section 2 and 4 refers to the literature survey we did.

2 Background

In this section, we give a complete background of all the general concepts and theories required to study the results of this paper. This includes XOR games, communication complexity, no-signalling correlations, amplification of GHZ correlations and linear programs.

2.1 XOR Games

Consider a game involving n spatially separated parties. Each of the parties receive an input bit x_i and, based on some protocol, output a bit a_i. The parties are not allowed to communicate during the game, but can share a resource, including shared randomness, quantum entanglement and superquantum correlations, which they have prepared collectively before the start of the game. Their mutual goal is to satisfy correlations of the following type

$$\oplus_{i=1}^{n} a_i = f(\hat{\mathbf{x}}) \tag{1}$$

where $f(\hat{\mathbf{x}})$ is an n-variable Boolean function.

One trivial representation of such games would be the $2^n \times 2^n$ probability matrix where the rows and columns represent all the possible inputs and outputs respectively and each element is a probability of producing an output given some input.

The bipartite general box is trivially represented as

$$\begin{pmatrix} p_{00|00} & p_{01|00} & p_{10|00} & p_{11|00} \\ p_{00|01} & p_{01|01} & p_{10|01} & p_{11|01} \\ p_{00|10} & p_{01|10} & p_{10|10} & p_{11|10} \\ p_{00|11} & p_{01|11} & p_{10|11} & p_{11|11} \end{pmatrix}$$

However, based on the assumptions we can definitely reduce the representation complexity of XOR games. Since the parity function is applied on the outputs produced by the players, only the parity of the output matters and not the actual outputs itself. This way the column size can be reduced to 2, columns indicate the possible parities of the output (either 0 or 1). Secondly, for the sake of this paper, we assume that the box we are studying is symmetric: that is, bits of parties can be exchanged without affecting the output probability; in other words, symmetric functions only depend on the hamming weight of the input. If we only consider hamming weights for columns instead of the actual inputs, we can represent any symmetric game using $n + 1$ rows, each corresponding to a hamming weight ranging from 0 to n.

Our representation for bipartite symmetric box is as follows.

$$\begin{pmatrix} p_{a_1 \oplus a_2 = 0 | h=0} & p_{a_1 \oplus a_2 = 1 | h=0} \\ p_{a_1 \oplus a_2 = 0 | h=1} & p_{a_1 \oplus a_2 = 1 | h=1} \\ p_{a_1 \oplus a_2 = 0 | h=2} & p_{a_1 \oplus a_2 = 1 | h=2} \end{pmatrix}$$

Example 1. Let us consider the box shared between Alice and Bob, with x and y as their respective inputs and a and b as their respective outputs. Assume that they possess the PR [19] correlations

$$a \oplus b = x \cdot y \tag{2}$$

This shared box is represented by the following matrix

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

where the columns represent the lexicographic outputs, while the row represents lexicographic inputs.

Since the nature is never perfect, it is almost impossible to share perfect correlations, which introduces the need for noise. We can say that the correlations does not adhere to the corresponding rule with ϵ probability. Then, we represent such box by the following matrix

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

PR boxes are extremely significant in terms of understanding nonlocality. Since they provide more powerful correlations than quantum entanglement, they are seen as general version of quantum correlations. This means that whatever is true for PR boxes is also true for quantum entanglement. Actually, the EPR pair, which has the highest quantum correlations, can be viewed as the noisy PR box. Furthermore, these boxes also upholds the no-signalling conditions. That is, these boxes cannot be used for communication, a feature essential for quantum correlations. The pictorial representation of these features can be seen in Figure 1.

These XOR games, even though are more powerful than quantum entanglement, must adhere to the no-signalling principles: a party may not be able to deduce

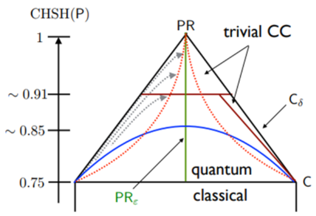

Figure 1: No-signalling Polytope

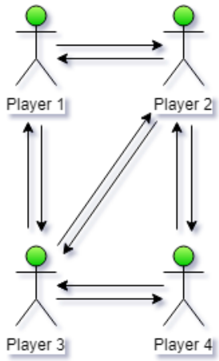

Figure 2: Communication Complexity Model for 4 parties

any information about the input/output of another party. This is formally written for the bipartite case as

$$\sum_a P_{ab|xy} = \sum_a P_{ab|x'y} = P_{b|y}$$

and,

$$\sum_b P_{ab|xy} = \sum_b P_{ab|xy'} = P_{a|x}$$

2.2 Communication Complexity

The communication complexity measure is applied when a task is to be done, collectively, by more than one party. It is defined as the minimum amount of communication required to perform a certain distributed task [13]. It is a measure that is particularly used to compare the complexity of tasks that require parallel computation and sharing of information. The model for two-party (C_I and C_{II}) tasks is shown below:

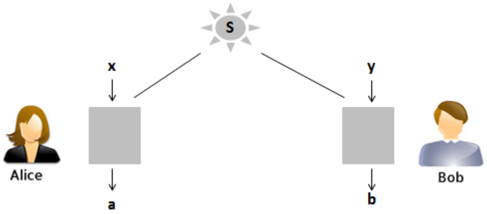

Figure 3: Bipartite Bell's Experiment

In general, we can suppose that the total number of input bits are n and it is shared equally among k parties (n/k bits will be input for each party). Their mutual goal would be to calculate the shared function

$$f(x_1, x_2, \cdots, x_n) \tag{3}$$

with minimum amount of communication among the parties.

2.3 No-signalling Correlations

Correlations are measured using probability distributions on the output of certain measurements conditioned upon certain inputs $P(AB|XY)$ as shown in Figure 3. We play a small three-party game to understand what is Bell's argument, which generally known as the strict upper bound on the correlations attainable classically. Suppose we have three players :Alice, Bob, and Charlie, and each of them have access to one classical bit $0, 1$. The constraints on the game are:

1. their bits $x \oplus y \oplus z = 0$.

2. they cannot communicate with each other.

We instantly know that there are four such possible inputs $000, 011, 101, 110$. There is a uniform distribution on these inputs; that is to say all three players can choose any of the four cases with equal probability.

The **task** for Alice, Bob,and Charlie to output three bits a, b, c such that the following rule is satisfied

$$a \oplus b \oplus c = x \vee y \vee z$$

What is the probability of winning classically? Let's investigate. If Alice and Bob decide to always output 1 and Charlie outputs 0 then $a \oplus b \oplus c$ will be equal to 1. Except for 000, which is 0 when disjunction is taken on it $0 \vee 0 \vee 0$, all the other cases give 1. Therefore this strategy satisfied 3 out of the 4 cases. Thus we say that the probability of winning is $\frac{3}{4}$. The natural question at this point is can we do better than this classically?

Suppose we can do better than this. Let us suppose that $a_x \oplus b_y \oplus c_z$ is the output on input x, y, z. We, therefore, have to satisfy the following conditions

$$\begin{aligned}
a_0 \oplus b_0 \oplus c_0 &= 0 \\
a_0 \oplus b_1 \oplus c_1 &= 1 \\
a_1 \oplus b_0 \oplus c_1 &= 1 \\
a_1 \oplus b_1 \oplus c_0 &= 1
\end{aligned}$$

If we simultaneously take the parity on both sides:

$$\begin{aligned}
(a_0 \oplus a_0) \oplus (a_1 \oplus a_1) \oplus \\
(b_0 \oplus b_0) \oplus (b_1 \oplus b_1) \oplus \\
(c_0 \oplus c_0) \oplus (c_1 \oplus c_1) \oplus \\
= 0 \vee 1 \vee 1 \vee 1
\end{aligned}$$

Because $a_x \oplus a_x = 0$, we have the following contradiction: $0 \neq 1$

This contradiction shows that without signalling the three players cannot perform better 0.75 classically. But as shown in the Figure 1, the quantum bound is $2\sqrt{2}$. Therefore, we formally derive the proofs for classical and quantum bounds.

2.3.1 Bell's Experiment

A Bipartite Bell's Experiment [14], which is also denoted as $(2, 2, 2)$, contains 2 inputs, 2 outputs, and 2 participants Alice and Bob as shown in Figure 3. Two participants who have interacted previously when receive systems from the source S measure their possible inputs on her system and receive outputs. Once the experiments has started, the two participants are not allowed to communicate with each other.

Given such an experimental setup, if Alice and Bob are performing Bell's Experiment "classically" without communicating with each other, to what extent can they their maximize the probability of guessing the outputs correctly?

It should be noted that because the outputs adhere to a particular joint conditional probability distribution $P(AB|XY)$, it is possible to have different outputs with the same input in the second go of an experiment. Like the coin-flip experiments, repeating such an experiment several times can give us certain probabilities for every input and output.

Because the systems may have interacted before, as they were generated from a source S, it is possible, and not because they are communicating with each other (as they are space-like separated), there is some interdependence between them and their joint probability distribution cannot be written as:

$$p(ab|xy) \neq p(a|x)(p|y) \tag{4}$$

We, like Bell, now formulate a classical theory. We introduce a random variable also called $past - factor$ in literature θ. This variable takes into account the past interaction and modify Eq (4) as follows:

$$p(ab|xy, \theta) = p(a|x, \theta)(p|y, \theta) \tag{5}$$

Eq(5) implies that the output a depends upon the input x and a past factor, and not on the space-like separated measurement and outcome of the other party. This construction makes sure that the no-signalling condition is not violated at all.

We assume that the the choice of measurement settings for a party is not dependent upon the past factor. We also know that θ i.e. the past-factor is not constant for all experiments. It is possible that θ itself is dependent upon some physical quantity that is variable. Therefore we can equate Eq (4) with:

$$p(ab|xy) = \int_{\theta} p(a|x, \theta)(p|y, \theta)q(\theta)d\theta \tag{6}$$

where q(θ) is the probability distribution on the values of the continuous random variable θ.

We can see that determinism is the essence of 6. It is assumed that the outcome is determined probabilistically be the measurement setting and input x as well as the past factor $p(a|x, \theta)$ without any influence of the causal effects of the physical system they are in. Bell's theorem is as follows:

Theorem 1. Any classical correlation following the decoupling of (6) is bounded above by 2.

The theorem is crucial because it promises that information-theoretic resources in the classical realm are always going restricted by the above bound .

Proof. First we obtain a bound on the correlations that admit the classical decoupling. For the sake of simplification, say x,y ∈ 0,1 and a,b ∈ +1,-1. We can obtain an expected value on the outcomes given certain input as follows:

$$\langle a_x, b_y \rangle = \sum_{a,b} ab \; p(ab|xy) \tag{7}$$

If $p(ab|xy)$ satisfies the local decomposition mentioned in Eq(6), it can be shown that the following equality arises:

$$M = \langle a_0, b_0 \rangle + \langle a_0, b_1 \rangle + \langle a_1, b_0 \rangle + \langle a_1, b_1 \rangle \leq 2 \tag{8}$$

This inequality is known as $CHSH$ inequality [5]. Following [5], to derive CHSH inequality, we write $p(ab|xy) = \int \langle a_x \rangle_\theta \; \langle b_y \rangle_\theta \; q(\theta) \; d\theta$. Where $\langle a_x \rangle_\theta$ is $\sum_a a \; p(a|x,\theta)$ (similar for $\langle b_y \rangle_\theta$) Given this, We define M_θ as follows:

$$M_\theta = \langle a_0 \rangle_\theta \langle b_0 \rangle_\theta + \langle a_0 \rangle_\theta \langle b_1 \rangle_\theta + \langle a_1 \rangle_\theta \langle b_0 \rangle_\theta - \langle a_1 \rangle_\theta \langle b_1 \rangle_\theta \tag{9}$$

This will help us Eq(8) as follows:

$$M = \int M_\theta \; q(\theta) \; d\theta \tag{10}$$

We also know that $\langle a_0 \rangle_\theta, \langle a_1 \rangle_\theta \in [-1, 1]$. It makes the following inequality greater than Eq(9):

$$|\langle b_0 \rangle_\theta + \langle b_1 \rangle_\theta| - |\langle b_1 \rangle_\theta - \langle b_0 \rangle_\theta| \tag{11}$$

In order to maximize M_θ, without loss of generality, we can obtain $M_\theta = 2|\langle b_0 \rangle_\theta \leq 2$ following Eq(11). Note that we are assuming $|\langle b_0 \rangle_\theta \geq |\langle b_0 \rangle_\theta \geq 0$ then $M \leq 2$ (M as defined in 10)

□

2.4 The Amplification of GHZ correlations

A tripartite input-output system characterized by a conditional probability distribution $P(abc|xyz)$ is *non-signaling* if one cannot communicate from one side to the other sides by the choice of the input. This implies that the marginal probabilities $P(a|x)$, $P(b|y)$, and $P(c|z)$ are independent of the inputs $\{y, z\}$, $\{x, z\}$, and $\{x, y\}$ respectively.

$$\sum_{b,c} P(abc|xyz) = \sum_{b,c} P(abc|xy'z') \equiv P(a|x) \ \forall a, x, y, y', z, z'$$
$$\sum_{a,c} P(abc|xyz) = \sum_{a,c} P(abc|x'yz') \equiv P(a|x) \ \forall a, x, x', y, z, z'$$
$$\sum_{b,c} P(abc|xyz) = \sum_{b,c} P(abc|xy'z') \equiv P(a|x) \ \forall a, x, x', y, y', z$$

In the no-signalling system as shown above for Alice the probability to receive a on input x is independent of what output the other two parties receive. This is one of the constraints that form the no-signalling polytope.

We represent a tripartite system by its probability distribution $P(abc|xyz)$ in matrix notation as

$$\begin{bmatrix} P(000|000) & P(001|000) & P(010|000) & ... & P(111|000) \\ P(000|001) & P(001|001) & P(010|001) & ... & P(111|001) \\ . & & . & & . \\ . & & & . & . \\ . & & & & . \\ P(000|111) & P(001|111) & P(010|111) & ... & P(111|111) \end{bmatrix}.$$

The matrix $\mathbf{p}$, with its rows having indices xyz and columns abc, gives the probability with which Alice,Bob, and Charlie output abc on inputs xyz, respectively. Along with *positivity* and *normalization* constraints, the no-signalling conditions are enforced on $\mathbf{p}$ as shown before. In a general tripartite scenario, the value attained for a strategy $\mathbf{p}$ is given by

$$V(\mathbf{p}) = \sum_{f(a,b,c)=g(x,y,z)} P(abc|xyz) - \sum_{f(a,b,c)\neq g(x,y,z)} P(abc|xyz) \tag{12}$$

The perfect nonlocal tripartite box is characterized to output a uniform distribution over the bits a, b, c on inputs x, y, z such that $f(a, b, c) = g(x, y, z)$.

2.5 CHSH game

In the last section, we looked at conditions that totally, depend or, do not depend on the input of the other party. Here, we analyze a condition, or a game, suggested by Clauser et al [5]. Again, the scenario is the same as above and the condition to be satisfied by both parties is $a \oplus b = x\dot{y}$. It is a bit more subtle than the conditions we looked at earlier. Since, we have demonstrated this as a game, we define the winning value as the average success rate on all of the inputs. We will name this value on the names of the inventors of the game and describe it as follows:

$$CHSH(P) = \frac{1}{4} \sum_{xy} a \oplus b = x\dot{y}$$

where P is our system $P(ab \mid xy)$.

The wining values for each input of this game is given in the table below:

x	y	a ⊕ b
0	0	0
0	1	0
1	0	0
1	1	1

Note that the third column is simply $x\dot{y}$. It tells us the required value of $a \oplus b$ on the given values of x and y for the players to win.

Analyzing it would reveal some intuition about it. We notice that on 3 of the 4 available combinations of x, y, we get $x\dot{y}$ to be 0. This suggests that there exists a trivial strategy that can yield a success rate of 75%. Let us explore a bit. What if both Alice and Bob output their inputs. Then, they will only succeed 1 of the 4 times (when the input is 00). Since we want $\frac{3}{4}$, let us try the negation of our strategy (i.e. when both of them output the negation of their inputs). They still succeed in only one case, which is the same as above. Why is this so? Since negating both the a and b in $a \oplus b$ will negate it twice and give a result similar to the case without negation. If only one of them, let's say Alice, outputs the negation of her input and the other, Bob in this case, outputs his input, then we can succeed on 3 occasions (when the input is 01, 10 and 11). Since we always succeed on 3 of the possible inputs and never succeed on the 4^{th} input, then we have a success rate of 75%.

Can we achieve a success rate of more than 75%? Can we have a strategy that is equivalent in terms of the success percentage to the above one? Note that at the moment, we are only considering classical strategies and we have done nothing quantum yet!! We answer the latter question first.

In fact, the strategy we describe here does not even depend on the input the system is getting. If both the parties output a 0, then $a \oplus b = x\dot{y}$ is false only when $x = y = 1$ and true otherwise. With the above reasoning, we deduce a success rate of 75%.

Our intuition tells us that there must exist a strategy with success rate higher than that of a strategy that does not depend on the input. Classically, it is not true. In other words, there does not exist any deterministic strategy that have a success rate of more than 75%. We now prove that there is an upper bound on the $CHSH$ value of our system.

Theorem 2. Classical bound on CHSH Any classical strategy on the system $P(ab \mid xy)$, restricted by no-signalling constraints, do not yield a $CHSH$

value greater than 0.75.

From the above table of the winning values, we express the winning conditions as:

a_0	$\oplus$	b_0	$=$	0
a_0	$\oplus$	b_1	$=$	0
a_1	$\oplus$	b_0	$=$	0
a_1	$\oplus$	b_1	$=$	1

We prove that the above 4 equations cannot be simultaneously true and at least 1 of them has to be false if all the others are true. If 2 equations are consistent, combining them using a standard binary operator would still yield a consistent system so we combine the above 4 equations using an $\oplus$ operator:

$$\begin{aligned} a_0 \oplus b_0 \oplus a_0 \oplus b_1 \oplus a_1 \oplus b_0 \oplus a_1 \oplus b_1 &= 0 \oplus 0 \oplus 0 \oplus 1 \\ a_0 \oplus a_0 \oplus a_1 \oplus a_1 \oplus b_0 \oplus b_0 \oplus b_1 \oplus b_1 &= 1 \\ 0 \oplus 0 \oplus 0 \oplus 0 &= 1 \\ 0 &= 1 \end{aligned}$$

We see a contradiction here. Hence, our assumption that the system is consistent is false. All the 4 equations cannot be true at the same time and so, at least 1 has to be false. Since at most 3 can be true, the maximum value of $CHSH$ we can have is 0.75.

Up till now, we considered only the deterministic strategies. One would argue that a probabilistic strategy may have the $CHSH$ value greater than what we have got taking only deterministic strategies into consideration. This is not true. Since a probabilistic strategy would be a convex combination of the deterministic strategies, the $CHSH$ value of any probabilistic strategy would be a weighted average of the $CHSH$ vale of the deterministic strategies. If every term in the weighted average is ≤ 0.75 and the weights sum up to 1, the weighted average would also be ≤ 0.75 and thus, the $CHSH$ value of a probabilistic strategy will always be ≤ 0.75.

The theorem above is often considered as the separator of the classical and non-classical domains. In other words, if any system violates this theorem, it does something not possible in the classical realm.

2.6 The quantum advantage

We saw above that classically, the maximum value of $CHSH$ we can get on our system is 0.75. Is it also true quantum mechanically? What if instead of sharing information bits, the parties share a quantum state?

Theorem 3. Quantum possibility of CHSH If a quantum state is allowed to be shared between Alice and Bob, we can attain a $CHSH$ value of $\cos^2(\frac{\pi}{8}) = 0.853$[BCMW09].

Suppose that when Alice and Bob met before their performance of the experiment, they shared the quantum entangled state $\frac{1}{\sqrt(2)}(|00\rangle - |11\rangle)$. During the experiment, they are allowed to perform operations on their qubit but not allowed to alter or measure the other party's qubit.

We saw earlier that the general qubit rotation matrix is $R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. Alice and Bob will apply this matrix to their corresponding qubits, with θ depending on their corresponding input bits, x and y as follows: If their input bit (x for Alice and y for Bob) is 0, they take their corresponding rotation θ to be $-\frac{\pi}{16}$; otherwise they take θ as $\frac{3\pi}{16}$.

For the sake of generality, we will call Alice's rotation θ as θ_a and Bob's rotation θ as θ_b. We calculate the final state of our system by applying $R(\theta_a)$ and $R(\theta_b)$ on Alice's and Bob's qubits respectively:

$$= \frac{1}{\sqrt(2)}(R(\theta_a)|0\rangle R(\theta_b)|0\rangle - R(\theta_a)|1\rangle R(\theta_b)|1\rangle)$$

$$= \frac{1}{\sqrt(2)}\Bigg(\begin{pmatrix} \cos\theta_a & -\sin\theta_a \\ \sin\theta_a & \cos\theta_a \end{pmatrix}|0\rangle \begin{pmatrix} \cos\theta_b & -\sin\theta_b \\ \sin\theta_b & \cos\theta_b \end{pmatrix}|0\rangle$$

$$- \begin{pmatrix} \cos\theta_a & -\sin\theta_a \\ \sin\theta_a & \cos\theta_a \end{pmatrix}|1\rangle \begin{pmatrix} \cos\theta_b & -\sin\theta_b \\ \sin\theta_b & \cos\theta_b \end{pmatrix}|1\rangle\Bigg)$$

$$= \frac{1}{\sqrt(2)}((\cos\theta_a|0\rangle + \sin\theta_a|1\rangle)(\cos\theta_b|0\rangle + \sin\theta_b|1\rangle)$$

$$- (-\sin\theta_a|0\rangle + \cos\theta_a|1\rangle)(-\sin\theta_b|0\rangle + \cos\theta_b|1\rangle))$$

$$= \frac{1}{\sqrt(2)}(\cos\theta_a\cos\theta_b|00\rangle + \cos\theta_a\sin\theta_b|01\rangle + \sin\theta_a\cos\theta_b|10\rangle + \sin\theta_a\sin\theta_b|11\rangle$$

$$- \sin\theta_a\sin\theta_b|00\rangle + \sin\theta_a\cos\theta_b|01\rangle + \cos\theta_a\sin\theta_b|10\rangle - \cos\theta_a\cos\theta_b|11\rangle)$$

$$= \frac{1}{\sqrt(2)}((\cos\theta_a\cos\theta_b - \sin\theta_a\sin\theta_b)|00\rangle + (cos\theta_a\sin\theta_b + \sin\theta_a\cos\theta_b)|01\rangle$$

$$+ (\sin\theta_a\cos\theta_b + \cos\theta_a\sin\theta_b)|10\rangle + (\sin\theta_a\sin\theta_b - \cos\theta_a\cos\theta_b)|11\rangle)$$

$$= \frac{1}{\sqrt(2)}(\cos(\theta_a+\theta_b)|00\rangle + \sin(\theta_a+\theta_b)|01\rangle$$

$$+ \sin(\theta_a+\theta_b)|10\rangle - \cos(\theta_a+\theta_b)|11\rangle)$$

$$= \frac{1}{\sqrt(2)}(\cos(\theta_a+\theta_b)(|00\rangle - |11\rangle) + \sin(\theta_a+\theta_b)(|01\rangle + |10\rangle)$$

Alice and Bob then measure their qubit in the computational standard basis

$(\{|0\rangle, |1\rangle\})$ after getting the above state. We see in the last equation that the probability of success is the square of the coefficient of the first term ($\cos^2(\theta_a + \theta_b)$) when we require $a \oplus b = 0$ and is the square of the coefficient of the second term ($\sin^2(\theta_a + \theta_b)$) when we require $a \oplus b = 1$. The former situation occurs when either x or y is 0 and the latter occurs when both are 1.

Now, we analyze the four possible cases. When the input of the system is 00 ($x = y = 0$), $\theta_a + \theta_b = -\frac{\pi}{8}$ and so, the probability of success is $\cos^2 -\frac{\pi}{8} = 0.853$. Similar result is obtained when one of the inputs is 0 and the other is 1 since $\theta_a + \theta_b = \frac{\pi}{8}$ and therefore, the probability of success is $\cos^2 \frac{\pi}{8} = 0.853$. The sum of the rotation angles of Alice and Bob ($\theta_a + \theta_b$) is $\frac{3\pi}{8}$ when $x = y = 1$ but the success probability is still the same as for other inputs ($\sin^2 \frac{3\pi}{8} = 0.853$). Since the success probability is 0.853 in all the possible cases, the $CHSH$ value of this system is 0.853. Woah! We have got the $CHSH$ value to be greater than it is classically possible.

2.7 Tsirelson's Bound

One would ask that what is the maximum possible value of $CHSH$ quantum mechanically? Boris Tsirelson, in 1980, showed that the value we got above (0.853) is the maximum value we can get [6]. We will prove the same bound as Tsirelson's but by using a different, and slightly intuitive, method [4].

Theorem 4. (Tsirelson's bound on CHSH) If the system $P(ab|xy)$ is restricted by the laws of quantum information, the maximum attainable value of $CHSH$ is $\cos^2 \frac{\pi}{8} = 0.853$.

We claim that any arbitrary strategy of Alice and Bob can be represented by applying projective measurements to four observables (A_0 and A_1 in Alice's possession, B_0 and B_1 in Bob's possession), each having an eigenvalue from $\{+1, -1\}$. It is because any non-projective measurement can be simulated by projective measurements in a larger Hilbert space. Note that the set of eigenvalues of the observables have been transformed form $\{0, 1\}$ by the -1^{input} function.

We call the state shared between Alice and Bob as $|\psi_{ab}\rangle$. Without loss of generality, we assume that when Alice receives 0 as input, she obtains her output bit by applying the projective measurement on the component of $|\psi_{ab}\rangle$ in her possession with respect to the eigenspace of A_0. The observables A_1, B_0 and B_1 are used for similar purposes.

The expected value of the product of Alice's and Bob's projective measurements is $\langle\psi_{ab}| A_x \otimes B_y |\psi_{ab}\rangle$. When $x\dot{y} = 0$, Alice and Bob only win when their projective measurement outcomes are equivalent so the above expression is the losing probability subtracted from the winning probability. When $x\dot{y} = 1$, they win when their measurement outcomes are anti-equivalent so this expression is the winning probability subtracted from the losing probability. Thus, the difference between the winning probability and the losing probability irrespective of the inputs is:

$$\frac{1}{4} \langle \psi_{ab} | \left(A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_1 - A_1 \otimes B_1 \right) | \psi_{ab} \rangle$$

Before proceeding further, let us discuss some facts from Linear Algebra which we will use later. From the definition of an observable, it follows that for any observable X, $X^2 = \mathbb{1}$. Another fact is that $(X_0 \otimes Y_0)(X_1 \otimes Y_1) = X_0 X_1 \otimes Y_0 Y_1$, where X_0, X_1 are observables of an arbitrary Hilbert space A and Y_0, Y_1 are observables of a different Hilbert space B.We now derive an upper bound on the eigenvalue of $A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_1 - A_1 \otimes B_1$.

$$
\begin{aligned}
M &= \frac{1}{4}(A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1) \\
M^2 &= \frac{1}{16}(((A_0 \otimes B_0)(A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1) \\
&\quad + (A_0 \otimes B_1)(A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1) \\
&\quad + (A_1 \otimes B_0)(A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1) \\
&\quad - (A_1 \otimes B_1)(A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - A_1 \otimes B_1)) \\
M^2 &= \frac{1}{16}((A_0 \otimes B_0)(A_0 \otimes B_0) + (A_0 \otimes B_0)(A_0 \otimes B_1) \\
&\quad + (A_0 \otimes B_0)(A_1 \otimes B_0) - (A_0 \otimes B_0)(A_1 \otimes B_1) \\
&\quad + (A_0 \otimes B_1)(A_0 \otimes B_0) + (A_0 \otimes B_1)(A_0 \otimes B_1) \\
&\quad + (A_0 \otimes B_1)(A_1 \otimes B_0) - (A_0 \otimes B_1)(A_1 \otimes B_1) \\
&\quad + (A_1 \otimes B_0)(A_0 \otimes B_0) + (A_1 \otimes B_0)(A_0 \otimes B_1) \\
&\quad + (A_1 \otimes B_0)(A_1 \otimes B_0) - (A_1 \otimes B_0)(A_1 \otimes B_1) \\
&\quad - (A_1 \otimes B_1)(A_0 \otimes B_0) - (A_1 \otimes B_1)(A_0 \otimes B_1) \\
&\quad - (A_1 \otimes B_1)(A_1 \otimes B_0) + (A_1 \otimes B_1)(A_1 \otimes B_1) \\
M^2 &= \frac{1}{16}(A_0A_0 \otimes B_0B_0 + A_0A_0 \otimes B_0B_1 + A_0A_1 \otimes B_0B_0 - A_0A_1 \otimes B_0B_1 \\
&\quad + A_0A_0 \otimes B_1B_0 + A_0A_0 \otimes B_1B_1 + A_0A_1 \otimes B_1B_0 - A_0A_1 \otimes B_1B_1 \\
&\quad + A_1A_0 \otimes B_0B_0 + A_1A_0 \otimes B_0B_1 + A_1A_1 \otimes B_0B_0 - A_1A_1 \otimes B_0B_1 \\
&\quad - A_1A_0 \otimes B_1B_0 - A_1A_0 \otimes B_1B_1 - A_1A_1 \otimes B_1B_0 + A_1A_1 \otimes B_1B_1) \\
M^2 &= \frac{1}{16}(\mathbb{1} \otimes \mathbb{1} + \mathbb{1} \otimes B_0B_1 + A_0A_1 \otimes \mathbb{1} - A_0A_1 \otimes B_0B_1 \\
&\quad + \mathbb{1} \otimes B_1B_0 + \mathbb{1} \otimes \mathbb{1} + A_0A_1 \otimes B_1B_0 - A_0A_1 \otimes \mathbb{1} \\
&\quad + A_1A_0 \otimes \mathbb{1} + A_1A_0 \otimes B_0B_1 + \mathbb{1} \otimes \mathbb{1} - \mathbb{1} \otimes B_0B_1 \\
&\quad - A_1A_0 \otimes B_1B_0 - A_1A_0 \otimes \mathbb{1} - \mathbb{1} \otimes B_1B_0 + \mathbb{1} \otimes \mathbb{1}) \\
M^2 &= \frac{1}{16}(4(\mathbb{1} \otimes \mathbb{1}) - A_0A_1 \otimes B_0B_1 + A_0A_1 \otimes B_1B_0 - A_1A_0 \otimes B_1B_0 + A_1A_0 \otimes B_0B_1)
\end{aligned}
$$

The maximum eigenvalue of M^2 is the sum of the eigenvalues of different terms in M^2 such that the total eigenvalue is maximized. In other words, every term's maximum addition to the total eigenvalue is the absolute value of its coefficient. Thus, the maximum eigenvalue of M^2 is $\frac{1}{16}(4+1+1+1+1) = \frac{1}{2}$. It follows that $\langle\psi_{ab}|\,(M^2)\,|\psi_{ab}\rangle = 1/2$ and so, $\langle\psi_{ab}|\,M\,|\psi_{ab}\rangle = \sqrt{\frac{1}{2}} = \frac{1}{\sqrt{(2)}}$.

The value we got here is the expected value, which is winning probability minus losing probability $(P(W) - P(L) = \frac{1}{\sqrt{(2)}}$. From the definition of losing probability, we know that it is 1 minus the winning probability $(P(L) = 1 - P(W))$. From the above statements, it follows that:

$$\begin{aligned}
P(W) - P(L) &= \frac{1}{\sqrt(2)}\\
P(W) - (1 - P(W)) &= \frac{1}{\sqrt(2)}\\
2P(W) - 1 &= \frac{1}{\sqrt(2)}\\
P(W) &= \frac{1}{2} + \frac{1}{2\sqrt(2)}\\
P(W) &= 0.853
\end{aligned}$$

Hence, the maximum bound on the value of $CHSH$ using quantum information is 0.853. Similar to the classical bound on $CHSH$, it is used for differentiating quantum and non-quantum systems. If some system violates this bound, it means that it cannot be explained by principles of quantum information.

2.8 Linear Programs

A linear program is an optimization problem with a linear objective function and linear (in)equality constraints [21]. It is formally written as

$$\begin{aligned}
\min \quad & \sum_{i=1}^{m} b_i y_i\\
\text{subject to} \quad & \\
& \sum_{i=1}^{m} a_{ij} y_i \geq c_j\\
& y_i \geq 0
\end{aligned} \tag{13}$$

where $i \in [1, n]$, $j \in [1, m]$ and a_{ij}, b_i and c_j are real and y_i are the variables that we want to optimize.

Note that each less-than-or-equal-to and equality constraint can be easily converted into the greater-than-or-equal-to constraint: if $a_{ij} y_i \leq c_j$, then $-a_{ij} y_i \leq -c_j$; if $a_{ij} y_i = c_j$, then $a_{ij} y_i \geq c_j$ and $-a_{ij} y_i \geq -c_j$ suffice.

A set of y_i that fulfills the inequalities is called a feasible solution and the corresponding value of the objective function the feasible value. A feasible solution is called an optimal solution if it minimizes the objective function ($\sum_{i=1}^{m} b_i y_i$). The corresponding value of the objective function is called the optimal value.

An interesting and significant feature of linear programming is duality. It states that for every linear program (called the primal), there exist another linear program (called the dual), which is the converse of the original one, (maximization

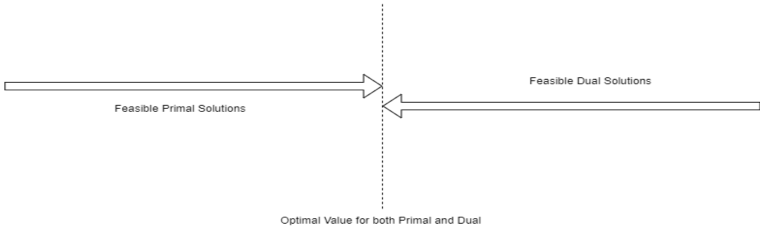
Feasible Primal Solutions
Feasible Dual Solutions
Optimal Value for both Primal and Dual

Figure 4: A Way to Determine the Optimal Solution of a Linear Program

if the original was minimization, and vise versa) that produces the same optimal value as the original one. The dual of the above general minimization problem is

$$\begin{aligned} \max \quad & \sum_{j=1}^{n} c_j x_j \\ \text{subject to} & \\ & \sum_{j=1}^{n} a_{ij} x_j \leq b_i \\ & x_j \geq 0 \end{aligned} \tag{14}$$

where x_i are the variables that we want to optimize and all the other variables are as defined for the primal.

The Weak Duality Theorem states that all the feasible solutions of a minimization primal are greater than or equal to all the feasible solutions of its dual. The Strong Duality Theorem states that if the solutions of the primal and the dual of a linear program are feasible, then, the optimal value of the objective value of the primal and the dual will be equivalent. Together these theorems imply that providing equivalent feasible solutions for the primal and the dual is sufficient to find the optimal solution of the linear program.

3 Problem Statement

From a computational point of view, we see if we have access to imperfect nonlocal correlations with certain noise ϵ, can we simulate them using resources that allow signalling. Secondly, we try to find if we can reduce the bounded error ϵ on the computation of the shared function $f(\hat{\mathbf{x}})$ using the simulated Nonlocal XOR correlation.

The question that we are researching on is: What is the communication complexity of generating an exact nonlocal no-signalling box shared between n parties, having a noise of ϵ and the corresponding rule to be the natural extension of the rule of the PR boxes

$$a_1 \oplus a_2 \oplus \cdots \oplus a_n = x_1 \cdot x_2 \cdot \cdots \cdot x_n.$$

For this particular rule, the n-party nonlocal no-signalling box with ϵ noise is represented, in general, by the following $n+1 \times 2$ matrix:

$$\begin{pmatrix} 1-\epsilon & \epsilon \\ 1-\epsilon & \epsilon \\ \vdots & \vdots \\ 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$$

The players are allowed to use a set of strategies which are categorized in subclasses of DET, LOC_i and SIG_i. Intuitively, the LOC_i and SIG_i classes refer to the local and signalling strategies respectively while the DET strategy is the deterministic strategy. We Define the DET, LOC_i for $i \in \{2, \cdots, n\}$ and SIG_i for $i \in \{1, \cdots, n\}$ strategies as follows:

DET: All the parties choose deterministic bits such that the parity of the output is 0. Formally,

$$\bigoplus_{j=1}^{n} a_j = 0$$

Since it always outputs 0, the box associated with this correlation have probabilities

$$p_{h,b} = \begin{cases} 1 & b = 0 \\ 0 & b = 1 \end{cases}$$

LOC_i: Given an i, the parties choose i input bits $\{x'_1, x'_2, \cdots, x'_i\}$ from the set of input bits $\{x_1, x_2, \cdots, x_n\}$ and produce an output such that

$$\bigoplus_{j=1}^{n} a_j = \begin{cases} \bigoplus_{j=1}^{i} x'_j \oplus 1 & i \text{ is even} \\ \bigoplus_{j=1}^{i} x'_j & i \text{ is odd} \end{cases}$$

Since XOR boxes are used, this correlation can be simulated using shared randomness as a resource only. The parties choose one of the combinations of outputs that satisfy this equation with uniform probability.

The box associated with this correlation has the following probability values (Lemma 11 and 12)

$$p_{h,b} = \begin{cases} \sum_{y \text{ is odd}}^{i} \binom{n-h}{y}\binom{h}{i-y}/\binom{n}{i} & b = 0 \\ \sum_{y \text{ is even}}^{i} \binom{n-h}{y}\binom{h}{i-y}/\binom{n}{i} & b = 1 \end{cases}$$

SIG_i: Given an i, the parties choose i input bits $\{x'_1, x'_2, \cdots, x'_i\}$ from the set of input bits $\{x_1, x_2, \cdots, x_n\}$ and produce an output such that

$$\bigoplus_{j=1}^{n} a_j = \prod_{j=1}^{i} x'_j$$

The above correlation can be simulated perfectly using exactly k-1 bits. (all the x'_j for $j \in \{1, \cdots n-1\}$ bits are send to the party possessing x'_n bit. The parties, again, choose uniformly one of the combinations of outputs that satisfy this equation.

The box associated with this correlation has the following probability values (Lemma 14 and 13)

$$p_{h,b} = \begin{cases} 1 - \binom{h}{i}/\binom{n}{i} & b = 0 \\ \binom{h}{i}/\binom{n}{i} & b = 1 \end{cases}$$

The players are allowed to use any of the local strategies specified (DET or LOC_i) without any cost and signalling (SIG_i) strategies with a cost of $i-1$. Note that the class of strategies used have the minimum error rows in comparison to the other strategies of the same class.

We enumerate all the above stated strategies D and associate with each a communication cost c_{λ_D} of simulating it and the probability p_{λ_D} with which it will be simulated. The resulting simulation must be exactly equivalent to the input box. We incorporate these constraints and the general probability constraints in the linear program below

$$\begin{aligned} \min \sum_{\lambda_D} c_{\lambda_D} p_{\lambda_D} \\ \textsf{subject to } \sum_{\lambda_D} p_{\lambda_D} \mathsf{D} = \mathsf{Box} \\ \sum_{\lambda_D} p_{\lambda_D} = 1 \\ p_{\lambda_D} \geq 0, \end{aligned} \tag{15}$$

As one may notice that the objective function is the communication cost of the simulation of the input box, which is to be minimized. The first constraint implies that the simulated distribution must be equivalent to the box to be simulated. The last two ones belong to the normalization and positivity constraints respectively.

Next, we provide the dual problem of the primal stated above. It is constructed as per the standard method provided in [21].

$$\begin{aligned} &\max \begin{pmatrix} q_{1,1} & q_{1,2} \\ \vdots & \vdots \\ q_{n+1,1} & q_{n+1,2} \end{pmatrix} \bullet \mathsf{Box} \\ &\mathsf{subject\ to} \begin{pmatrix} q_{1,1} & q_{1,2} \\ \vdots & \vdots \\ q_{n+1,1} & q_{n+1,2} \end{pmatrix} \bullet \mathsf{D} \leq c_{\lambda_D} \end{aligned} \tag{16}$$

where the $\bullet$ operation is defined as

$$\begin{pmatrix} u_{1,1} & u_{1,c'} \\ \vdots & \vdots \\ u_{r',1} & u_{r',c'} \end{pmatrix} \bullet \begin{pmatrix} v_{1,1} & v_{1,c'} \\ \vdots & \vdots \\ v_{r',1} & v_{r',c'} \end{pmatrix} = \sum_{x=1}^{r'} \sum_{y=1}^{c'} u_{x,y} v_{x,y}$$

Our approach would be to provide feasible solutions for both the primal and the dual and prove that the objective value corresponding to both of them is equivalent. This value, as we argued above in the implication of the duality theorems, will be the optimal value.

4 Related Work

The question of communication complexity of Boolean functions was first addressed by Andrew Yao [1], with the motivation of studying a domain in VLSI circuits. In the following years, excessive work was done in the field of Communication Complexity [15]. The reasons for its rapid development are [13]: i) it provides general methods for finding lower bounds for problems, which is otherwise extremely rare in Computer Science; ii) it is already known that non-deterministic communication complexity is strictly more powerful than its deterministic counterpart.

Apart from the significance of communication complexity provided above, the communication complexity of nonlocal boxes are used as a measure of nonlocality. The most natural measure of nonlocality is provided by the maximum violation of Bell inequalities: Bell proved that only those correlations are local which satisfy all the Bell inequalities [14] so the violation of these inequalities tells how un-local any given correlation is. However, since this violation does not have any direct relation with information theory, it is not the best information-theoretic measure of nonlocality. Also, a subset of these inequalities may not be able to distinguish between local and non-local correlations.

The first person to realize the limitations of the above measure in the information-theoretic case was Maudlin [22], who, surprisingly, was a Philosopher of Science, and understood the need of quantifying nonlocal correlations explicitly. Brassard, Cleve and Tapp [10] provided a method to simulate quantum entanglement which, as argued above, is a subset of nonlocality. Regev and Toner [20] generalized this result for the d-dimensional quantum states (which means that instead of the usual binary states, the system is allowed to be in a superposition of d states) using a power series method.

Pironio, in his thesis, gave a linear program construction which proved the equivalency of the two measures of nonlocality stated above [17]: maximum Bell inequality violation and communication complexity. The proof was based on the idea that the primal and the dual problems are equivalent. Using the same idea, we can interpret our dual solution as the coefficients of the maximum Bell inequality violation. In the same thesis, he determined the exact nonlocality of bipartite XOR games using the Bell inequality violation measure. Although his proof was general enough to be applied on any arbitrary bipartite distribution, it was not possible to be generalized to the multipartite case, which is the aim of this paper.

Recently, a couple of influential manuscripts has been produced regarding lower bounds for communication complexity in nonlocality simulations: Degorre et al produced a new technique based on affine combinations of lower-complexity distributions, which provided direct lower bounds on quantum and classical communication requirement for simulating non-local distributions [12]; Montina and Wolf provided a lower bound on a different quantity, which they call nonlocal capacity, defined as the minimum communication cost such that the probability distributions refer to entropy [16].

Although these results have been greatly celebrated, they are only valid for bipartite distributions and there does not exist any general method for the multipartite case. Secondly, most of the above methods provides bounds on the problem at stake without providing much insight to how close these are to the actual solution. Fortunately, our method is general enough to take care of the multipartite case and provides exact simulation protocols rather than bounded ones.

As discussed in the Introduction, Bell's Experiment [3] contains two inputs, two outputs, and two players, we attain two outputs. The Bell's Experiment helps us understand bounds on nature. The supremacy of quantum correlations is understood because the results of Bell's Theorem. Following Bell, Tsirelson's bound [6] gives us the bound on the quantum correlations. In [18], the extremal tripartite correlations have been analyzed. This opens space for further understanding of tripartite polytope and it also gives us insights toward multipartite polytope and thus multipartite unit of nonlocality.

In [7] [9] [11] [8], the concept of correlations amplification or distillation has been thoroughly discussed and several results have been reported. In [9] introduced a method to distill correlations. In [2], it is proved that non-local correlations are closed under wiring, which create possibility of distillation without going out of the set of nonlocal correlations. In [11] showes how isotropic correlations are not distillable. We follow the same methodology to prove our results of 5.6.

5 Results

In this section, we present our novel technical contribution. The first 4 sections contain the optimal communication protocol, its feasibility proof of the primal and dual and the equivalency of the predecied feasible solutions for both the programs. The last 2 sections consists of the optimality results of distillation of GHZ boxes and all XOr games.

5.1 Protocol

Given n players and a nonlocal box with error probability ϵ, the parties must choose a strategy from a set of strategies, each with a certain probability depending on ϵ. The parties must follow the following protocol in order to minimize communication cost.

1. Let $k \in \{1, 2, \dots, n\}$ and set N, K, β_k and ϵ_k to the following values,

$$N = 2^{n-1} - 1, \qquad K = 2^{k-1} - 1,$$
$$\beta_k = N\binom{n-1}{k-1} - K\binom{n}{k}, \qquad M_k = 2\beta_k + (N+1)\binom{n-1}{k-1}$$
$$\text{and } \epsilon_k = \frac{\beta_k}{M_k} = \frac{Nk - Kn}{(3N+1)k - 2Kn}.$$

   ϵ_k is a decreasing function of k (Lemma 10).

2. Let k^* be the largest k such that $\epsilon < \epsilon_k$ and determine α such that

$$\epsilon = \alpha\epsilon_{k^*} + (1-\alpha)\epsilon_{k^*+1} \text{ for } \alpha \in [0, 1].$$

3. The parties use the strategies DET, LOC_i and SIG_i with the following probabilities:

$$p_{det} = \epsilon$$

$$p_{loc_i} = \alpha l_{i,k^*} + (1-\alpha) l_{i,k^*+1}$$

where

$$l_{i,k} = \binom{n}{i} \frac{1}{M_k} \sum_{j=2}^{i} \binom{n-j}{n-k-1}$$

is the probability of using the LOC$_i$ strategy at the k^{th} breakpoint.

$$p_{sig_i} = \alpha s_{i,k^*} + (1-\alpha) s_{i,k^*+1}$$

where

$$s_{i,k} = \binom{n}{i} \frac{K+1}{M_k} \, [i = k]$$

is the probability of using the SIG$_i$ strategy at the k^{th} breakpoint.

Notice that SIG$_1$ = LOC$_1$

5.2 Primal

We now show that the protocol we provided in the last section is valid. That is, it satisfies all the contraints of the Primal. It is easy to notice that all the probabilities defined in the protocol are positive and trivially satisfy the positivity constraint.

To prove the satisfaction of the normalization constraint, we notice that we only need to prove it for all the breakpoints because all the other points are a convex combination of the nearest breakpoints. We simply sum over the probabilities (p_i) for all the strategies used.

$$
\begin{aligned}
\sum_i p_i &= p_{det} + \sum_{i=2}^{n} p_{loc_i}\binom{n}{i} + \sum_{x=1}^{n} p_{sig_i}\binom{n}{i} \\
&= \epsilon_{k^*} + \sum_{i=2}^{n} l_{i,k^*}\binom{n}{i} + \sum_{x=1}^{n} s_{i,k^*}\binom{n}{i} \\
&= \epsilon_{k^*} + \frac{1}{M_{k^*}}\left(\sum_{i=2}^{n}\binom{n}{i}\sum_{j=2}^{i}\binom{n-j}{n-k^*-1} + (K^*+1)\binom{n}{k^*}\right) \\
&= \epsilon_{k^*} + \frac{1}{M_{k^*}}\Bigg(\sum_{i=2}^{n}\binom{n}{i}\left(\sum_{j=2}^{k^*+1}\binom{n-j}{n-k^*-1} - \sum_{j=i+1}^{k^*+1}\binom{n-j}{n-k^*-1}\right) \\
&\quad + (K^*+1)\binom{n}{k^*}\Bigg) \\
&= \epsilon_{k^*} + \frac{1}{M_{k^*}}\left(\sum_{i=2}^{n}\binom{n}{i}\left(\binom{n-1}{k^*-1} - \binom{n-i}{k^*-i}\right) + (K^*+1)\binom{n}{k^*}\right) \\
&\quad \text{Using Lemma 3} \\
&= \epsilon_{k^*} + \frac{1}{M_{k^*}}\Bigg(\binom{n-1}{k^*-1}(2N-n+1) - \sum_{i=2}^{n}\binom{n}{i}\binom{n-i}{k^*-i} \\
&\quad + (K^*+1)\binom{n}{k^*}\Bigg) \text{ Using Lemma 1} \\
&= \epsilon_{k^*} + \frac{1}{M_{k^*}}\left((2N-n+1)\binom{n-1}{k^*-1} - \binom{n}{k^*}\sum_{i=2}^{n}\binom{k^*}{i} + (K^*+1)\binom{n}{k^*}\right) \\
&\quad \text{Using Identity 3} \\
&= \epsilon_{k^*} + \frac{1}{M_{k^*}}\Bigg((2N-n+1)\binom{n-1}{k^*-1} - (2K^*-k^*+1)\binom{n}{k^*} \\
&\quad + (K^*+1)\binom{n}{k^*}\Bigg) \text{ Using Lemma 1} \\
&= \frac{1}{M_{k_*}}\left(N\binom{n-1}{k^*-1} - K^*\binom{n}{k^*}\right) + \frac{1}{M_{k^*}}\Bigg((2N-n+1)\binom{n-1}{k^*-1} \\
&\quad - (K^*-k^*)\binom{n}{k^*}\Bigg) \\
&= \frac{1}{M_{k_*}}\left(\left((3N+1)\binom{n-1}{k^*-1} - 2K^*\binom{n}{k^*}\right) - n\binom{n-1}{k^*-1} + k^*\binom{n}{k^*}\right) \\
&= 1 \text{ Using identity 2}
\end{aligned}
$$

It only remains to show that the the convex combination of the strategies provided by our protocols is equivalent to the ϵ noise isotropic box we aim to construct. From Lemma 12 and 13, it is clear that for the LOC_i and SIG_i strategies the probability of outputting a 1 is 1 for the case when the input of all the parties is 1 (hamming weight h of the input is n). This means that all the errors arising in the case of $h = n$ are from the DET strategy. Our protocol dictates the probability of the DET vector to be ϵ which means that when $h = n$, the constructed box outputs a correct value with probability $1 - \epsilon$ which is as per our requirement. We now need to show that for all other values of h, our constructed box produces the correct output with probability $1 - \epsilon$ or with probability equal to the probability of the case when $h = n$. Again, we only prove this for all the breakpoints (ϵ_k) since every other point (ϵ) is the convex combination of the nearest breakpoints. Notice that for the case $h < n$, the output produced by the box is considered an error if and only if the output is 1 as the AND in this case will be 0. Since DET will always output a 0, it does not produce any error. Lemma 12 and 13 dictates the probabilities of the LOC_i and SIG_i to output a 1 and hence output an error. We sum up, for each strategy, the products of the probability with which it is used and the probability with which it produces an error on a hamming weight $h < n$.

$$\sum_{i=2}^{n} \sum_{y \text{ is even}}^{i} loc_{i,k^*} \frac{\binom{n-h}{y}\binom{h}{i-y}}{\binom{n}{i}} + \sum_{i=1}^{n} sig_{i,k^*} \frac{\binom{h}{i}}{\binom{n}{i}}$$

$$= \frac{1}{M_{k^*}} \left(\sum_{i=2}^{n} \sum_{y \text{ is even}}^{i} \binom{n-h}{y}\binom{h}{i-y} \sum_{j=2}^{i} \binom{n-j}{n-k^*-1} + (K^*+1)\binom{h}{k^*} \right)$$

$$= \frac{1}{M_{k^*}} \left(\sum_{i=1}^{n} \sum_{y \text{ is even}}^{i} \binom{n-h}{y}\binom{h}{i-y} \left(\sum_{j=2}^{k^*+1} \binom{n-j}{n-k^*-1} - \sum_{j=i+1}^{k^*+1} \binom{n-j}{n-k^*-1} \right) \right.$$

$$\left. + (K^*+1)\binom{h}{k^*} \right)$$

$$= \frac{1}{M_{k^*}} \left(N\binom{n-1}{k^*-1} - \left(\sum_{i=1}^{n} \binom{n-i}{k-i} \sum_{y \text{ is even}}^{i} \binom{n-h}{y}\binom{h}{i-y} - (K^*+1)\binom{h}{k^*} \right) \right)$$

Using Identity 7 & Lemma 3

$$= \frac{N\binom{n-1}{k^*-1} - K^*\binom{n}{k^*}}{M_{k^*}} \text{ Using Lemma 8}$$

$$= \epsilon_{k^*} \text{ Using Identity 2}$$

5.3 Dual

The dual only involves determining the values of the $q_{i,j}$ such that the Bell inequalities associated with each of the vectors considered are valid. We provide $n-1$ dual solutions, each associated with a linear equation between two consecutive breakpoints in the primal structure. To obtain this, we represent the dual variables as $q_{l,h,b}$ where l represents the l^{th} solution and ranges from 0 to $n-1$. We claim that the following assignment to the dual variables suffices:

$$q_{l,h,b} = \begin{cases} 0 & b = 0 \\ Nf_l - N'g_l & b = 1 \,\&\, h = n \\ d_h g_l - f_l & \text{otherwise} \end{cases}$$

where

$$L = 2^{l-1},$$

$$N = 2^{n-1} - 1,$$

$$N' = 2^{n-2}\left(n-2\right) + 1,$$

$$d_y = n - y - 1,$$

$$D_y = 2^{n-y-1},$$

$$g_l = \frac{2L - l - 1}{D_l\left(L\left(l-1\right)+1\right)},$$

and

$$f_l = \frac{1}{D_l} + \frac{d_l g_l}{2}.$$

Even though the dual takes ϵ as input, it is only used in the objective function and not in inequality constraints. This gives an intuitive justification as to why the dual solution is independent of ϵ.

We now prove that the dual solution presented above is actually a valid solution. Notice that the there is a single constraint applied on all the vectors considered. To prove that the dual solution is valid, we need to show that this inequality constraint is satisfied by all the classes of vectors considered (DET, LOC_i and SIG_i).

For the DET class, the probabilities of the column belonging to $b = 2$ have all entries 0 and hence the left-hand side of the inequality associated with it will produce a 0.

Below, we evaluate the dual expression for the LOC_i class. It is the sum, over each hamming weight h, of the products of the probability of producing a 1 by the relevant LOC_i strategy and the dual coefficient.

$$\begin{aligned}
&q_{l,n,1} + \sum_{h=0}^{n-1} \sum_{z \text{ is even}}^{h} \binom{n-i}{z}\binom{i}{h-z} q_{l,h,1} \\
=&Nf_l - N'g_l + \sum_{h=0}^{n-1} \sum_{z \text{ is even}}^{h} \binom{n-i}{z}\binom{i}{h-z} d_h g_l - f_l \\
=&Nf_l - N'g_l + g_l \sum_{h=0}^{n-1} \sum_{z \text{ is even}}^{h} \binom{n-i}{z}\binom{i}{h-z} d_h \\
&- f_l \sum_{h=0}^{n-1} \sum_{z \text{ is even}}^{h} \binom{n-i}{z}\binom{i}{h-z} \\
\leq&Nf_l - N'g_l + N'g_l - Nf_l \text{ Using Lemma 6 \& 7} \\
=&0
\end{aligned}$$

Similar to LOC_i strategies, we express below the dual expression for the SIG_i strategy and show that the associated constraint is satisfied.

$$\begin{aligned}
&q_{l,n,1} + \sum_{h=i}^{n-1} \binom{n-i}{n-h} q_{l,h,1} \\
=&Nf_l - N'g_l + \sum_{h=i}^{n-1} \binom{n-i}{n-h} d_h g_l - f_l \\
=&Nf_l - N'g_l + \sum_{h=1}^{n-i} \binom{n-i}{h} (h-1)\, g_l - f_l \\
=&Nf_l - N'g_l + g_l \sum_{h=1}^{n-i} \binom{n-i}{h} (h-1) - f_l \sum_{h=1}^{n-i} \binom{n-i}{h} \\
=&Nf_l - N'g_l + N_i'g_l - N_if_l \text{ Using Identity 5 \& Lemma 2} \\
=&(N - N_i)\, f_l - (N' - N_i')\, g_l \\
\leq&i - 1 \text{ Using Lemma 9}
\end{aligned}$$

where

$$N_i = 2^{n-i} - 1,$$

$$N_i' = 2^{n-i-1}\left(n - i - 2\right) + 1,$$

5.4 Equivalency of the Primal and Dual Solutions

We have already noted that the dual solution provides $n-1$ linear equations. All of these predict the behaviour of the communication complexity of the input box for a range of ϵ. On the other hand, the primal solution provides the n breakpoints, each of the $n-1$ consecutive pairs and their signalling and local decomposition point toward a linear equation given by the dual. In this section, we complete the optimality proof of our protocol by giving an argument that shows that the linear equations entailed by the consecutive pairs of breakpoints from the primal is exactly equivalent to the ones given by the dual. Each linear equation is of the form $c_\lambda + m_\lambda \epsilon$. We give separate arguments for the different components of the linear equations (the rate of change of signalling m and the intercept c).

Since all the solutions are linear and only one class of signalling strategies is involved in the decomposition of each breakpoint, we calculate the rate of change of signalling between consecutive breakpoints by simply calculating the gradient of communication complexity between the breakpoints b_λ and $b_{\lambda+1}$. Note that each term in the numerator of the first equation is the amount of communication required to simulate the corresponding breakpoint (product of amount of signalling and the probability with which it is done).

$$
\begin{aligned}
m_\lambda =& \frac{\lambda s_{\lambda+1,\lambda+1} - (\lambda-1)s_{\lambda,\lambda}}{\epsilon_{\lambda+1} - \epsilon_\lambda} \\
=& \frac{M_\lambda \lambda 2^\lambda - M_{\lambda+1}(\lambda-1)2^{\lambda-1}}{M_\lambda((\lambda+1)2^{n-1} - n2^\lambda) - M_{\lambda+1}(\lambda 2^{n-1} - n2^{\lambda-1})} \\
=& \frac{2n(2^\lambda - \lambda - 1) - (\lambda^2+1)(3\cdot 2^{n-1} - 2)}{(\lambda-1)2^{n-1} + 2^{n-\lambda}} \\
=& - q_{\lambda,n,1} - (2^n - 1)f_\lambda + (2^{n-1}(n-2)+1)g_\lambda \\
=& - q_{\lambda,n,1} - \sum_{h=0}^{n-1}\binom{n}{h} f_\lambda + \sum_{h=0}^{n-1}(n-h-1)\binom{n}{h} g_\lambda \\
=& - q_{\lambda,n,1} - \sum_{h=0}^{n-1}\binom{n}{h}(f_\lambda + (n-h-1)g_\lambda) \\
=& \sum_{h=0}^{n} -q_{\lambda,h,1}
\end{aligned}
$$

For the intercept calculation, we extend the line connecting the breakpoints b_λ and $b_{\lambda+1}$ and check where it bisects the y-axis. In other words, we determine the communication of the λ^{th} linear equation at $\epsilon = 0$ if we extend the line connecting ϵ_λ and $\epsilon_{\lambda+1}$.

$$
\begin{aligned}
c_\lambda =& (\lambda-1)s_{\lambda,\lambda} - m_\lambda \epsilon_\lambda \\
=& - M_{\lambda+1}(\lambda-1)2^{\lambda-1} + m_\lambda M_{\lambda+1}(\lambda 2^{n-1} - n2^{\lambda-1}) \\
=& \frac{(\lambda^2+1)(2^{n-1}-1) - n(2^\lambda - \lambda - 1)}{(\lambda-1)2^{n-1} + 2^{n-\lambda}} \\
=& q_{\lambda,n,1}
\end{aligned}
$$

The equivalency of the concerned linear programs provides the statement of our main theorem. We express the communication complexity of the n-partite PR boxes in terms of the intermediate expressions we have for the intersect and the gradient above. Precisely, the expanded expression in the third step of each of the above proofs provides a succinct representation of the problem at stake.

Theorem 1. The communication complexity of simulating a non-local noisy box, with error parameter ϵ satisfying the rule

$$a_1 \oplus a_2 \oplus \ldots \oplus a_n = x_1 \cdot x_2 \cdot \ldots \cdot x_n$$

is given by

$$\max_{\lambda} \frac{(c_\lambda - \epsilon(2c_\lambda + N(\lambda^2+1))}{D}$$

where

$$c_\lambda = (\lambda^2+1)(N-1) - n(2\Lambda - \lambda - 1),$$

$$N = 2^{n-1},$$

$$\Lambda = 2^{\lambda-1},$$

and,

$$D = N\left(\lambda - 1 + \frac{1}{\Lambda}\right)$$

5.5 GHZ Results

The bipartite input-output case is unique as it is the only non-local vertex in the bipartite no-signalling polytope. This guarantees that all bipartite non-local correlations can be simulated by that one non-local box, which is Popescu-Rohrlich box [19], in the following way:

$$P_{nl}(ab|xy) = wP_{a\oplus b=xy}(ab|xy) + (1-w)P_{a\oplus b=xy}(ab|xy). \tag{17}$$

where w has constraints of normalization and positivity.

However, the system becomes completely different if we increase the number of parties.

Pironio *et al.* found 53856 vertices of a tripartite no-signalling polytope that belong to 46 unique classes [18]. For this paper, we are considering a restricted class of tripartite non-local correlations proposed by Greenberger *et al.* called *GHZ* correlations.

We define the *GHZ* non-local box as follows

$$p_{abc|xyz=000} = \begin{cases} 1+\epsilon, & \text{if } a\oplus b\oplus c = 0 \\ 1-\epsilon, & \text{if } a\oplus b\oplus c = 1 \end{cases}$$

$$p_{abc|xyz\neq 000} = \begin{cases} 1+\delta, & \text{if } a\oplus b\oplus c = 1 \\ 1-\delta, & \text{if } a\oplus b\oplus c = 0 \end{cases}$$

such that the input

$$x\oplus y\oplus z = 0$$

That is to say, the three players Alice, Bob, and Charlie respectively take inputs x, y, z from the set $\{000, 011, 101, 110\}$ and output a, b, c following the rule

$$a\oplus b\oplus c = x\vee y\vee z$$

. The matrix $\mathbf{p}$ is the following

$$\frac{1}{8}\begin{pmatrix} 1+\epsilon & 1-\epsilon & 1-\epsilon & 1+\epsilon & 1-\epsilon & 1+\epsilon & 1+\epsilon & 1-\epsilon \\ 1-\delta & 1+\delta & 1+\delta & 1-\delta & 1+\delta & 1-\delta & 1-\delta & 1+\delta \\ 1-\delta & 1+\delta & 1+\delta & 1-\delta & 1+\delta & 1-\delta & 1-\delta & 1+\delta \\ 1-\delta & 1+\delta & 1+\delta & 1-\delta & 1+\delta & 1-\delta & 1-\delta & 1+\delta \end{pmatrix}$$

The non-local value of ths box is

$$V(\mathbf{p}) = 3\delta + \epsilon$$

We were able to find following *seven* unique distillation protocols and their distillation values. The protocols may be different for the input to the second box. Therefore we only show the inputs to the second box:

1. $$V'(\mathbf{p}) = \frac{1}{8}(-\epsilon^2 + 24\delta^2 + 7\delta\epsilon + \delta + \epsilon)$$

   for the following protocol

   $$x_2 = x_1 \vee \overline{a_1} \quad y_2 = y_1 \vee \overline{b_1} \quad z_2 = z_1 \vee \overline{c_1}$$

   as shown in the FIG. 1

2. $$V'(\mathbf{p}) = 3\delta^2 - \epsilon^2$$

   for the following protocol

   $$x_2 = \overline{x_1} \wedge \overline{a_1} \quad y_2 = y \quad z_2 = z$$

3. $$V'(\mathbf{p}) = -\epsilon^2 + \frac{9}{4}\delta^2 - \frac{3}{4}\delta\epsilon$$

   for the following protocol

   $$x_2 = x_1 \wedge a \quad y_2 = y_1 \wedge b \quad z_2 = z_1 \wedge c$$

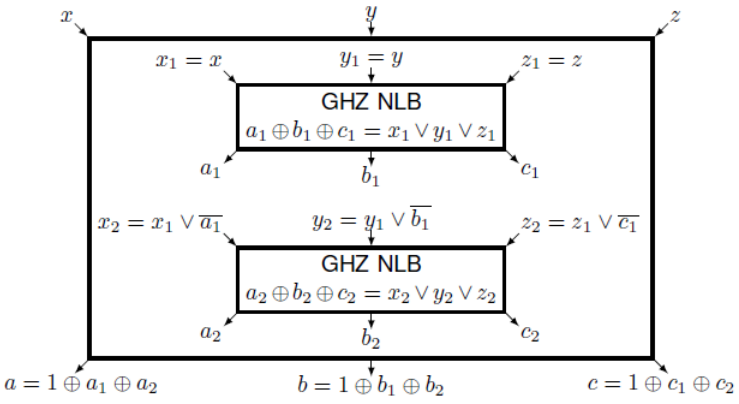
x
y
z
$x_1 = x$
$y_1 = y$
$z_1 = z$
GHZ NLB
$a_1 \oplus b_1 \oplus c_1 = x_1 \vee y_1 \vee z_1$
a_1
b_1
c_1
$x_2 = x_1 \vee \overline{a_1}$
$y_2 = y_1 \vee \overline{b_1}$
$z_2 = z_1 \vee \overline{c_1}$
GHZ NLB
$a_2 \oplus b_2 \oplus c_2 = x_2 \vee y_2 \vee z_2$
a_2
b_2
c_2
$a = 1 \oplus a_1 \oplus a_2$
$b = 1 \oplus b_1 \oplus b_2$
$c = 1 \oplus c_1 \oplus c_2$

Figure 5: A depth 2 GHZ distillation protocol.

4.
$$V'(\mathbf{p}) = -\epsilon^2 + \frac{11}{4}\delta^2 - \frac{1}{4}\delta\epsilon$$
for the following protocol
$$x_2 = x_1 \wedge a \quad y_2 = y_1 \wedge \bar{b} \quad z_2 = z_1$$

5.
$$V'(\mathbf{p}) = -\frac{1}{2}\epsilon^2 + \frac{23}{8}\delta^2 + \frac{3}{8}\delta\epsilon + \frac{1}{8}\delta + \frac{1}{8}\epsilon$$
for the following protocol
$$x_2 = x_1 \wedge a \quad y_2 = y_1 \wedge \bar{b} \quad z_2 = \overline{a} \vee (x \wedge a)$$

6.
$$V'(\mathbf{p}) = -\frac{1}{4}\epsilon^2 + \frac{23}{8}\delta^2 + \frac{5}{8}\delta\epsilon + \frac{1}{8}\delta + \frac{1}{8}\epsilon$$
for the following protocol
$$x_2 = x_1 \wedge a \quad y_2 = a_1 \quad z_2 = \overline{z_1} \wedge c \vee z_1$$

7.
$$V'(\mathbf{p}) = -\frac{1}{4}\epsilon^2 + 3\delta^2 + \frac{3}{4}\delta\epsilon$$
for the following protocol
$$x_2 = x_1 \quad y_2 = y_1 \oplus b_1 \quad z_2 = \overline{z_1} \wedge c \vee z_1$$

We in particular analyze the protocol 1 and the protocol 2 for their outstanding performances as shown in the figure 6.

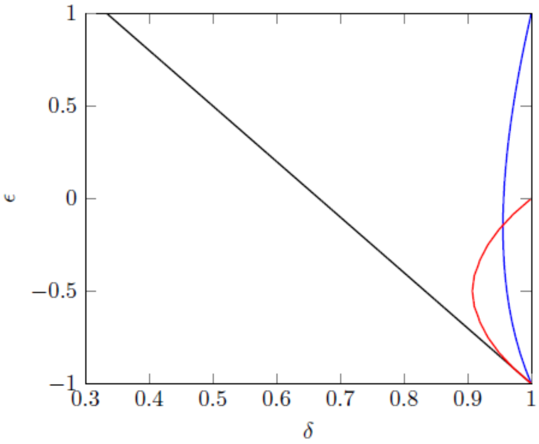

Figure 6: It shows the region of distillable GHZ correlations. The black line corresponds to the boundary between local and non-local region. The blue and red contour curves correspond to distilled region using protocol 1 and protocol 2. As shown, it is clear that the protocol 1 covers more range whereas the protocol 2 has a high magnitude

5.6 Parity is Optimal for XOR Games

In this Section we prove Theorem 1 by first calculating the value achieved by the parity protocol over m NLBs in Lemma 1. This is using the method of Theorem 2 from [9] which can be proved by considering the parity of the number of heads obtained by flipping a coin with bias δ a number of k times. In addition, we need the following Observation that allows us to apply this claim to all XOR games.

Observation 1. Let matrix N be a $2^n \times 2^n$ matrix that follows a rule of the form $\oplus_{i=1}^{m} a_i = f(\hat{\mathbf{x}})$, with the input string $\hat{\mathbf{x}}$ acting as index over the rows, while the output string $\hat{\mathbf{a}}$ indexes the columns. An entry in the matrix is 1 if the rule is satisfied and 0 otherwise. We note that the matrix has one 2 possible unique rows. Therefore, when $f(\hat{\mathbf{x}})$ is equal to 0 for a row, we have one particular distribution on the outputs, otherwise we have the complement of the distribution. The amount of correlation the players are able to produce by a given strategy is measured by the variable V, given by

$$V = \sum_{\bigoplus_{i=1}^{n} a_i = f(\hat{x})} p_{\hat{a}|\hat{x}} - \sum_{\bigoplus_{i=1}^{n} a_i \neq f(\hat{x})} p_{\hat{a}|\hat{x}}. \tag{18}$$

Combining the above observations with Theorem 2 from [9], we obtain that for the parity protocol on m NLBs given by Equation 18, we attain the expected value δ_j^m when $f(\hat{\mathbf{x}}) = 0$ and $-\delta_j^m$ otherwise. This proves Lemma 1 regarding the value of the parity protocol on XOR games.

Lemma 1. The parity protocol over m NLB s of the form in Equation 18, attains the value

$$\sum_{i=1}^{2^n} (-1)^{r_i} \delta_i^m,$$

Finally, we complete the proof for Theorem 2 by proving the value of all non-adaptive protocols is bounded by the value of the parity protocol.

Theorem 2. For m copies of a multipartite no-signalling nonlocal box given by Equation 18, the value of a non-adaptive protocol using at most m NLBs is upper bounded by the value

$$\max_{1 \le k \le m} \left| \sum_{i=1}^{2^n} (-1)^{r_i} \delta_i^k \right|,$$

where r_i is 0 if $f(i)$ evaluates to 0 on input i and 1 otherwise.

Our proof of Theorem 2 follows the proof structure for the bipartite case considered in [11] but requires an additional observation that allows us to generalize the argument to arbitrary XOR games.

Proof. Theorem 2 Let $\hat{\mathbf{a}}_{\mathbf{i}} = (a_i^1, a_i^2, \ldots, a_i^n)$ be the tuple of the n party output from the i^{th} box.The tuple is taken from $\{0,1\}^n$ where for the output string $\hat{\mathbf{a}}_{\mathbf{i}}$ and the input string $\hat{\mathbf{x}}_{\mathbf{i}} = (x_i^1, x_i^2, \ldots, x_i^n)$ the probability distribution is given by Equation 18.

We consider the situation where for initial input $\hat{\mathbf{x}} = (x_1, x_2, \ldots, x_n)$, all n parties output $\hat{\mathbf{0}} = (0, 0, \ldots, 0)$. For a fixed input $\hat{\mathbf{x}}$, we drop the subscript from δ for the corresponding row. Let each A_i in $\hat{\mathbf{A}} = (A_1, A_2, \ldots, A_n)$ represent the set of strings $\{0,1\}^m$ such the players final output is 0. Given that the n players input $\hat{\mathbf{x}}$ into the m NLBs, the probability to output $\hat{\mathbf{a}}$ is

$$\begin{aligned} P(a_1, a_2, \ldots, a_n | x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{m} \left(\frac{1-\delta}{2^n} + \frac{\delta}{2^{n-1}} \left[\oplus_{i=1}^m a_i = 0 \right] \right) \\ &= \frac{1}{2^{mn}} (1-\delta)^{|\oplus_{i=1}^m a_i|} (1-\delta)^{m - |\oplus_{i=1}^m a_i|} \end{aligned}$$

The probability, $Q(\hat{\mathbf{A}})$ of obtaining output $\hat{\mathbf{0}}$ can be obtained by summing over all output bit strings $a_i \in A_i$, i.e.,

$$Q(\hat{\mathbf{A}})(\delta) = \sum_{a_1 \in A_1} \sum_{a_2 \in A_2} \cdots \sum_{a_n \in A_n} \frac{1}{2^{mn}} (1-\delta)^{|\oplus_{i=1}^m a_i|} (1+\delta)^{m - |\oplus_{i=1}^m a_i|}$$

Transforming into the Fourier basis, we obtain

$$\begin{aligned} Q(\hat{\mathbf{A}})(\delta) &= \frac{1}{2^{nm}} \sum_{a_1 \in A_1} \sum_{a_2 \in A_2} \cdots \sum_{a_n \in A_n} \sum_{z \in \{0,1\}^m} \chi_z \left(\oplus_{i=1}^{m} a_i \right) \delta^{|z|} \\ &= \frac{1}{2^{nm}} \sum_{a_1 \in A_1} \sum_{a_2 \in A_2} \cdots \sum_{a_n \in A_n} \sum_{z \in \{0,1\}^m} \prod_{i=1}^{m} \chi_z \left(a_i \right) \delta^{|z|} \\ &= \sum_{z \in \{0,1\}^m} \delta^{|z|} \prod_{i=1}^{m} \left(\sum_{a_i \in A_i} \frac{1}{2^n} \chi_z \left(a_i \right) \right) \\ &= \sum_{z \in \{0,1\}^m} \delta^{|z|} \prod_{i=1}^{m} \sum_{s_i} \frac{1}{2^n} \chi_z \left(s_i \right) \left[s_i \in A_i \right] \end{aligned}$$

We introduce a set of functions $f_1, f_2, \ldots f_n$ that take the value $+1$ if $s_i \in A_i$ and -1 otherwise, to determine

$$\begin{aligned} Q(\hat{\mathbf{A}})(\delta) &= \sum_{z \in \{0,1\}^m} \delta^{|z|} \left(\prod_{i=1}^{n} \sum_{s_i} \frac{1}{2^n} \chi_z \left(s_i \right) \left(\frac{f_i(s_i) + 1}{2} \right) \right) \\ &= \sum_{z \in \{0,1\}^m} \frac{\delta^{|z|}}{2^n} \prod_{i=1}^{n} \left({}^{i}\hat{f}_z + [z = 0] \right), \end{aligned}$$

where ${}^{i}\hat{f}_z$ is the Fourier coefficient. In order to determine $R(\hat{\mathbf{A}})(\delta)$, which is the probability that the n players have even parity output, we now need to compute additional Q probabilities for all the remaining inputs. Let k be a bit string in $\{0,1\}^n$ with even parity. We define the i^{th} element of $\hat{\mathbf{A}}^k$ be given by A_i if $k_i = 0$ and $\overline{A_i}$ if $k_i = 1$, where $\overline{A_i}$ is the complement of A_i, i.e., the set containing bit strings on which player i outputs 1. Here we equate $Q(\hat{\mathbf{A}}^{\mathbf{k}=\hat{\mathbf{0}}})$ to $Q(\hat{\mathbf{A}})$. For a fixed k, the corresponding output probability is given by,

$$\begin{aligned} Q(\hat{\mathbf{A}}^{\mathbf{k}})(\delta) &= \sum_{z \in \{0,1\}^m} \delta^{|z|} \left(\prod_{i=1}^{n} \sum_{s_i} \frac{1}{2^n} \chi_z \left(s_i \right) \left(\frac{(-1)^{k_i} f_i(s_i) + 1}{2} \right) \right) \\ &= \sum_{z \in \{0,1\}^m} \frac{\delta^{|z|}}{2^n} \prod_{i=1}^{n} \left({}^{i}\hat{f}_z (-1)^{k_i} + [z = 0] \right). \end{aligned}$$

The probability $R(\hat{\mathbf{A}})(\delta)$ to output even parity bit string is obtained by summing over all Q's,

$$R(\hat{\mathbf{A}})(\delta) = \sum_{|\mathbf{k}|=0} Q(\hat{\mathbf{A}}^{\mathbf{k}})(\delta).$$

Working out the summation, all the cross terms in the middle terms cancel out leaving us with

$$R(\hat{\mathbf{A}})(\delta) = \frac{1}{2} \left(1 + \sum_{z \in \{0,1\}^m} \prod_{i}^{n} {}^{i}\hat{f}_z \delta^{|z|} \right).$$

The expected value for a fixed δ can then be expressed as

$$R(\hat{\mathbf{A}})(\delta) - (1 - R(\hat{\mathbf{A}})(\delta) = 2R(\hat{\mathbf{A}})(\delta) - 1.$$

This expression can now be used to determine the bound on the value V that any non-adaptive distillation protocol may attain for a NLB, given the biases δ_i.

$$\begin{aligned}
V &= \sum_{i=1}^{2^n} (-1)^{r_i} (2R(\delta_i) - 1)) \\
&= \sum_{z \in \{0,1\}^m} \prod_{i}^{n} {}^{i}\hat{f}_z \left(\sum_{i}^{2^n} (-1)^{r_i} \delta_i^{|z|} \right) \\
&\leq \sum_{z \in \{0,1\}^m} \prod_{i}^{n} \left| {}^{i}\hat{f}_z \right| \left(\left| \sum_{i}^{2^n} (-1)^{r_i} \delta_i^{|z|} \right| \right) \\
&\leq \max_{1 \leq k \leq m} \left| \sum_{i}^{2^n} (-1)^{r_i} \delta_i^{k} \right| \sum_{z \in \{0,1\}^m} \prod_{i}^{n} \left| {}^{i}\hat{f}_z \right| \\
&\leq \max_{1 \leq k \leq m} \left| \sum_{i}^{2^n} (-1)^{r_i} \delta_i^{k} \right|,
\end{aligned}$$

where r_i is 0 when f outputs 0 and 1 otherwise. The last inequality comes from the fact the ${}^{i}\hat{f}_z$ function is normalized and the dot product between the normalized functions will give 1. □

This concludes the proof for Theorem 2

6 Conclusion and Open Questions

This paper provides four significant results in the domain of multipartite nonlocality. Firstly, it identifies a novel structure of the communication complexity of multipartite PR boxes. It was not known in the literature that whether there exist a better protocol than the trivial protocol. We provided a breakpoint-type structure consisting of $n-1$ linear equations in comparison to the single linear equation in the case of the trivial protocol.

Secondly, we provide a novel protocol for the simulation of general PR boxes which is optimal over the set of most natural strategies having coefficients either sum of combinations for the case of local strategies and powers of 2 for the case of signalling strategies.

Thirdly, we show multiple correlation amplification protocols for a GHZ game.

Finally, we show that for any XOR correlation, what is the value attainable by parity protocol. We shown that this value is upper bound on the value attainable through correlation amplification in the non-adaptive space.

A natural open question would be to extend the domain of available strategies to the set of all Boolean functions and then determine the communication structure of mutipartite PR boxes in terms of this extended space. More generally, what is the communication complexity of the simulation of an XOR game associated with any Boolean function?

In the context of XOR correlation amplification, two important questions emerge: Is Parity a Universally Optimal Protocol or do we have a protocol that performs better than parity for a certain correlation with a certain error parameterization? The second question is what would be the optimal protocol for the correlations of the form:

$$f(\mathbf{a}) = g(\mathbf{x}) \tag{19}$$

7 Appendix

7.1 Identities

Identity 1.

$$\binom{n}{k} = \binom{n+1}{k} - \binom{n}{k-1}$$

Identity 2.

$$n\binom{n}{k} = k\binom{n-1}{k-1}$$

Identity 3.

$$\binom{n}{i+a}\binom{i+a}{i} = \binom{n}{i}\binom{n-i}{a}$$

Identity 4.

$$\sum_{j=0}^{a}\binom{n-j}{a-j} = \binom{n+1}{a}$$

Identity 5.

$$\sum_{k=0}^{n}\binom{n}{k} = 2^n$$

Identity 6.

$$\sum_{k=0}^{n} k\binom{n}{k} = 2^{n-1}n$$

Identity 7.

$$\sum_{y=0}^{n} \binom{n-h}{y}\binom{h}{i-y} = \binom{n}{i}$$

Lemma 1.

$$\sum_{i=2}^{n} \binom{n}{k} = 2^{n} - n - 1$$

Proof.

$$\begin{aligned}\sum_{i=2}^{n} \binom{n}{k} =& \sum_{i=0}^{n} \binom{n}{k} - \binom{n}{1} - \binom{n}{0}\\ =& 2^{n} - n - 1\end{aligned}$$

□

Lemma 2.

$$\sum_{k=1}^{n} (k-1)\binom{n}{k} = 2^{n-1}(n-2) + 1$$

Proof.

$$\begin{aligned}\sum_{k=1}^{n} (k-1)\binom{n}{k} =& \sum_{k=0}^{n} (k-1)\binom{n}{k} + 1\\ =& \sum_{k=0}^{n} k\binom{n}{k} - \sum_{k=0}^{n} \binom{n}{k} + 1\\ =& 2^{n-1}n - 2^{n} + 1\\ =& 2^{n-1}(n-2) + 1\end{aligned}$$

□

Lemma 3.

$$\sum_{j=i}^{a} \binom{n-j}{a-j} = \binom{n+1-i}{a-i}$$

Proof. We prove this by induction. The base case is true by **Identity** 4.

Assume that the above equation is true for $i = i'$. We prove it for $i = i' + 1$:

$$\begin{aligned}
\sum_{j=i'+1}^{a} \binom{n-j}{a-j} &= \sum_{j=i'}^{a} \binom{n-j}{a-j} - \binom{n-i'}{a-i'} \\
&= \binom{n+1-i'}{a-i'} - \binom{n-i'}{a-i'} \\
&= \binom{n-i'}{a-i'-1} \text{ Using Identity 1}
\end{aligned}$$

□

Lemma 4.

$$T(n,a,b) = \sum_{y \text{ is even}} \binom{n-a}{y}\binom{a}{b-y}$$

T has the following generating function:

$$T(n,a,b) = \begin{cases} 1 & b = 0 \text{ or } (b = n \text{ and } n-a \text{ is even}) \\ 0 & n = 0 \text{ or } n < b \text{ or } (b = n \text{ and } n-a \text{ is odd}) \\ \binom{n}{b}\delta(b \bmod 2, 0) & a = 0 \\ T(n-1,a-1,b) + T(n-1,a-1,b-1) & \text{otherwise} \end{cases}$$

Proof. The bases cases can be directly verified. The third case can be easily proved by observing that the summation only has a positive term if $b = y$.

We prove the recursive relation (the last case):

$$\begin{aligned}
&T(n-1,a-1,b)+T(n-1,a-1,b-1)\\
=&\sum_{y \text{ is even}}\binom{n-a}{y}\binom{a-1}{b-y}+\sum_{y \text{ is even}}\binom{n-a}{y}\binom{a-1}{b-y-1}\\
=&\sum_{y \text{ is even}}\binom{n-a}{y}\left(\binom{a-1}{b-y}+\binom{a-1}{b-y-1}\right)\\
=&\sum_{y \text{ is even}}\binom{n-a}{y}\binom{a}{b-y} \text{ Using Identity 1}\\
=&T(n,a,b)
\end{aligned}$$

□

Lemma 5.

$$Y_{n,a}=\sum_{b=1}^{n}T(n,a,b)=2^{n-1}-1$$

Proof. We prove the claim by induction. The base cases follows directly from the definition of $T(n,a,b)$. We assume the claim to be true for the tuple $(n-1,a-1)$. We prove it for $(n,\ a)$.

$$\begin{aligned}
Y_{n,a}=&\sum_{b=1}^{n}T(n,a,b)\\
=&\sum_{b=1}^{n}\left(T(n-1,a-1,b)+T(n-1,a-1,b-1)\right) \text{ Using Lemma 2}\\
=&\sum_{b=1}^{n}T(n-1,a-1,b)+\sum_{b=0}^{n-1}T(n-1,a-1,b)\\
=&2\sum_{b=1}^{n-1}T(n-1,a-1,b)+T(n-1,a-1,n)+T(n-1,a-1,0)\\
=&2Y_{n-1,a-1}+1\\
=&2^{n-1}-1
\end{aligned}$$

□

Lemma 6.

$$Y'_{n,a}=\sum_{b=0}^{n-1}T(n,a,b)=\begin{cases}2^{n-1} & n-a \text{ is odd}\\ 2^{n-1}-1 & \text{otherwise}\end{cases}$$

Proof. We use the connection between the definition of $Y_{n,a}$ and $Y'_{n,a}$.

$$\begin{aligned}
Y'_{n,a} =& \sum_{b=0}^{n-1} T(n,a,b) \\
=& \sum_{b=1}^{n} T(n,a,b) + T(n,a,0) - T(n,a,n) \\
=& 2^{n-1} - \begin{cases} 0 & n-a \text{ is odd} \\ 1 & \text{otherwise} \end{cases} \\
=& \begin{cases} 2^{n-1} & n-a \text{ is odd} \\ 2^{n-1}-1 & \text{otherwise} \end{cases}
\end{aligned}$$

□

Lemma 7.

$$Z_{n,a,b} = \sum_{b=0}^{n-1} T(n,a,b)\,(n-b-1) \leq 2^{n-2}\,(n-2)+1$$

Proof. We prove the claim by induction. The base cases follow from the fact that $T(2,0,0) = T(2,1,0) = 1$. We assume the above claim for the tuples $(n-1,$ $a-1,$ $b)$ and $(n-1,$ $a-1,$ $b-1)$ and prove it for the tuple $(n-1,$ $a-1,$ $b)$.

$$
\begin{aligned}
Z_{n,a,b} &= \sum_{b=0}^{n-1} T(n,a,b)\,(n-b-1) \\
&= \sum_{b=0}^{n-1} \left(T(n-1,a-1,b) + T(n-1,a-1,b-1)\right)(n-b-1) \\
&= \sum_{b=0}^{n-2} T(n-1,a-1,b)\,(n-b-1) \\
&\quad + \sum_{b=0}^{n-2} T(n-1,a-1,b-1)\,(n-b-1) \\
&= \sum_{b=0}^{n-2} T(n-1,a-1,b)\,(n-b-2) + \sum_{b=0}^{n-2} T(n-1,a-1,b) \\
&\quad + \sum_{b=0}^{n-2} T(n-1,a-1,b-1)\,(n-b-1) \\
&= \sum_{b=0}^{n-2} T(n-1,a-1,b)\,(n-b-2) \\
&\quad + \sum_{b=0}^{n-2} T(n-1,a-1,b-1)\,(n-b-1) \\
&\quad + \sum_{b=1}^{n-1} T(n-1,a-1,b-1) \\
&\le Z_{n-1,a-1,b} + Z_{n-1,a-1,b-1} + Y_{n-1,a-1} \\
&= 2\left(2^{n-3}(n-3)+1\right) + 2^{n-2} - 1 \\
&= 2^{n-2}(n-3) + 2 + 2^{n-2} - 1 \\
&= 2^{n-2}(n-2) + 1
\end{aligned}
$$

□

Lemma 8.

$$X_{n,k,a} = \sum_{b=1}^{n} \binom{n-b}{k-b} T(n,a,b) = 2^{k-1}\binom{a}{k} + \left(2^{k-1}-1\right)\binom{n}{k}$$

Proof. We prove the claim by induction. The base cases follows directly from the definition of $T(n,a,b)$. We assume the claim to be true for the tuples $(n-1,$ $k,$ $a-1)$ and $(n-1,$ $k-1,$ $a-1)$. We prove it for $(n,$ $k,$ $a)$.

$$\begin{aligned}
X_{n,k,a} =& \sum_{b=1}^{n} \binom{n-b}{k-b} T(n,a,b) \\
=& \sum_{b=1}^{n} \left(\binom{n-b-1}{k-b} + \binom{n-b-1}{k-b-1} \right) \\
& (T(n-1,a-1,b) + T(n-1,a-1,b-1)) \\
& \text{Using Identity 1 and Lemma 2} \\
=& \sum_{b=1}^{n} \left(\binom{n-b-1}{k-b} + \binom{n-b-1}{k-b-1} \right) T(n-1,a-1,b) \\
& + \sum_{b=0}^{n-1} \left(\binom{n-b-2}{k-b-1} + \binom{n-b-2}{k-b-2} \right) T(n-1,a-1,b) \\
=& \sum_{b=1}^{n-1} \binom{n-b-1}{k-b} T(n-1,a-1,b) \\
& + \sum_{b=1}^{n-1} \binom{n-b-1}{k-b-1} T(n-1,a-1,b) \\
& + \binom{n-1}{k-1} + \sum_{b=1}^{n-1} \binom{n-b-1}{k-b-1} T(n-1,a-1,b) \\
& \text{Using Identity 1 and Lemma 2} \\
=& X_{n-1,k,a-1} + 2X_{n-1,k-1,a-1} + \binom{n-1}{k-1} \\
=& 2^{k-1} \binom{a-1}{k} + \left(2^{k-1} - 1\right) \binom{n-1}{k} \\
& + 2^{k-1} \binom{a-1}{k-1} + \left(2^{k-1} - 2\right) \binom{n-1}{k-1} + \binom{n-1}{k-1} \\
=& 2^{k-1} \binom{a}{k} + \left(2^{k-1} - 1\right) \binom{n}{k}
\end{aligned}$$

□

Lemma 9.

$$(N - N_i)\, f_l - (N' - N_i')\, g_l \leq i - 1$$

Proof. Expanding g_l and f_l, we get

$$g_l = \frac{2\left(2^l - l - 1\right)}{2^{n-l}\left((l-1)\, 2^{l-1} + 1\right)}$$

$$f_l = \frac{\left(2^l - l - 1\right)(n-2) + l^2 + 1}{2^{n-l}\left(\left(l-1\right)2^{l-1}+1\right)}$$

Re-writing the inequality in the statement of the Lemma in other terms using the above expansion of g_l and f_l

$$\begin{aligned} W(n,l,i) = {} & (N - N_i)\, f_l - (N' - N_i')\, g_l - i + 1 \leq 0 \\ & = l^2 2^{n-1} - l^2 2^{n-i} - il2^{n-1} + il2^{n-i} + i2^{n-i} - 2^{n-i} - i2^{n-i+l} \\ & \quad + i2^{n-1} - i2^{n-l} + l2^{n-1} + 2^{n-l} \leq 0 \end{aligned}$$

We prove this inequality by induction on n. The base case is simply $W(2,1,0)$ and $W(2,1,1)$ which on evaluation gives 0. Our induction hypothesis is that the claim is true for all values of $W(n-1,l,i)$ $(W(n-1,l,i) \leq 0)$. The inductive step is to prove for all values of $W(n-1,l.i)$.

$$\begin{aligned} & W(n,l,i) \\ & = l^2 2^{n-1} - l^2 2^{n-i} - il2^{n-1} + il2^{n-i} + i2^{n-i} - 2^{n-i} - i2^{n-i+l} \\ & \quad + i2^{n-1} - i2^{n-l} + l2^{n-1} + 2^{n-l} \\ & = 2l^2 2^{n-2} - l^2 2^{n-i-1} - il2^{n-2} + il2^{n-i-1} + i2^{n-i-1} - 2^{n-i-1} \\ & \quad - i2^{n-i+l-1} + i2^{n-2} - i2^{n-l-1} + l2^{n-2} + 2^{n-l-1}) \\ & = 2W(n-1,l,i) \\ & \leq 0 \end{aligned}$$

□

Lemma 10. Assuming constant n, ϵ_k is a decreasing function of k.

Proof. We proof this claim by showing that the derivative of ϵ_k is negative for the range of k. Since ϵ_k is a fraction, applying quotient rule would mean that the denominator of the derivative is a square function and will always be positive. Hence, showing that the numerator of the derivative is negative is sufficient:

$$\begin{aligned}
&\operatorname{sgn}\left(\frac{d\epsilon_k}{dk}\right)\\
&= \operatorname{sgn}\Bigg(\frac{d\left(Nk-Kn\right)}{dk}\left(\left(3N+1\right)k-2Kn\right)\\
&\quad -\left(Nk-Kn\right)\frac{d\left(\left(3N+1\right)k-2Kn\right)}{dk}\Bigg)\\
&= \operatorname{sgn}\Bigg(\frac{d\left(Nk-2^{k-1}n+n\right)}{dk}\left(3Nk+k-2^{k}n+2n\right)\\
&\quad -\left(Nk-2^{k-1}n+n\right)\frac{d\left(3Nk+k-2^{k}n+2n\right)}{dk}\Bigg)\\
&= \operatorname{sgn}(\left(N-\left(\ln 2\right)2^{k-1}n\right)\left(3Nk+k-2^{k}n+2n\right)\\
&\quad -\left(Nk-2^{k-1}n+n\right)\left(3N+1-\left(\ln 2\right)2^{k}n\right))\\
&= \operatorname{sgn}\left(n\left(N+1\right)\left(2^{k-1}-\left(2^{k-1}k\left(\ln 2\right)+1\right)\right)\right)\\
&= \operatorname{sgn}\left(2^{k-1}-\left(2^{k-1}k\left(\ln 2\right)+1\right)\right)\\
&= -1
\end{aligned}$$

The last step is due to the fact that the second term is greater than the first term for all $k \geq 1$. □

Lemma 11. Given an n length string of hamming weight h, applying the LOC_i strategy would produce the output 0 with probability $\sum_{y \text{ is odd}} \binom{n-h}{y}\binom{h}{i-y}/\binom{n}{i}$.

Proof. Since in the LOC_i strategy, we randomly select i parties from n (select i bits form n input bits), the probability of getting a 0 output will be the percentage of $\binom{n}{i}$ selections that output 0.

We write the LOC_i strategy as follows:

$$\text{Output} \begin{cases} \bigoplus_{j=1}^{i} x'_j \oplus 1 & i \text{ is even} \\ \bigoplus_{j=1}^{i} x'_j & i \text{ is odd} \end{cases}$$

Note that the output will only be 0 when from the parties having 0 as an input, odd number of parties are selected in the set of i parties: if i is even and the number of 0 terms in the above expression are odd, the number of 1 terms will be odd so the parity expression will result in a 1 ($\bigoplus_{j=1}^{i} x'_j = 1$). Hence, the output will be $1 \oplus 1 = 0$; the output will be 1 if the number of 0 terms in the above expression are even; similar is the case for odd i.

The numerator expression in the statement of the Lemma follows directly: for each odd $y \leq i$, we choose y bits from $n - h$ 0s and the rest of the $i - y$ bits are selected from h 1s; all these possible distinct combinations are then summed over to give the total number of positive combinations. □

Lemma 12. Given an n length string of hamming weight h, applying the LOC_i strategy would produce the output 1 with probability $\sum_{y \text{ is even}} \binom{n-h}{y}\binom{h}{i-y}/\binom{n}{i}$.

Proof. Since the probability of outputting a 1 is 1 minus the probability of outputting a 0, we have the following expression for the former probability following from the previous Lemma

$$
\begin{aligned}
&1 - \frac{\sum_{y \text{ is odd}} \binom{n-h}{y}\binom{h}{i-y}}{\binom{n}{i}} \\
=&\frac{\binom{n}{i} - \sum_{y \text{ is odd}} \binom{n-h}{y}\binom{h}{i-y}}{\binom{n}{i}} \\
=&\frac{\sum_{y=0}^{n} \binom{n-h}{y}\binom{h}{i-y} - \sum_{y \text{ is odd}} \binom{n-h}{y}\binom{h}{i-y}}{\binom{n}{i}} \\
=&\frac{\sum_{y \text{ is even}} \binom{n-h}{y}\binom{h}{i-y}}{\binom{n}{i}}
\end{aligned}
$$

□

Lemma 13. Given an n length string of hamming weight h, applying the SIG_i strategy would produce the output 1 with probability $\binom{h}{i}/\binom{n}{i}$.

Proof. Since in the SIG_i strategy, we randomly select i parties from n (select i bits form n input bits), the probability of getting a 1 output will be the percentage of $\binom{n}{i}$ selections that output 1.

We write the SIG_i strategy as follows:

$$\text{Output } \prod_{j=1}^{i} x'_j$$

Note that the output will only be 1 if the input bit of all the i selected parties is 1. For a fixed string, this means that all the i bits belong to the h bits which are 1. The numerator expression in the statement of the Lemma follows directly: we choose i parties from h possible parties. □

Lemma 14. Given an n length string of hamming weight h, applying the SIG_i strategy would produce the output 0 with probability $1 - \binom{h}{i}/\binom{n}{i}$.

Proof. Since the probability of outputting a 0 is 1 minus the probability of outputting a 1, the statement of this Lemma follows directly from the previous Lemma. □

References

[1] A.C.Yao. Some complexity questions related to distributive computing. *11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.

[2] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vertési. Closed sets of nonlocal correlations. *Physical Review A*, 80:062107, 2009.

[3] J. S. Bell. On the Einstein–Podolsky–Rosen paradox. *Physics*, 1:195, 1964.

[4] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86:419–478, 2014.

[5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden–variable theories. *Physical Review Letters*, 23:880, 1969.

[6] B. S. Csirel'son. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4:93–100, 1980.

[7] D. Dukaric. Nonlocality distillation is impossible for isotropic quantum systems. 2011.

[8] D. Dukaric and S. Wolf. A limit on nonlocality distillation. 2008.

[9] M. Forster. Bounds for nonlocality distillation protocols. *Physical Review A*, 83:062114, 2011.

[10] R. Cleve G. Brassard and A. Tapp. Cost of exactly simulating quantum entanglement with classical communication. *Phys. Rev. Lett*, 83:1874–1877, 1999.

[11] P. Høyer and J. Rashid. Optimal protocols for nonlocality distillation. *Physical Review A*, 82:042118, 2010.

[12] S. Laplante J. Degorre, M. Kaplan and J. Roland. The communication complexity of nonsignaling distributions. *Quantum Information and Computation*, 11:649–676, 2011.

[13] J.Hromkovic. *Communication Complexity and Parallel Computing.* Springer, 1997.

[14] J.S.Bell. On the einstein podolsky rosen paradox. *Physics*, 1:195, 1964.

[15] E. Kushilevitz and N. Nisan. *Communication Complexity.* Cambridge University Press, New York, 1997.

[16] A. Montina and S. Wolf. Information-based measure of nonlocality. *New J. Phys*, 18:013035, 2016.

[17] S. Pironio. Violations of bell inequalities as lower bounds on the communication cost of nonlocal correlations. *Phys. Rev. A*, 68(6):062102, 2003.

[18] S. Pironio, J. D. Bancal, and V. Scarani. Extremal correlations of the tripartite no-signaling polytope. *Journal of Physics A: Mathematical and Theoretical*, 44(6):065303, 2011.

[19] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3), 1994.

[20] O. Regev and B. Toner. Simulating quantum correlations with finite communication. *SIAM J. Comput*, 39(4):1562, 2010.

[21] S.Boyd and L.Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

[22] T.Maudlin. Bell's inequality, information transmission, and prism models. *In Biennal Meeting of the Philosophy of Science Association*, pages 404–417, 1992.