Parallel repetition of local simultaneous state discrimination

Llorenç Escolà Farràs[1,2], Jaròn Has[3], Maris Ozols[1,3,4], Christian Schaffner[1,2], and Mehrdad Tahmasbi[5]

[1] *QuSoft, Amsterdam, The Netherlands*
[2] *Informatics Institute, University of Amsterdam, The Netherlands*
[3] *Korteweg-de Vries Institute (KdVI), University of Amsterdam, The Netherlands*
[4] *Institute for Logic, Language, and Computation (ILLC), University of Amsterdam, The Netherlands*
[5] *Tufts University, Medford, MA, USA*

Local simultaneous state discrimination (LSSD) is a recently introduced problem in quantum information processing. Its classical version is a non-local game played by non-communicating players against a referee. Based on a known probability distribution, the referee generates one input for each of the players and keeps one secret value. The players have to guess the referee's value and win if they all do so. For this game, we investigate the advantage of no-signalling strategies over classical ones. We show numerically that for three players and binary values, no-signalling strategies cannot provide any improvement over classical ones. For a certain LSSD game based on a binary symmetric channel, we show that no-signalling strategies are strictly better when multiple simultaneous instances of the game are played. Good classical strategies for this game can be defined by codes, and good no-signalling strategies by list-decoding schemes. We expand this example game to a class of games defined by an arbitrary channel, and extend the idea of using codes and list decoding to define strategies for multiple simultaneous instances of these games. Finally, we give an expression for the limit of the exponent of the classical winning probability, and show that no-signalling strategies based on list-decoding schemes achieve this limit.

Contents

1 Introduction

The task of discriminating between states is of fundamental importance in information processing and cryptography [1, 2, 3]. A rich and extensive literature exists on this fundamental problem under the name of state discrimination or hypothesis testing [4, 5, 6]. In quantum cryptography and quantum information theory, a natural extension of state-discrimination problem is to distinguish *quantum states*. In the context of non-local games, the state-discrimination problem arises in a multi-player setting. In these scenarios, it is interesting to study how non-local resources such as shared randomness, quantum entanglement or no-signaling correlations can help the players to succeed in the state-discrimination task. Authors of [7, 8] have studied the scenario where local operation and classical communication are allowed between two parties, and they have shown that entanglement can help the players.

The authors of [9] studied another variant of distributed state discrimination in which multiple parties cannot communicate and have to estimate the state locally and simultaneously, hence calling the problem *local simultaneous state discrimination* (LSSD). LSSD problems naturally arise in the context of uncloneable cryptography [10, 11, 12, 13], where we encode classical data into a quantum state such that an adversary cannot copy it. In such scenarios, successfully copying translates into successfully distinguishing quantum states. LSSD problems also appear in the study of monogamy of entanglement games [14], where two parties prepare a tripartite state and perform a measurement to guess the outcome of a measurement performed by a third party. Optimal performance of such games has been crucial to prove the security of uncloneable cryptographic schemes [10]. Depending on the resources shared between the parties, one can consider various strategies. The authors of [9] showed that even when the state has a classical description, quantum entanglement could enhance the probability of simultaneous state discrimination, and a more powerful resource of no-signaling correlations could enhance it even further.

As [9] have shown that finding the optimal strategy for three-party LSSD is NP-hard, it is likely to be challenging to study LSSDs in general. One could, however, characterize the optimal probability of winning and optimal strategies for LSSDs with some specific structure. One natural structure of interest is when an LSSD problem consists of several independent and identical LSSDs, and the parties have to win all these games at once in parallel. We call this type of LSSDs *parallel repetition* of LSSDs, for which we establish several results in this article. Studying parallel LSSD games might have cryptographic implications. Many protocols have product structures, and if we restrict the adversaries only to applying a "product" attack, then the performance of such protocols is governed by parallel repetition of LSSDs. Furthermore, the monogamy of entanglement games with product structure have been important to understand. If we restrict the strategies to those with product states, then the problem can be formulated in terms of parallel repetition of LSSD games.

1.1 Our contributions

As a first simple observation, we show in Theorem 3.1 that for symmetric LSSD problems with classical inputs (as depicted in Fig. 2), there exists an optimal symmetric strategy. In other words, for an LSSD problem defined by a joint distribution P_{XAB} such that P_{X} is uniform over $\mathscr{X}$, $P_{\mathsf{AB|X}} = P_{\mathsf{A|X}} P_{\mathsf{B|X}}$, and $P_{\mathsf{A|X}} = P_{\mathsf{B|X}}$, there exist optimal classical deterministic strategies for Alice and Bob that are identical.

In Section 4 we analyze an example of an LSSD game introduced in [9], where the referee sends a bit x over a *binary symmetric channel* (BSC), see Fig. 3, to Alice and Bob. We use the symmetry observation above to find optimal classical strategies for two and three parallel repetitions of this game in Theorems 4.3 and 4.5, respectively. We also give optimal no-signalling strategies for two and three copies (our results for two copies are depicted in Fig. 1). Finally, in Section 4.3, we consider the n-fold parallel repetition of this game, and argue how the classical strategies relate to (regular) error-correcting codes and the no-signaling strategies relate to list-decoding schemes.

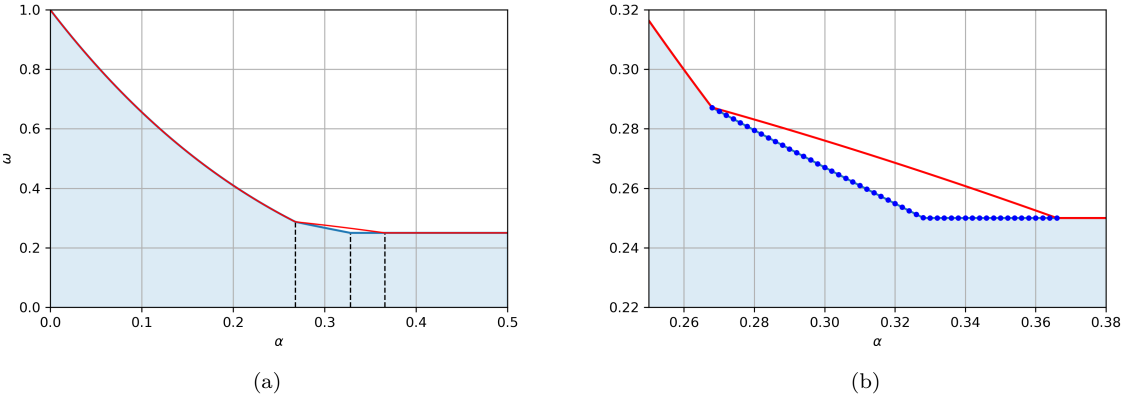

Figure 1: (a) Optimal classical (blue) and no-signalling (red) winning probabilities for the two-fold parallel repetition of the BSC game. The light blue area represents the values below the optimal classical winning probability. (b) Closeup of (a) with an additional numerical upper bound on the optimal quantum winning probability (blue dots) from the level "$1 + NM$" of the NPA hierarchy for the values of α where the classical and no-signalling values differ. The numerical quantum upper bound is in excellent agreement with the classical value, suggesting its optimality (see Conjecture 4.4).

In Section 5 we introduce the notion of channel games, which are an extension of the LSSD problem in Section 4. We then define classical strategies based on codes and no-signalling strategies based on list-decoding schemes. In Theorem 5.2 we provide an expression for the limit of the exponent of the classical winning probability, where we make use of strategies based on codes. Furthermore, we show that no-signalling strategies based on list-decoding schemes achieve the same limit as classical strategies. As a result, the optimal probability of winning for that class of LSSD games is asymptotically the same for all three types of resources available to the players. This allows one to solve the optimization problem for no-signalling strategies, for which there is an efficient algorithm, to find the asymptotic classical or quantum value which are otherwise computationally expensive to evaluate.

As an extension, in Appendix C we analyze three-party LSSD problems with *binary* inputs and outputs. Lemma C.2 extends the two-party characterization from [9] for the classical winning probability of binary LSSD games to three parties. The main result of this appendix, Theorem C.6, shows that no-signalling resources cannot improve the winning probability of the players in this setting.

1.2 Open problems

It would be interesting to examine the settings when there is a gap between the no-signalling and classical winning probabilities in the BSC game. Wherever there is a gap, it is interesting to look for a quantum strategy that also performs better than classical.

In the context of channel games, as introduced in Section 5, can we show, like for the BSC, that no-signalling strategies based on list-decoding schemes are asymptotically optimal? Are there more examples of channels for which there is a gap in winning probability between classical and no-signalling strategies in a finite number of parallel repetitions? Can the results be extended to classical-quantum channels where Alice and Bob receive a quantum state? For this last question, we would need to extend the idea of no-signalling to the case where the inputs and outputs can be quantum states.

Section 5 also gives rise to a new area within information theory: simultaneous decoding. Within this setting, a sender tries to send a message to two receivers using identical channels, and the communication is successful if both receivers decode correctly. We can allow the receivers to share some quantum or no-signalling resources and examine whether this leads to better coding schemes. There are similar settings that have already been researched. In one such setting, the messages sent to the receivers are not necessarily the same, or two different channels are used (like in the book of El Gamal and Kim [15, Part 2]). In another similar setting, we allow the sender and the receiver to share some entanglement (like in the book by Holevo [16, Section 9]). There is even very recent research in a setting with two senders and one receiver that all share a no-signalling box (see the paper by Fawzi and Fermé [17]).

For the case of multi-player LSSD games with binary inputs and outputs as (see Appendix C), it is an open problem whether this result holds for any number of players. However, extending our numerical analysis to a larger number of players requires enumerating over all extrema of the corresponding no-signalling polytope. This polytope quickly grows in the number of vertices, making the analysis infeasible at the moment.

2 Preliminaries

For $n \in \mathbb{N}$, we denote the set $\{0, \dots, n-1\}$ by $[n]$ and the set of all permutations of $[n]$ by S_n. We denote by δ the indicator function, which is 1 if its argument is true and 0 otherwise. Throughout, we use binary logarithms and denote them by log rather than $\log_2$. We denote the bitwise XOR operator on bitstrings by $\oplus$ and the all-zero and all-one bitstrings of length n by 0^n and 1^n, respectively. Let X be a random variable over a finite set $\mathscr{X}$. We denote its probability distribution by P_{X} where X is used to label the register that stores the random variable X. For any $n \geq 1$, we denote by $P_{\mathsf{X}}^{\times n} = (P_{\mathsf{X}})^{\times n}$ the product distribution of n copies of X on $\mathscr{X}^n := \mathscr{X} \times \cdots \times \mathscr{X}$ defined by

$$P_{\mathsf{X}}^{\times n}(x^n) := \prod_{i=1}^{n} P_{\mathsf{X}}(x_i),$$

where $x^n = x_1 \dots x_n$ is an element of $\mathscr{X}^n$. We sometimes omit writing the subscript in P_{X}, when it is obvious over which set P is a distribution. For $\mathscr{A} \subset \mathscr{X}$, we denote by $P_{\mathsf{X}}(\mathscr{A})$ the probability of random variable X taking on a value in $\mathscr{A}$:

$$P_{\mathsf{X}}(\mathscr{A}) = \sum_{x \in \mathscr{A}} P_{\mathsf{X}}(x).$$

Lastly, for an arbitrary function $f : \mathscr{X} \to \mathscr{Y}$, we define $f^{-1}(y) := \{x \in \mathscr{X} : f(x) = y\}$.

2.1 Quantum information

A *quantum state* on $\mathbb{C}^d$ is a $d \times d$ positive semi-definite matrix of unit trace, i.e., $\rho \in \mathbb{C}^{d \times d}$ such that $\rho \succeq 0$ and $\operatorname{tr} \rho = 1$. We denote the set of all quantum states on $\mathbb{C}^d$ by $\mathrm{D}(\mathbb{C}^d)$. Operations on quantum states are described by *unitary* matrices, i.e., $U \in \mathbb{C}^{d \times d}$ such that $U^\dagger U = \mathbb{I}$ where $\mathbb{I}$ is the identity matrix. We denote the set of all unitaries on $\mathbb{C}^d$ by $\mathrm{U}(\mathbb{C}^d)$.

An n-outcome *measurement* or POVM on $\mathbb{C}^d$ is a collection of n positive semi-definite $d \times d$ matrices that sum to identity. We will denote a measurement by $M = \{M_1, \dots, M_n\}$ where $M_i \succeq 0$ and $\sum_{i=1}^n M_i = \mathbb{I}$. We denote the set of all n-outcome measurements on $\mathbb{C}^d$ by $\mathrm{M}(\mathbb{C}^d)$ (since the outcome set is always clear from the context, we do not specify it). If $M_i^2 = M_i$ for all $i = 1, \dots, n$, we call the measurement *projective*. We denote the set of all n-outcome projective measurements on $\mathbb{C}^d$ by $\mathrm{PM}(\mathbb{C}^d)$.

2.2 Linear programming

Linear programming is a technique for optimizing a linear function over a convex polytope. A polytope is a generalization of a polygon to any number of dimensions. There are two ways of describing a convex polytope: by giving its extreme points (and rays), called the vertex representation, or by linear constraints, called the half-space representation.

The half-space representation of a convex polytope is a collection of (closed) half-spaces, such that their intersection is the convex polytope. A half-space can be described by a linear inequality

$$a_1 x_1 + \cdots + a_n x_n \leq c. \tag{1}$$

Using this description, the convex polytope can be represented as a system of linear inequalities, which can be written as a matrix inequality

$$Ax \leq d.$$

Here, A is the matrix containing all coefficients a_i and d the vector containing all constants c, for all inequalities (1) representing the polytope. Note that we can also include linear equalities, as they can be described by two opposite inequalities.

Given a vertex representation, the corresponding convex polytope is the convex hull of the extreme points. The convex hull of a set of points is the smallest convex set that contains all the points, or simply the set of all convex combinations of the points (i.e., all weighted averages). This representation is especially interesting, since a linear function always has a global maximum in (at least) one of the extreme points of a convex polytope. We make use of this fact in Appendix C.2.

3 Local simultaneous state discrimination (LSSD)

In this section, we define the local simultaneous state discrimination (LSSD) task, originally introduced in [9]. In particular, we discuss strategies with classical, quantum and no-signalling resources for LSSD, and show that the optimal classical success probability can be attained by a symmetric strategy if certain conditions are fulfilled. Here we only consider the case of two players, Alice and Bob, but all definitions can easily be generalized to any number of players.

An LSSD game played by two players and a referee is defined by a classical-quantum-quantum (cqq) state ρ_{XAB}, where the referee's register X is classical while the Alice and Bob's registers A and B can generally be quantum. We denote the underlying spaces of X, A and B by $\mathcal{X} = \mathbb{C}^{\mathscr{X}}$, $\mathcal{A} = \mathbb{C}^{\mathscr{A}}$ and $\mathcal{B} = \mathbb{C}^{\mathscr{B}}$, respectively, where $\mathscr{X}$, $\mathscr{A}$ and $\mathscr{B}$ are some finite sets. We can always write the state ρ_{XAB} as

$$\rho_{\mathsf{XAB}} = \sum_{x \in \mathscr{X}} P_{\mathsf{X}}(x) \, |x\rangle\langle x|_{\mathsf{X}} \otimes \rho^x_{\mathsf{AB}},$$

where P_{X} is a probability distribution over $\mathscr{X}$ and each ρ^x_{AB} is a bipartite quantum state on $\mathcal{A} \otimes \mathcal{B}$. The state ρ_{XAB} is known to Alice and Bob, and they try to guess the referee's value x based on their reduced states ρ_{A} and ρ_{B}. We denote their guesses by x_A and x_B. In general, Alice and Bob may share some additional resources before the game, but they are not allowed to communicate with each other during the game. They win the game if both guesses are correct: $x_A = x_B = x$.

In most of this paper, we are going to consider the case where ρ_{XAB} is entirely classical. Meaning that there exists an orthonormal basis $\{|a\rangle : a \in \mathscr{A}\}$ of $\mathcal{A}$ and $\{|b\rangle : b \in \mathscr{B}\}$ of $\mathcal{B}$ that are independent of $x \in \mathscr{X}$, and probability distributions P^x_{AB} over $\mathscr{A} \times \mathscr{B}$ such that

$$\rho^x_{\mathsf{AB}} = \sum_{\substack{a \in \mathscr{A} \\ b \in \mathscr{B}}} P^x_{\mathsf{AB}}(a,b) \, |a\rangle\langle a|_{\mathsf{A}} \otimes |b\rangle\langle b|_{\mathsf{B}}.$$

In this case, it is useful to rephrase the problem. Instead of describing the game by a cqq state, we can describe it by a probability distribution P_{XAB} on $\mathscr{X} \times \mathscr{A} \times \mathscr{B}$. The referee picks elements $x \in \mathscr{X}, a \in \mathscr{A}$ and $b \in \mathscr{B}$ according to this distribution and gives a and b to Alice and Bob, respectively. Alice and Bob know the distribution P_{XAB} and both try to guess the value x. Again, they may share some resources, but are not allowed to communicate during the game, and they win if they both guess x correctly. A schematic representation of LSSD is shown in Fig. 2.

We now describe different types of strategies based on three different possible shared resources: classical, quantum and no-signalling. While these additional resources can be of different types, the strategies themselves are in general quantum since the LSSD game is based on a quantum state.

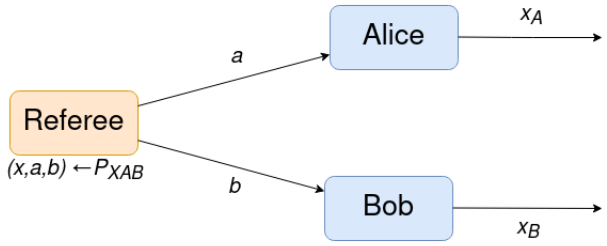

Figure 2: A schematic of the LSSD game. On inputs a and b, Alice and Bob make guesses x_A and x_B respectively, and win if $x = x_A = x_B$.

3.1 Classical resources

While strategies for LSSD may in general take advantage of shared randomness, this does not help in increasing the winning probability. Indeed, after a random value is generated, we are left with a deterministic strategy that depends on this value. Thus instead of the original randomized strategy, the players can just use one of the deterministic strategies that achieves the highest winning probability. Hence in the following, we assume that the players do not use shared randomness.

In the quantum case of the LSSD game (meaning that the game is described by a cqq state ρ_{XAB}), a strategy is completely defined by two measurements $M = \{M_x : x \in \mathscr{X}\}$ and $N = \{N_x : x \in \mathscr{X}\}$ on $\mathcal{A}$ and $\mathcal{B}$, respectively. Alice and Bob perform these measurements on their subsystems to produce their guesses for x. Given the measurements M and N, their winning probability is

$$\sum_{x\in\mathscr{X}} P_{\mathsf{X}}(x)\,\mathrm{tr}[\rho^x_{\mathsf{AB}}(M_x \otimes N_x)],$$

and the optimal winning probability is denoted by

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})_\rho := \sup_{\substack{M\in\mathrm{M}(\mathcal{A})\\ N\in\mathrm{M}(\mathcal{B})}} \sum_{x\in\mathscr{X}} P_{\mathsf{X}}(x)\,\mathrm{tr}[\rho^x_{\mathsf{AB}}(M_x \otimes N_x)],$$

where $\mathrm{M}(\mathcal{A})$ and $\mathrm{M}(\mathcal{B})$ denote the sets of all measurements on $\mathcal{A}$ and $\mathcal{B}$, respectively.

In case ρ_{XAB} is purely classical and described by a probability distribution P_{XAB}, the strategy of Alice and Bob is given by two conditional probability distributions $Q_{\mathsf{X}_A|\mathsf{A}}$ and $Q_{\mathsf{X}_B|\mathsf{B}}$ describing their local behaviour. The winning probability is then given by

$$\sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)Q_{\mathsf{X}_A|\mathsf{A}}(x|a)Q_{\mathsf{X}_B|\mathsf{B}}(x|b).$$

The optimal winning probability can now be obtained by maximizing over all conditional probabilities. However, we can restrict this optimization to maximizing over all deterministic strategies, i.e., strategies that can be described by two functions $f\colon \mathscr{A} \to \mathscr{X}$ and $g\colon \mathscr{B} \to \mathscr{X}$. Similarly to shared randomness, Alice and Bob can condition any local randomness on the realization that maximizes their probability of winning. Now, the optimal winning probability is given by

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P = \max_{f,g} \sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)\delta[f(a) = g(b) = x].$$

We say that a strategy is *symmetric* if Alice and Bob perform the same local strategy, i.e., if $f = g$. In the following theorem, we show that symmetric strategies attain optimal classical values for classical LSSD games (see Appendix A.1 for proof).

Theorem 3.1. *Let P_{XAB} be a distribution over $\mathscr{X} \times \mathscr{A} \times \mathscr{B}$, with $\mathscr{A} = \mathscr{B}$, satisfying the following:*

(i) The marginal distribution P_{X} over $\mathscr{X}$ is uniform.

(ii) $P_{\mathsf{AB|X}} = P_{\mathsf{A|X}}P_{\mathsf{B|X}}$.

(iii) $P_{\mathsf{A|X}} = P_{\mathsf{B|X}}$.

Then the classical LSSD game defined by P_{XAB} has an optimal deterministic strategy that is symmetric.

3.2 Quantum resources

In this case, Alice and Bob can share an entangled state prior to receiving their inputs. Let $\mathcal{A}' = \mathcal{B}' = \mathbb{C}^d$ be two complex Euclidean spaces of dimension d. Alice and Bob first jointly prepare a quantum state $\sigma_{\mathsf{A}'\mathsf{B}'}$ on $\mathcal{A}' \otimes \mathcal{B}'$, after which Alice and Bob keep systems A' and B', respectively. After receiving their inputs, Alice and Bob determine their output by measuring the registers AA' and BB' with local measurements M and N, respectively (this is the most general strategy because no communication is allowed).

When the local dimensions of the shared entangled state $\sigma_{\mathsf{A}'\mathsf{B}'}$ are limited to d for both parties, the optimal probability of winning is

$$\omega_{\mathrm{q}}^d(\mathsf{X}|\mathsf{A};\mathsf{B})_\rho := \sup_{\sigma_{\mathsf{A}'\mathsf{B}'}\in \mathrm{D}(\mathbb{C}^d\otimes\mathbb{C}^d)} \sup_{\substack{M\in\mathrm{M}(\mathcal{A}\otimes\mathbb{C}^d)\\ N\in\mathrm{M}(\mathcal{B}\otimes\mathbb{C}^d)}} \sum_{x\in\mathscr{X}} P_{\mathsf{X}}(x)\,\mathrm{tr}\big[(\rho^x_{\mathsf{AB}}\otimes\sigma_{\mathsf{A}'\mathsf{B}'})(M_x\otimes N_x)\big]. \tag{2}$$

When the dimensions of A' and B' are not limited, the optimal winning probability is

$$\omega_{\mathrm{q}}(\mathsf{X}|\mathsf{A};\mathsf{B})_\rho := \sup_{d\geq 1} \omega_{\mathrm{q}}^d(\mathsf{X}|\mathsf{A};\mathsf{B})_\rho. \tag{3}$$

When ρ_{XAB} is classical and described by a probability distribution P_{XAB}, we can simplify Eq. (2) as follows:

$$\begin{aligned} \omega_{\mathrm{q}}^d(\mathsf{X}|\mathsf{A};\mathsf{B})_P &= \sup_{\sigma_{\mathsf{A}'\mathsf{B}'}\in \mathrm{D}(\mathbb{C}^d\otimes\mathbb{C}^d)} \sup_{\substack{M:\mathscr{A}\to\mathrm{M}(\mathbb{C}^d)\\ N:\mathscr{B}\to\mathrm{M}(\mathbb{C}^d)}} \sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)\,\mathrm{tr}\big[\sigma_{\mathsf{A}'\mathsf{B}'}\big(M_x(a)\otimes N_x(b)\big)\big] \qquad (4)\\ &= \sup_{\substack{M:\mathscr{A}\to\mathrm{M}(\mathbb{C}^d)\\ N:\mathscr{B}\to\mathrm{M}(\mathbb{C}^d)}} \Bigg\| \sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b) M_x(a)\otimes N_x(b) \Bigg\|, \qquad (5) \end{aligned}$$

where M and N are collections of measurements, i.e., for every input $a \in \mathscr{A}$ and $b \in \mathscr{B}$, we have that $M(a) = \{M_x(a) : x \in \mathscr{X}\}$ and $N(b) = \{N_x(b) : x \in \mathscr{X}\}$ are measurements on $\mathbb{C}^d$ with outcomes in $\mathscr{X}$.

3.3 No-signalling resources

We define strategies with no-signaling resources only when ρ_{XAB} is classical and described by a probability distribution P_{XAB}. Given classical inputs $a \in \mathscr{A}$ and $b \in \mathscr{B}$ for Alice and Bob, respectively, they output their estimates x_A and x_B of $x \in \mathscr{X}$ according to a conditional probability distribution $Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}$ on $\mathscr{X} \times \mathscr{X} \times \mathscr{A} \times \mathscr{B}$ satisfying

$$\begin{aligned} \forall x_B, a, a', b: \quad & \sum_{x_A\in\mathscr{X}} Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x_A, x_B|a,b) = \sum_{x_A\in\mathscr{X}} Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x_A, x_B|a',b), \qquad (6)\\ \forall x_A, a, b, b': \quad & \sum_{x_B\in\mathscr{X}} Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x_A, x_B|a,b) = \sum_{x_B\in\mathscr{X}} Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x_A, x_B|a,b'). \qquad (7) \end{aligned}$$

An optimal no-signaling strategy succeeds with probability

$$\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P := \sup_{Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}} \sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b) Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x,x|a,b). \tag{8}$$

The set of classical correlations is a subset of the set of quantum correlations, and the latter is a subset of the set of no-signalling correlations, see [18] for more details. Therefore, we have that

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P \leq \omega_{\mathrm{q}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P \leq \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})_P. \tag{9}$$

Notice that the winning probability for a given no-signalling strategy is a linear function in the values $Q_{\mathsf{X}_A\mathsf{X}_B|\mathsf{AB}}(x_A, x_B|a,b)$. This, together with the fact that the set of no-signalling correlations forms a convex polytope, see e.g. [18], implies that we can use linear programming to find the optimal no-signalling winning probability of an LSSD game. It also implies that there is always an optimal strategy at one of the extreme points of the no-signalling polytope.

This last fact is what Majenz et al. used to prove that there exists no probability distribution P_{XAB} with binary x, a and b, such that the corresponding LSSD game can be won with higher probability using no-signalling strategies [9, Proposition 3.3]. They showed that none of the no-signalling correlations at the extreme points of the no-signalling polytope could ever perform better than the simple classical strategy of outputting the most likely value for x. We do something similar in Appendix C for the tripartite case. However, it turns out that this argument is not enough in the tripartite case, and we take a numerical approach to finish the argument.

4 The binary-symmetric-channel game

A binary symmetric channel (BSC) with error $\alpha \in [0, 1/2]$ is a channel with a single bit of input that transmits the bit without error with probability $1-\alpha$ and flips it with probability α, see Fig. 3. In this section, we study a particular LSSD problem: the binary-symmetric-channel game, originally introduced in [9, Example 1], where a referee sends a bit to Alice and Bob over two identical and independent binary symmetric channels, both with error probability α, see Definition 4.1 for a formal definition. In [9], an explicit optimal classical strategy for this game is shown and its corresponding optimal winning probability for every α is obtained. Moreover, the authors show that the winning probability cannot be improved by any quantum nor no-signalling strategy. In addition, they show that if two copies of the game are played in parallel for $\alpha = 1 - \frac{1}{\sqrt{2}}$, there is an explicit optimal classical strategy that performs better than repeating the optimal classical strategy for a single copy of the game twice and, as a consequence, quantum and no-signalling optimal strategies must perform better than repeating the respective optimal strategies for a single copy of the game.

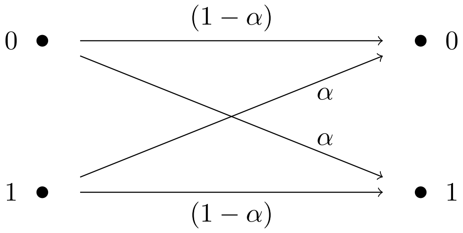

Figure 3: Schematic representation of a binary symmetric channel with error probability α.

In Section 4.1, we study the parallel repetition of the BSC game and, for the case of two copies, we provide the optimal classical, quantum and no-signalling values, showing that for most α the three values coincide (and in most of the cases the optimal values are obtained just by repeating the optimal strategy for a single copy of the BSC game). Nevertheless, for certain values of α, the classical and quantum values coincide but there is a no-signalling advantage.

In Section 4.2, we provide the optimal no-signalling winning probabilities for the three-fold parallel repetition of the BSC game. We study the 'good' classical and no-signalling strategies for arbitrary number n of parallel rounds of the BSC game in Section 4.3.

Definition 4.1 (Example 1 in [9])**.** *Let X, Y and Z be independent binary random variables such that X is uniformly random, i.e.,* $\Pr[X = 1] = 1/2$, *and* $\Pr[Y = 1] = \Pr[Z = 1] = \alpha$ *for* $\alpha \in [0, 1/2]$. *Let $A := X \oplus Y$ and $B := X \oplus Z$, and denote the joint probability mass function of (X, A, B) by $P^{\alpha}_{\mathsf{XAB}}$. The* binary-symmetric-channel *(BSC) game is defined as the task of simultaneously guessing X from A and B.*

Proposition 4.2 (Example 1 in [9])**.** *For every $\alpha \in [0, 1/2]$, the optimal classical, quantum and no-signalling winning probabilities for the BSC game P^{α} are equal and given by*

$$\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P^\alpha} = \omega_{\mathrm{q}}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P^\alpha} = \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P^\alpha} = \begin{cases} (1-\alpha)^2 & \textit{if } \alpha \in [0, 1-\frac{1}{\sqrt{2}}], \\ \frac{1}{2} & \textit{if } \alpha \in (1-\frac{1}{\sqrt{2}}, \frac{1}{2}]. \end{cases} \tag{10}$$

The optimal winning probability for $\alpha \in [0, 1 - 1/\sqrt{2}]$ is achieved by the strategy where Alice and Bob output the input they received. The intuition behind this strategy is that for 'small' α, the bits they receive most likely have not been flipped. Notice that if Alice and Bob were playing this game without having to coordinate their answers, such a strategy would be optimal for all α. In fact, the optimal strategy for 'high'-noise BSC channels, $\alpha \in (1 - 1/\sqrt{2}, 1/2]$, is achieved by both parties outputting some previously agreed bit.

4.1 Two-fold parallel repetition of the binary-symmetric-channel game

Let (X', A', B') be an independent copy of (X, A, B), as described in Definition 4.1. The two-fold parallel repetition of the BSC game consists of simultaneously guessing (X, X') from (A, A') and (B, B'). This game is described by the probability distribution $P^{\alpha}_{\mathsf{XAB}} \otimes P^{\alpha}_{\mathsf{X'A'B'}}$. According to [9], the optimal classical winning probability for the two-fold parallel repetition of the BSC game for $\alpha = 1 - \frac{1}{\sqrt{2}}$ is

$$\frac{1}{4}(1-\alpha^2)^2 + \frac{1}{4}(1-\alpha)^4. \tag{11}$$

Hence, for $\alpha = 1 - \frac{1}{\sqrt{2}}$, $\omega_{\mathrm{c}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha \otimes P^\alpha} > \omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})^2_{P^\alpha}$ and, from (9) and (10), we also have

$$\omega_{\mathrm{q}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha \otimes P^\alpha} > \omega_{\mathrm{q}}(\mathsf{X}|\mathsf{A};\mathsf{B})^2_{P^\alpha}, \tag{12}$$

$$\omega_{\mathrm{ns}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha \otimes P^\alpha} > \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})^2_{P^\alpha}. \tag{13}$$

Here we study the full range of α (namely, $\alpha \in [0, 1/2]$). In the following Theorem, we provide the optimal classical and no-signalling winning probabilities for the two-fold parallel repetition of the BSC game, graphically represented in Fig. 1. The Theorem shows that for most values of α, the classical and no-signalling optimal success probabilities coincide (and therefore so does the quantum value).

Theorem 4.3. *Let (X', A', B') be an independent copy of (X, A, B). Let $\alpha_0 < 1$ be the real solution of $(1-\alpha^2)^2 + (1-\alpha)^4 = 1$, i.e. $\alpha_0 \simeq 0.32814$, and let $I_1 = [0, 2-\sqrt{3}]$, $I_2 = (2-\sqrt{3}, \alpha_0]$, $I_3 = (\alpha_0, \frac{\sqrt{3}-1}{2}]$ and $I_4 = (\frac{\sqrt{3}-1}{2}, \frac{1}{2}]$. Then, for the two-fold parallel repetition of the BSC game, we have*

$$\omega_{\mathrm{c}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha \otimes P^\alpha} = \begin{cases} (1-\alpha)^4 & \text{if } \alpha \in I_1, \\ \frac{1}{4}(1-\alpha^2)^2 + \frac{1}{4}(1-\alpha)^4 & \text{if } \alpha \in I_2, \\ \frac{1}{4} & \text{if } \alpha \in I_3 \cup I_4, \end{cases} \tag{14}$$

and

$$\omega_{\mathrm{ns}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha \otimes P^\alpha} = \begin{cases} (1-\alpha)^4 & \text{if } \alpha \in I_1, \\ \frac{(1-\alpha^2)^2}{3} & \text{if } \alpha \in I_2 \cup I_3, \\ \frac{1}{4} & \text{if } \alpha \in I_4. \end{cases} \tag{15}$$

Proof. Since the BSC game fulfills the conditions of Theorem 3.1, a symmetric strategy will provide the optimal classical value. We determine ω_{c} by considering all deterministic classical strategies. For each strategy, we compute the winning probability as a function of α. Then we obtain the analytical value (14) by taking the maximum and applying the `PiecewiseExpand` command. For more details on this derivation, see the *Mathematica* file "`BSC classical strategy n=2.nb`" in [19].

The optimal no-signalling value can be found via a linear program, i.e., a maximization of a linear function subject to linear constraints. In *Mathematica*, the standard form to represent a linear program that optimizes over $x \in \mathbb{R}^n$ is

$$\begin{aligned} &\textit{Primal problem:} \qquad && \text{minimize: } \langle c, x\rangle = \sum_{i=1}^{n} c_i x_i \\ & && \text{subject to: } Ax + b \geq 0, \\ & && \qquad\qquad A_{\mathrm{eq}} x + b_{\mathrm{eq}} = 0, \end{aligned} \tag{16}$$

where $x, c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $A_{\text{eq}} \in \mathbb{R}^{k \times n}$, $b_{\text{eq}} \in \mathbb{R}^k$ (see LinearOptimization for more details). Its dual, which optimizes over $\lambda \in \mathbb{R}^m$ and $\nu \in \mathbb{R}^k$, is given by

$$\begin{aligned} \textit{Dual problem:} \qquad & \text{maximize: } -\big(\langle b, \lambda\rangle + \langle b_{\text{eq}}, \nu\rangle\big) = -\sum_{i=1}^{m} b_i \lambda_i - \sum_{i=1}^{k} b_{\text{eq},i} \nu_i \\ & \text{subject to: } A^{\mathsf{T}} \lambda + A_{\text{eq}}^{\mathsf{T}} \nu - c = 0, \\ & \qquad \lambda \geq 0. \end{aligned} \tag{17}$$

A common technique in linear programming is to use one of the two problems to obtain a bound on the other. In the above formulation, any feasible solution to the dual problem (17) provides a lower bound on the optimal solution of the primal problem (16). The optimal value of both problems can be determined by finding feasible primal and dual solutions that have the same value. Then, as a consequence of strong duality, both solutions must be optimal.

Since the original linear program for computing ω_{ns} for the BSC game is quite large, see Eqs. (6) to (8), we first simplify it by reducing the number of parameters. We do this by imposing the following symmetries on Alice's and Bob's no-signalling strategy Q:[1]

1. By Lemma 5.11 below, there is an optimal no-signalling strategy that is invariant under any permutation of the instances of the game, i.e., $Q\big(\sigma(x), \sigma(y)|\sigma(a), \sigma(b)\big) = Q(x, y|a, b)$, for any permutation σ of positions within a string.

2. Since the BSC game is symmetric under exchanging Alice and Bob, we can also exchange Alice's and Bob's strategies, i.e., $Q(y, x|b, a) = Q(x, y|a, b)$.

3. Since the BSC game is symmetric under negating any subset of input and output bits, we can do the same to Alice's and Bob's strategy, i.e., $Q(x \oplus s, y \oplus s|a \oplus s, b \oplus s) = Q(x, y|a, b)$ for any bit string s.

After performing the above symmetry reductions, we need to find feasible primal and dual solutions of equal value. These solutions should be α-dependent, i.e., work not just for a single value of α but for whole intervals of α. We managed to find such solutions with the help of *Mathematica*, and we have provided them in the format of Eqs. (16) and (17) in the notebook "`BSC no-signalling strategy n=2.nb`" [19]. The primal and dual objective values of these solutions match and agree with Eq. (15) in each of the intervals $I_1, \dots, I_4$ (occasionally we could not obtain a single α-dependent solution for a whole interval, in which case we broke it into smaller subintervals).

Finding these exact α-dependent solutions required some numerical tricks. Indeed, while it is easy to solve the linear program for any particular value of α, obtaining continuous α-dependent solutions is nontrivial – it requires interpolating from a small number of solutions, or often even a single solution. We used a combination of the following numerical tricks to cover all cases in Eq. (15) (often obtaining the same solution with different methods):

- *Rational multiples of π*: We chose a rational number r so that $\alpha = r\pi$ lies in a given interval I_i. Using LinearOptimization we then find a symbolic solution that is polynomial in π.[2] Substituting back $\pi = \alpha/r$ gives us an exact polynomial α-dependent solution. This is quite remarkable since we have effectively interpolated a polynomial function from a single irrational point. This strategy unfortunately did not work for 3 repetitions of the game since the linear program was too large.

- *Rational solutions*: We choose a sequence of equally spaced rational values of α and find exact rational solutions for these values by using LinearOptimization. We then interpolate between them by using FindSequenceFunction. This method generally requires some fiddling with the chosen sequence since nearby values of α can lead to completely different and unrelated solutions.

[1] Here we consider only two parallel repetitions of the BSC game. But the same symmetry reductions can be performed for any number of repetitions (see Theorem 4.5).

[2] This works since on one hand *Mathematica* treats π symbolically, while on the other it can compare π to any other number by calculating its numerical value to arbitrary accuracy. It is also important that *Mathematica* can manipulate rational numbers symbolically and that π is irrational.

- *Algebraic solutions*: We choose an algebraic α from the given interval I_i and find a numerical solution for this α to extremely high accuracy (300 digits). Then we use `RootApproximant` to turn this numerical solution into exact algebraic numbers. Reconstructing the minimal polynomial for each of these numbers gives us an interpolated α-dependent solution that is polynomial. This trick effectively interpolates from a single algebraic point.

Checking the primal and dual constraints of the resulting interpolated solution gives us constraints on α that capture the interval in which this solution holds.

It is important to note that, irrespective of how dirty the above numerical methods are, once an exact α-dependent solution is found, it can be easily verified that it satisfies all constraints and gives equal primal and dual values, hence implying optimality. For more details, see "`BSC no-signalling strategy n=2.nb`" in [19]. □

Notice that, unlike a single copy of the BSC game, the optimal winning probabilities have different behaviors split into three different intervals. We see that

$$\omega_{\mathrm{c}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha} = \omega_{\mathrm{ns}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha} = \omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})^2_{P^\alpha} \qquad \forall \alpha \in I_1 \cup I_4, \tag{18}$$

and therefore, due to (9), the quantum value is the same value as the classical. Analogously to the single copy of the BSC game, for 'small' α, $\alpha \in I_1$, an optimal classical and no-signalling strategy is given by Alice and Bob outputting their input. The intuition behind it is that, due to 'low' noise, every bit has low probability of being flipped, $(1-\alpha)$, and thus the winning probability using this strategy is $(1-\alpha)^4$. On the other hand, an optimal classical and no-signalling strategy for a 'high' noisy channel, $\alpha \in I_3 \cup I_4$ and $\alpha \in I_4$, respectively, is that both Alice and Bob output some previously agreed bit string. This leads to the conclusion that the corresponding optimal winning probabilities for these values of α can be achieved by just repeating the optimal classical and no-signalling strategies mentioned above for a single copy of the BSC game. Nevertheless, this is not always the case, since

$$\omega_{\mathrm{c}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha} < \omega_{\mathrm{ns}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha} \qquad \forall \alpha \in I_2 \cup I_3. \tag{19}$$

An optimal classical strategy for $\alpha \in I_2$ is given by Alice and Bob both outputting 00 if their input contains a 0 and outputting 11, otherwise, which gives an optimal winning probability of $\frac{1}{4}(1-\alpha^2)^2 + \frac{1}{4}(1-\alpha)^4$, which was already given in [9] for $\alpha = 1 - \frac{1}{\sqrt{2}}$. An optimal no-signalling strategy for $\alpha \in I_2 \cup I_3$ is given by

$$Q_2(x,y|a,b) = \begin{cases} \frac{1}{3} & \text{if } (x = y \text{ or } x \oplus b = 11 = y \oplus a) \text{ and } (x \oplus a \neq 11 \neq y \oplus b), \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

This strategy, see Section 4.3.2, has winning probability $(1-\alpha^2)^2/3$. More specifically, for $\alpha \in I_2$ and for $\alpha \in I_2 \cup I_3$ there exist classical and no-signalling strategies, respectively, that perform better than repeating the optimal strategy, i.e.

$$\begin{aligned} \omega_{\mathrm{c}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha} &> \omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})^2_{P^\alpha} \ \forall \alpha \in I_2, \\ \omega_{\mathrm{ns}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha} &> \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})^2_{P^\alpha} \ \forall \alpha \in I_2 \cup I_3. \end{aligned} \tag{21}$$

We are left with characterizing the value $\omega_{\mathrm{q}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha}$ for $\alpha \in I_2 \cup I_3$. From (19), the optimal quantum value for $\alpha \in I_2 \cup I_3$ has to be in between the two values. Based on strong numerical evidence (see Fig. 1), in Conjecture 4.4 below we conjecture that there is no quantum advantage with over the optimal classical strategy for any α.

Unlike the set of classical and the set of no-signaling correlations, the set of quantum correlations, $\mathcal{Q}$, has uncountably many extremal points, see e.g. [18], making the optimization problem a tough task. In [20], Navascués, Pironio and Acín (NPA) introduced an infinite hierarchy of conditions necessarily satisfied by any set of quantum correlations with the property that each of them can be tested using semidefinite programming (SDP) and thus they can be used to exclude non-quantum correlations, see Appendix B. The authors introduced a recursive way to construct subsets $\mathcal{Q}_\ell \supset \mathcal{Q}_{\ell+1} \supset \mathcal{Q}$ for all $\ell \in \mathbb{N}$, each of them can be tested using semidefinite programming and are such that $\cap_{\ell\in\mathbb{N}}\mathcal{Q}_\ell = \mathcal{Q}$, i.e. they converge to the set of quantum correlations.

By using an intermediate level between the first and the second levels of the NPA hierarchy, the so-called level “$1+MN$” (see Appendix B for a detailed explanation and “`NPA_hierarchy_BSC_Game.py`” [19] for the numerical code), we find that for $\alpha \in I_2$, $\omega_{\mathrm{q}}(\mathsf{XX'|AA';BB'})_{P^\alpha\otimes P^\alpha}$ is upper bounded by $\omega_{\mathrm{c}}(\mathsf{XX'|AA';BB'})_{P^\alpha\otimes P^\alpha}$, see Fig. 1 (b). Therefore, this shows that the values coincide in the interval I_2. The reason to restrict ourselves to the level “$1+MN$” is that it requires less computational resources than computing the level 2 and it already provides tight bounds. Based on the fact that the numerical upper bounds on the quantum value obtained by solving the semidefinite programs match the (analytical) lower bounds given by the classical values, we state the following conjecture.

Conjecture 4.4. *There is no quantum advantage over the best classical strategy for the two-fold parallel repetition of the BSC game for any value of α.*

4.2 Three-fold parallel repetition of the BSC game

Consider the three-fold parallel repetition of the BSC game. In the following Theorem, we provide the optimal classical and no-signalling winning probabilities, and we will see that for a vast range of values of α they coincide and therefore so does the quantum.

Theorem 4.5. *Let (X',A',B') and (X'',A'',B'') be two independent copies of (X,A,B) and let α_1 be the root of the polynomial $2(1-\alpha)^4(1+2\alpha)-1$ taking the value $\alpha_1 \simeq 0.358121$, $\alpha_2 = \frac{1}{8}(3-\sqrt{7}+\sqrt{2(32-11\sqrt{7})})$ and $\alpha_3 = 2^{-\frac{2}{3}}(4-\sqrt{14})^{\frac{1}{3}}$. Then, for three copies of the BSC game,*

$$\omega_{\mathrm{c}}(\mathsf{XX'X''|AA'A'';BB'B''})_{P^\alpha\otimes P^\alpha\otimes P^\alpha} = \begin{cases} (1-\alpha)^6 & \text{if } \alpha \in [0,\frac{1}{4}], \\ \frac{1}{4}(1-\alpha)^4(1+2\alpha) & \text{if } \alpha \in (\frac{1}{4},\alpha_1], \\ \frac{1}{8} & \text{if } \alpha \in (\alpha_1,\frac{1}{2}], \end{cases} \tag{22}$$

$$\omega_{\mathrm{ns}}(\mathsf{XX'X''|AA'A'';BB'B''})_{P^\alpha\otimes P^\alpha\otimes P^\alpha} = \begin{cases} (1-\alpha)^6 & \text{if } \alpha \in [0,\frac{1}{4}] =: J_1, \\ \frac{1}{4}(1-\alpha)^4(1+2\alpha)^2 & \text{if } \alpha \in (\frac{1}{4},\alpha_2] =: J_2, \\ \frac{1}{7}(1-\alpha^3)^2 & \text{if } \alpha \in [\alpha_2,\alpha_3] =: J_3, \\ \frac{1}{8} & \text{if } \alpha \in [\alpha_3,\frac{1}{2}] =: J_4. \end{cases} \tag{23}$$

Proof. The proof is analogous to the proof of Theorem 4.3 for two parallel repetitions. In “`BSC classical strategy n=3.nb`” [19] we perform an optimized search over all symmetric classical strategies leading to (22). In “`BSC no-signalling strategy n=3.nb`” [19] we provide explicit analytic α-dependent solutions for the primal and dual linear programs for the no-signalling value. Both solutions have identical objective value that agrees with (23). □

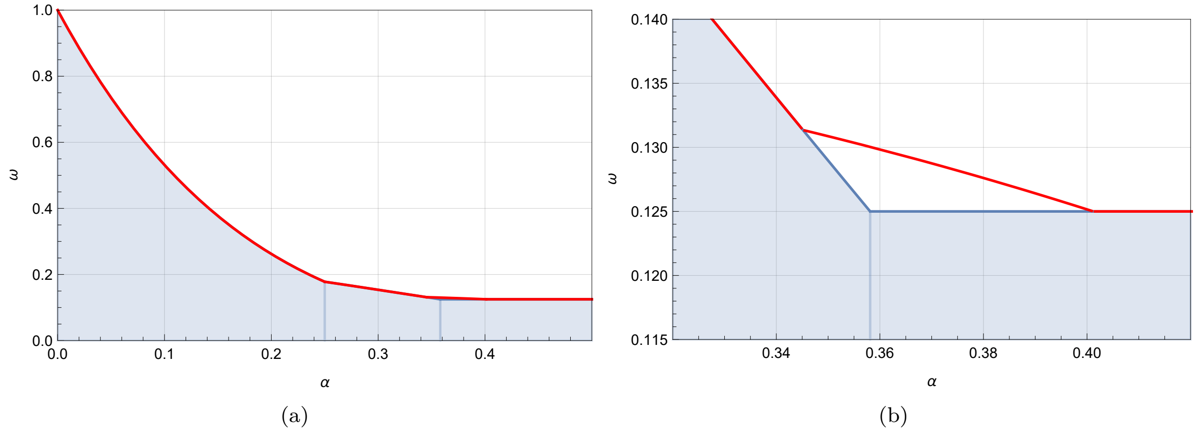

Figure 4: (a) Optimal classical (blue) and no-signalling (red) winning probabilities for the three-fold parallel repetition of the BSC game. The blue area represents the values below the optimal classical winning probabilities. (b) Zoom in of (a) for the values of α around 0.37 where the classical and no-signalling values differ.

See Fig. 4 for a graphical representation of the optimal values from Theorem 4.5. For 'low' noise, $\alpha \in J_1$, the optimal value is attained by the classical strategy consisting on Alice and Bob outputting the received bit, i.e. repeating three times the optimal classical strategy for a single copy of the game. On the other side, for 'high' noise, $\alpha \in J_4$, the optimal value is attained by the classical strategy where Alice and Bob output a pre-agreed bit, which is also obtained by repeating the optimal strategy for a single copy. Therefore,

$$\omega_{\mathrm{c}}(\mathsf{XX'X''}|\mathsf{AA'A''};\mathsf{BB'B''})_{P^\alpha\otimes P^\alpha\otimes P^\alpha} = \omega_{\mathrm{ns}}(\mathsf{XX'X''}|\mathsf{AA'A''};\mathsf{BB'B''})_{P^\alpha\otimes P^\alpha\otimes P^\alpha} = \omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A};\mathsf{B})^3_{P^\alpha},\ \forall\alpha \in J_1 \cup J_4. \tag{24}$$

For $\alpha \in J_2$, the no-signalling optimal value can be attained by the deterministic strategy consisting on Alice and Bob outputting 111 if they receive an input with more zeros than ones and outputting 000 otherwise. See Section 4.3 for no-signalling and classical strategies attaining this optimal value. For this interval, the optimal strategy for three copies is better than any combination of optimal two and one copies of the BSC game, i.e.

$$\begin{aligned}&\omega_{\mathrm{c}}(\mathsf{XX'X''}|\mathsf{AA'A''};\mathsf{BB'B''})_{P^\alpha\otimes P^\alpha\otimes P^\alpha} = \omega_{\mathrm{ns}}(\mathsf{XX'X''}|\mathsf{AA'A''};\mathsf{BB'B''})_{P^\alpha\otimes P^\alpha\otimes P^\alpha}\\ &> \omega_{\mathrm{ns}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha}\omega_{\mathrm{ns}}(\mathsf{X''}|\mathsf{A''};\mathsf{B''})_{P^\alpha} \geq \omega_{\mathrm{c}}(\mathsf{XX'}|\mathsf{AA'};\mathsf{BB'})_{P^\alpha\otimes P^\alpha}\omega_{\mathrm{c}}(\mathsf{X''}|\mathsf{A''};\mathsf{B''})_{P^\alpha}\\ &> \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A};\mathsf{B})^3_{P^\alpha},\ \forall\alpha \in J_2.\end{aligned} \tag{25}$$

For $\alpha \in J_3$ the following no-signalling strategy achieves the optimal value, as we explain in Section 4.3,

$$Q_3(x,y|a,b) = \begin{cases}\frac{1}{7} & \text{if } (x = y \text{ or } x\oplus b = 111 = y\oplus a) \text{ and } (x\oplus a \neq 111 \neq y\oplus b),\\ 0 & \text{otherwise.}\end{cases} \tag{26}$$

4.3 Arbitrary parallel repetition

In this section, we will look to find classes of good strategies, both classical and no-signalling, for the n-fold parallel repetition of the BSC game.

4.3.1 Classical strategies

We have already seen some similarities in classical strategies between one, two and three copies of the game. For small α, the best strategy is always to output the input (identity strategy). For α close to 1/2 the best strategy is to output some fixed bitstring regardless of the input (constant strategy). The winning probabilities of these strategies for n copies are $(1-\alpha)^{2n}$ and 2^{-n}, respectively. For two and three copies, we also found similar strategies "in between" the identity and constant strategies. These strategies can also be extended to n copies: outputting 0^n if the input contains at least as many zeros as ones and outputting 1^n otherwise (majority strategy). For odd n, the winning probability of the majority strategy is given by

$$\frac{1}{2^{n-1}}\left(\sum_{i=0}^{(n-1)/2}\binom{n}{i}\alpha^i(1-\alpha)^{n-i}\right)^2. \tag{27}$$

An error-correcting code for the BSC consists of a message set M and two functions $\mathrm{Enc}\colon M \to \{0,1\}^n$ and $\mathrm{Dec}\colon \{0,1\}^n \to M$. The objective of an error-correcting code is to send a message m over the BSC by first encoding it using Enc, sending the result over the BSC and recovering m using Dec, such that the probability of a correct recovery of m is maximized. We will look at error-correcting codes more formally in Section 5. The readers already familiar with error-correcting codes will notice that the majority strategy is exactly applying $\mathrm{Enc}\circ\mathrm{Dec}$ from the repetition code to the input: the repetition code encodes messages 0 and 1 to 0^n and 1^n respectively and decodes by picking the bit that appears the most in the input. This motivates us to look at error-correcting codes to define strategies for n repetitions of the BSC game.

Example 4.6. We consider the (7,4)-Hamming code, perhaps the most famous code for the BSC, introduced by Richard Hamming [21]. This code encodes bitstrings $d_1d_2d_3d_4$ of length 4 as bitstrings of length 7 by appending three parity bits: $d_1d_2d_3d_4p_1p_2p_3$. These bits represent the parity (XOR) of three of the original 4 bits (see Fig. 5).

Decoding works by checking if the parity bits are still correct (still equal to the parity of the corresponding 3 bits). If this is the case, we just remove the last three bits of the received bitstring. Now suppose an error occurred in exactly one bit.

- If the error occurred in d_4, all the parity bits are incorrect.
- If the error occurred in d_1, d_2 or d_3, two of the parity bits are incorrect (p_1 and p_2 for d_1, p_1 and p_3 for d_2 and p_2 and p_3 for d_3).
- If the error occurred in one of the parity bits, only that parity bit will be incorrect.

Using the above, we can perfectly deduce in which bit the error occurred and correct it accordingly. If more than one error occurs, this method never decodes correctly.

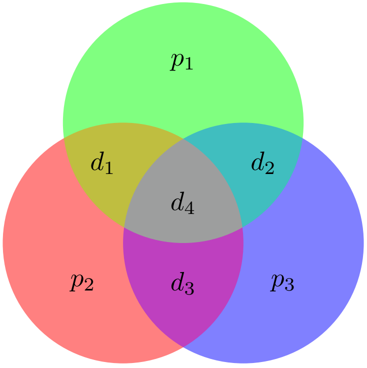

Figure 5: The Hamming code visualized: The bitstring $d_1 d_2 d_3 d_4$ is encoded by appending the parity bits p_1, p_2 and p_3, where each parity bit represents the parity of the three bits inside their circle. A single error in one of the seven bits can be perfectly detected by checking which parity bits are incorrect.

Since the Hamming code corrects exactly 0 or 1 error, we can write the average success probability of this code as

$$(1-\alpha)^7 + 7\alpha(1-\alpha)^6.$$

Now consider the following strategy for 7 copies of the BSC game based on the Hamming code: both players perform the correction part of the Hamming code on their input and output the result (this is the same as decoding and then encoding again). It is obvious that the players win if and only if the initial bitstring x is in the range of the encode function and the decoding of both players was successful. This observation results in the following winning probability:

$$\frac{2^4}{2^7}\left((1-\alpha)^7 + 7\alpha(1-\alpha)^6\right)^2.$$

It turns out that this Hamming code strategy is strictly better for a large range of α than the identity, constant and majority strategy for 7 copies of the game. This confirms the idea that error-correcting codes define good classical strategies.

4.3.2 No-signalling strategies

For two and three copies of the BSC game, we found the optimal no-signalling strategies Q_2 and Q_3 (described in Eqs. (20) and (26)). We can extend these no-signalling strategies to n copies as follows:

$$Q(x,y|a,b) = \begin{cases} \frac{1}{2^n-1} & \text{if } (x=y \text{ or } x\oplus b = 1^n = y \oplus a) \text{ and } (x\oplus a \neq 1^n \neq y\oplus b), \\ 0 & \text{otherwise.} \end{cases} \tag{28}$$

There is, however, a more intuitive way to describe this no-signalling correlation. Alice outputs uniformly at random any bit string, except the negation of her input. Bob outputs the same string as Alice, except when that string happens to be the negation of his input, in which case he outputs the

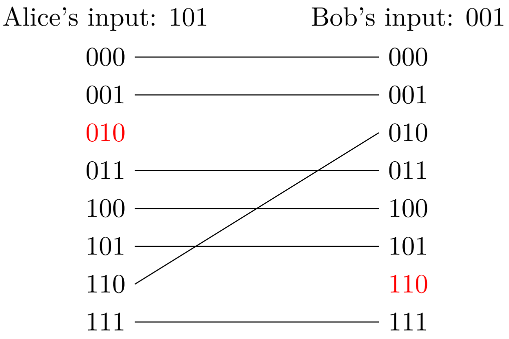

Figure 6: An example of a pairing of elements between the output sets of Alice and Bob, for three simultaneous copies. Each line represents a pair, and at the end of the process, one pair is chosen uniformly at random.

negation of Alice's input (see Fig. 6). Note that the roles of Alice and Bob in this description can be exchanged.

This formulation makes it obvious that we can define a more general class of no-signalling strategies: instead of the output sets consisting of everything apart from the opposite of the input, we can let the output sets consist of all bitstrings within Hamming distance d from the input. We can then pair up the elements from the output sets and say that each of those pairs is output with equal probability. Again, if an element occurs in both lists, we pair it with itself. This description defines a no-signalling strategy, since Alice and Bob always output each of the elements of their output sets with the same probability, regardless of the input of the other. We denote by Q_n^d a no-signalling strategy for n copies of the BSC game defined by Hamming distance d. Note that for $d \in \{1, \ldots, n-2\}$ the strategy Q_n^d is not unique, but they all achieve the same winning probability.

Let us find the winning probability of a strategy Q_n^d. Suppose that x is the bitstring generated by the referee. The only way the players could output the combination (x, x) is if both $d(x, a) \leq d$ and $d(x, b) \leq d$, in which case it is output with probability $\left(\sum_{i=0}^d \binom{n}{d}\right)^{-1}$, since the sum is the size of their output sets. The probability that a lies within distance d from x is $\sum_{i=0}^d \binom{n}{i}\alpha^i(1-\alpha)^{n-i}$. We conclude that the winning probability of Q_n^d is given by

$$\frac{1}{\sum_{i=0}^d \binom{n}{i}} \left(\sum_{i=0}^d \binom{n}{i} \alpha^i (1-\alpha)^{n-i} \right)^2 .$$

It turns out that all the optimal winning probabilities for one, two and three simultaneous copies of the BSC game can be achieved by a strategy of the form Q_n^d. If we pick $d = 0$ we get exactly the identity strategy. If we pick $d = n$, we get the average of all possible constant strategies (and by linearity, this achieves the same winning probability as a constant strategy). If we pick $d = n - 1$, we get exactly the strategy defined in Eq. (28). This strategy achieves winning probability

$$\frac{1}{2^n - 1} \left(\sum_{i=0}^n \binom{n}{i} \alpha^i (1-\alpha)^{n-i} - \alpha^n \right)^2 = \frac{1}{2^n - 1} (1 - \alpha^n)^2 .$$

We are left with segment two for three copies. The strategy Q_3^1 achieves winning probability

$$\frac{1}{4} \left((1-\alpha)^3 + 3\alpha(1-\alpha)^2 \right)^2 .$$

This probability is exactly the same winning probability as the majority strategy, which we found to be optimal in this segment. We conclude that all optimal winning probabilities for one, two and three copies of the game can be achieved by a strategy of the form Q_n^d.

It turns out that the class of strategies defined in this section can be described using a list-decoding scheme for the BSC channel. In the next section, we discuss strategies for a general channel $P_{\mathsf{A|X}}$ based on error-correcting codes and list-decoding schemes.

5 Channel LSSD games

In the previous section, we constructed an LSSD game based on a BSC. In this section, we extend this construction and define an LSSD game based on an arbitrary channel. For n parallel instances of these games, we discuss classical strategies based on error-correcting codes and no-signalling strategies based on list-decoding schemes. We also investigate the asymptotic behaviour of the optimal winning probability as n approaches infinity. Note that for any non-local game with optimal no-signalling winning probability smaller than 1 (and no promise on the input distribution), the optimal winning probability for n parallel instances of the game exponentially goes to 0 [22, Theorem 16]. Thus, we will be considering the limit of the exponent of the winning probability normalized by n.

We briefly recap basic concepts from information theory that we need in this section including entropic quantities and method of types. For a more in-depth introduction, see [23, Chapter 2] and [24, Chapter 2].

Let P be a probability distribution over $\mathscr{X}$, and let X be a random variable distributed according to P. We define the *entropy* $H(X)_P = H(P)$ of X as

$$H(X)_P := -\sum_{x\in\mathscr{X}} P(x)\log(P(x)),$$

with the convention that $P(x)\log(P(x)) = 0$ wherever $P(x) = 0$. We drop subscript P whenever the distribution of X is clear from the context. Let X and Y be two random variables with joint probability distribution P_{XY}. The *joint entropy* of X and Y is $H(X,Y)_P = H(P_{\mathsf{XY}})$ and the *conditional entropy* is

$$H(X|Y)_P := H(X,Y)_P - H(Y)_P.$$

The *mutual information* of two random variables X and Y is

$$I(X;Y)_P := H(X)_P + H(Y)_P - H(X,Y)_P.$$

For two probability distributions P and Q over $\mathscr{X}$, the *relative entropy* is

$$D(P\|Q) := \sum_{x\in\mathscr{X}} P(x)\log\left(\frac{P(x)}{Q(x)}\right).$$

If $P^1_{\mathsf{X|Y}}$ and $P^2_{\mathsf{X|Y}}$ are two conditional distributions over $\mathscr{X}\times\mathscr{Y}$ and Q_{Y} is a distribution over $\mathscr{Y}$, the corresponding *conditional relative entropy* is

$$D(P^1_{\mathsf{X|Y}}\|P^2_{\mathsf{X|Y}} \mid Q_{\mathsf{Y}}) := \sum_{y\in\mathscr{Y}} Q_{\mathsf{Y}}(y) D(P^1_{\mathsf{X|Y}=y}\|P^2_{\mathsf{X|Y}=y}).$$

We next introduce preliminaries on the *method of types* (see [24, Chapter 2] for further reading). Let $\mathscr{X}$ be a finite set and n be a positive integer. For a sequence $x^n \in \mathscr{X}^n$, its *type* is a probability distribution P over $\mathscr{X}$ defined as

$$P(x) := \frac{|\{i : x_i = x\}|}{n}. \tag{29}$$

Let $\mathcal{P}_n(\mathscr{X})$ denote the set of all types of sequences in $\mathscr{X}^n$. For a given distribution P over $\mathscr{X}$, we denote by $\mathcal{T}_P$ all sequences in $\mathscr{X}^n$ whose type is P. If P_{AX} is a joint probability distribution on $\mathscr{A}\times\mathscr{X}$ and x^n is a sequence in $\mathcal{T}_{P_{\mathsf{X}}}$, we let $\mathcal{T}_{P_{\mathsf{A|X}}}(x^n) := \{a^n : (a^n, x^n) \in \mathcal{T}_{P_{\mathsf{AX}}}\}$. We need the following inequalities whose proofs are in [24]:

$$|\mathcal{P}_n(\mathscr{X})| \le (n+1)^{|\mathscr{X}|}, \tag{30}$$

$$|\mathcal{T}_{P_{\mathsf{X}}}| \le 2^{nH(X)_P}, \tag{31}$$

$$|\mathcal{T}_{P_{\mathsf{A|X}}}(x^n)| \ge \frac{2^{nH(A|X)_P}}{(n+1)^{|\mathscr{A}|}}. \tag{32}$$

We are now ready to define the main object of this section, channel LSSD games.

Definition 5.1. *The channel LSSD game defined by $P_{\mathsf{A}|\mathsf{X}}$ is given by the probability distribution*

$$P_{\mathsf{XAB}} = P_{\mathsf{X}} P_{\mathsf{A}|\mathsf{X}} P_{\mathsf{B}|\mathsf{X}},$$

with P_{X} the uniform distribution over $\mathscr{X}$, $\mathscr{A} = \mathscr{B}$, and $P_{\mathsf{B}|\mathsf{X}} = P_{\mathsf{A}|\mathsf{X}}$.

Playing n parallel copies of this channel game is the same as playing the channel game defined by the channel $P_{\mathsf{A}|\mathsf{X}}^{\times n}$, which can be thought of as the referee generating a string $x^n \in \mathscr{X}^n$ and sending it to Alice and Bob by n independent uses of their channels. Our main result of this section is the following characterization of the exponent of the optimal probability of winning for all three classes of strategies.

Theorem 5.2. *Let $P_{\mathsf{A}|\mathsf{X}}$ be a channel and let $P_{\mathsf{XAB}}^{\times n}$ be the probability distribution defining the channel game corresponding to the channel $P_{\mathsf{A}|\mathsf{X}}^{\times n}$. We have*

$$\begin{aligned}\lim_{n\to\infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} = \lim_{n\to\infty} \frac{\log(\omega_{\mathrm{q}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} &= \lim_{n\to\infty} \frac{\log(\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \\ &= \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|).\end{aligned}$$

Note that to prove the theorem it is enough to prove the following two lemmas because of Eq. (9).

Lemma 5.3 (Achievability)**.** *We have*

$$\liminf_{n\to\infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \geq \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|). \tag{33}$$

Lemma 5.4 (Converse)**.** *We have*

$$\limsup_{n\to\infty} \frac{\log(\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \leq \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|). \tag{34}$$

Our proof for these two lemmas is based on tools from information theory that we introduce in Section 5.1. We prove the first lemma in Section 5.2 by constructing a classical LSSD strategy from a code for the corresponding channel, and then by choosing an appropriate sequence of codes that optimize the winning probability. We prove the second lemma in Section 5.3 by first relating the winning probability of an arbitrary no-signalling strategy to a list-decoding code, and then using a converse for list-decoding codes.

5.1 Tools from information theory

We recall here basic definitions concerning error-correction codes. A code for n uses of channel $P_{\mathsf{A}|\mathsf{X}}$ operates as follows. The sender has a message set M of possible messages. He picks one message $m \in M$ to send and encodes m as a codeword x^n of $\mathscr{X}^n$, using a function $\mathrm{Enc}\colon M \to \mathscr{X}^n$. Next, he transmits each of the symbols x_i of this codeword to the receiver by consecutive uses of the channel; the receiver receives an a^n in $\mathscr{A}^n$ and decodes it to a message m' using a function $\mathrm{Dec}\colon \mathscr{A}^n \to M$. The communication was successful if $m = m'$.

The minimum success probability of a code is given by

$$\min_{m\in M} P_{\mathsf{A}|\mathsf{X}}^{\times n}(\mathrm{Dec}^{-1}(m)|\,\mathrm{Enc}(m)),$$

and the rate of a code is $\frac{1}{n}\log|M|$.

Definition 5.5. *We call a code* (Enc, Dec) *for a channel $P_{\mathsf{A}|\mathsf{X}}$ an* $(n, 2^{nR}, \alpha)$-code *if*

$$\mathrm{Enc}\colon [2^{nR}] \to \mathscr{X}^n, \tag{35}$$

$$\mathrm{Dec}\colon \mathscr{A}^n \to [2^{nR}], \tag{36}$$

$$P_{\mathsf{A}|\mathsf{X}}^{\times n}(\mathrm{Dec}^{-1}(m)|\,\mathrm{Enc}(m)) \geq \alpha, \quad \forall m \in [2^{nR}]. \tag{37}$$

We know since Shannon's groundbreaking work [25] that there exists a sequence of codes with rate less than the capacity of the channel and probability of success tending to one. We also know from the strong-converse results [26] that if the rate is above the capacity, the success probability exponentially tends to zero. In [27], the optimal exponent of the success probability has been characterized. The following lemma is what we require in our achievability proof. Its proof resembles the proof in [27], but we need to modify it because we consider the minimum success probability, not the average. We leave the proof to Appendix A.2.

Lemma 5.6. *Let $P_{\mathsf{A}|\mathsf{X}}$ be a channel, Q_{XA} a probability distribution over $\mathscr{X} \times \mathscr{A}$ and $\delta > 0$. For $n \geq n_0(|\mathscr{X}|, |\mathscr{A}|, \delta)$, there exists an*

$$\left(n, 2^{n(I(X;A)_Q - \delta)}, \frac{2^{-nD(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} | Q_{\mathsf{X}})}}{p(n)}\right)$$

code for the channel $P_{\mathsf{A}|\mathsf{X}}$ where $p(n)$ is a polynomial depending only on $|\mathscr{A}|$.

We next recall the definition of list decoding. The decoder here outputs a list of L messages, instead of a single message. The decoding is successful if the list contains the correct message. We denote the list output by the decoder on input a^n by C_{a^n}. The minimum success probability is then

$$\min_{m \in M} \sum_{a^n \in \mathscr{A}^n \colon C_{a^n} \ni m} P_{\mathsf{A}|\mathsf{X}}^{\times n}(a^n | \operatorname{Enc}(m)).$$

Definition 5.7. *We call a list-decoding code an $(n, 2^{nR}, L, \alpha)$-code if* Dec *maps elements a^n of $\mathscr{A}^n$ to subsets C_{a^n} of $[2^{nR}]$ of size L and*

$$\begin{aligned} &\operatorname{Enc}\colon [2^{nR}] \to \mathscr{X}^n, \\ &\sum_{a^n \in \mathscr{A}^n \colon C_{a^n} \ni m} P_{\mathsf{A}|\mathsf{X}}^{\times n}(a^n | \operatorname{Enc}(m)) \geq \alpha, \quad \forall m \in [2^{nR}]. \end{aligned}$$

We have the following converse for list-decoding schemes, see Appendix A.3 for the proof.

Lemma 5.8. *For any list-decoding $(n, 2^{nR}, 2^{nR_L}, 2^{-n\zeta_n})$ code for $P_{\mathsf{A}|\mathsf{X}}$, we have*

$$\zeta_n \geq \min_{Q_{\mathsf{XA}}} \left[D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) + \max\{R - R_L - I(X;A)_Q, 0\} \right] + O\left(\frac{\log n}{n}\right),$$

where the constant hidden in $O(\cdot)$ depends only on $|\mathscr{X}|$ and $|\mathscr{A}|$.

5.2 Achievability: Classical strategies from error-correction codes

We prove Lemma 5.3 in this section, which we re-state for readers' convenience.

Lemma 5.9. *We have*

$$\liminf_{n \to \infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n | \mathsf{A}^n; \mathsf{B}^n)_{P^{\times n}})}{n} \geq \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A}|\mathsf{X}} \| P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|). \tag{38}$$

Proof. We first explain how to use error-correction codes to find classical strategies for the parallel repetition of a channel LSSD game. Let $(\operatorname{Enc}, \operatorname{Dec})$ be an $(n, 2^{nR}, \alpha)$-code for the channel $P_{\mathsf{A}|\mathsf{X}}$. We consider a classical LSSD strategy for n parallel repetitions of the channel LSSD game in which both players use the estimation function $f := \operatorname{Enc} \circ \operatorname{Dec}$. This strategy can be interpreted as the players decoding directly to the codeword of a message instead of to the message itself. We lower bound the winning probability of the strategy given by f as

$$\frac{1}{|\mathscr{X}|^n} \sum_{x^n \in \mathscr{X}^n} P_{\mathsf{A}|\mathsf{X}}^{\times n}(f^{-1}(x^n) | x^n)^2 = \frac{1}{|\mathscr{X}|^n} \sum_{x^n \in \mathscr{X}^n} \left(\sum_{a^n \in \mathscr{A}^n : f(a^n) = x^n} P_{\mathsf{A}|\mathsf{X}}^{\times n}(a^n | x^n) \right)^2 \tag{39}$$

$$\overset{(a)}{\geq} \frac{1}{|\mathscr{X}|^n} \sum_{x^n \in \operatorname{Im}(\operatorname{Enc})} \alpha^2 = \frac{2^{nR}}{|\mathscr{X}|^n} \alpha^2, \tag{40}$$

where (a) follows since $(\mathrm{Enc}, \mathrm{Dec})$ is an $(n, 2^{nR}, \alpha)$-code. Notice that there is a trade-off between the success probability and the number of messages. We simultaneously want the success probability and the number of messages to be large. However, increasing one necessarily means decreasing the other.

Let Q_{XA} be a distribution on $\mathscr{X} \times \mathscr{A}$ and $\delta > 0$. Let $n_0(|\mathscr{X}|, |\mathscr{A}|, \delta)$ and $p(n)$ be as in Lemma 5.6 where $p(n)$ is a polynomial only depending on $|\mathscr{A}|$. By Lemma 5.6, for $n \geq n_0(|\mathscr{X}|, |\mathscr{A}|, \delta)$, there exists an $\left(n, 2^{n(I(X;A)_Q - \delta)}, \frac{2^{-nD(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}})}}{p(n)}\right)$-code for $P_{\mathsf{A|X}}$. Let $f = \mathrm{Enc} \circ \mathrm{Dec}$ be the strategy defined by this code. The winning probability of this strategy is at most the optimal classical winning probability, so by using Eq. (40) we find

$$\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}} \geq \frac{2^{n(I(X;A)_Q - \delta - D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} | Q_{\mathsf{X}}))}}{|\mathscr{X}|^n p(n)},$$

and therefore

$$\frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \geq I(X;A)_Q - \delta - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|) - \frac{\log(p(n))}{n}. \tag{41}$$

Since Eq. (41) holds for any Q_{XA} and $\delta > 0$, and $\lim_{n\to\infty} \frac{\log(p(n))}{n} = 0$, we conclude that

$$\lim_{n\to\infty} \frac{\log(\omega_{\mathrm{c}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \geq \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|).$$

This completes the proof of the achievability of the error exponent. □

5.3 Converse: No-signalling LSSD strategies and list-decoding codes

We prove Lemma 5.4 in this section which we restate for readers' convenience.

Lemma 5.10. *We have*

$$\limsup_{n\to\infty} \frac{\log(\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}})}{n} \leq \max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|). \tag{42}$$

We first prove the existence of an optimal strategy invariant under permutations of inputs and outputs. To this end, we need the following notation: For a permutation $\sigma \in S_n$ and a sequence $x^n \in \mathscr{X}^n$ we denote by $\sigma(x^n) \in \mathscr{X}^n$ the sequence obtained from x^n by permuting its entries according to σ.

Lemma 5.11. *For n parallel repetitions of a channel LSSD game, there is an optimal no-signalling strategy Q such that*

$$\forall \sigma \in S_n: \quad Q(\sigma(x^n), \sigma(y^n)|\sigma(a^n), \sigma(b^n)) = Q(x^n, y^n|a^n, b^n). \tag{43}$$

Proof. Let Q be an optimal strategy and $\sigma \in S_n$. The strategy Q_σ defined by $Q_\sigma(x^n, y^n|a^n, b^n) = Q(\sigma(x^n), \sigma(y^n)|\sigma(a^n), \sigma(b^n))$ has the same winning probability as Q, since the n-fold probability distribution is invariant under permutations: $P_{\mathsf{XAB}}^{\times n}(x^n, a^n, b^n) = P_{\mathsf{XAB}}^{\times n}(\sigma(x^n), \sigma(a^n), \sigma(b^n))$. We define

$$\hat{Q} := \frac{1}{n!} \sum_{\sigma \in S_n} Q_\sigma.$$

The strategy $\hat{Q}$ satisfies (43): for any $\tau \in S_n$,

$$\begin{aligned}
\hat{Q}(\tau(x^n), \tau(y^n)|\tau(a^n), \tau(b^n)) &= \frac{1}{n!} \sum_{\sigma \in S_n} Q_\sigma(\tau(x^n), \tau(y^n)|\tau(a^n), \tau(b^n)) \\
&= \frac{1}{n!} \sum_{\sigma \in S_n} Q(\sigma(\tau(x^n)), \sigma(\tau(y^n))|\sigma(\tau(a^n)), \sigma(\tau(b^n))) \\
&= \frac{1}{n!} \sum_{\pi \in S_n} Q(\pi(x^n), \pi(y^n)|\pi(a^n), \pi(b^n)) \\
&= \frac{1}{n!} \sum_{\pi \in S_n} Q_\pi(x^n, y^n|a^n, b^n) \\
&= \hat{Q}(x^n, y^n|a^n, b^n).
\end{aligned}$$

Finally, by linearity of the winning probability, $\hat{Q}$ also achieves the same winning probability as Q, which means that it is optimal. □

Proof of Lemma 5.10. Let Q be an optimal strategy satisfying (43). Its marginal distributions $Q(x^n|a^n)$ and $Q(y^n|b^n)$ only depend on the joint type of (x^n, a^n) and (y^n, b^n), respectively. In particular, we can write the winning probability of Q as follows:

$$\begin{aligned}\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}} &= \sum_{x^n,a^n,b^n} P_{\mathsf{XAB}}^{\times n}(x^n,a^n,b^n)Q(x^n,x^n|a^n,b^n)\\ &= \sum_{\substack{Q_{\mathsf{XA}}\in\mathcal{P}_n(\mathscr{X}\times\mathscr{A})\\ Q'_{\mathsf{XB}}\in\mathcal{P}_n(\mathscr{X}\times\mathscr{B})}} \sum_{\substack{x^n,a^n,b^n:\\ (x^n,a^n)\in\mathcal{T}_{Q_{\mathsf{XA}}}\\ (x^n,b^n)\in\mathcal{T}_{Q'_{\mathsf{XA}}}}} P_{\mathsf{XAB}}^{\times n}(x^n,a^n,b^n)Q(x^n,x^n|a^n,b^n), \qquad (44)\end{aligned}$$

where $\mathcal{P}_n(\cdot)$ denotes the set of types of length-n strings over a given set, and $\mathcal{T}_\cdot$ denotes all sequences of a given type.

Since there are $(n+1)^{2|\mathscr{X}||\mathscr{A}|}$ terms in the first sum of (44), there must exist $Q_{\mathsf{XA}} \in \mathcal{P}_n(\mathscr{X}\times\mathscr{A})$ and $Q'_{\mathsf{XB}} \in \mathcal{P}_n(\mathscr{X}\times\mathscr{B})$ such that

$$\sum_{\substack{x^n,a^n,b^n:\\ (x^n,a^n)\in\mathcal{T}_{Q_{\mathsf{XA}}}\\ (x^n,b^n)\in\mathcal{T}_{Q'_{\mathsf{XA}}}}} P_{\mathsf{XAB}}^{\times n}(x^n,a^n,b^n)Q(x^n,x^n|a^n,b^n) \geq \frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^{2|\mathscr{X}||\mathscr{A}|}}. \qquad (45)$$

Let us define for each a^n and b^n

$$C_{a^n} := \{x^n : (x^n,a^n) \in \mathcal{T}_{Q_{\mathsf{XA}}}\}, \qquad (46)$$

$$D_{b^n} := \{x^n : (x^n,b^n) \in \mathcal{T}_{Q'_{\mathsf{XB}}}\}. \qquad (47)$$

Now consider the following strategy $\tilde{Q}$:

- on input (a^n, b^n), Alice and Bob generate (x^n, y^n) according to Q;
- Alice checks if $x^n \in C_{a^n}$ and if not, uniformly generates a new output $\tilde{x}^n$ from C_{a^n} (if C_{a^n} is the empty set, Alice generates an arbitrary output);
- Bob checks if $y^n \in D_{b^n}$ and if not, uniformly generates a new output $\tilde{y}^n$ from D_{b^n} (if D_{b^n} is the empty set, Bob generates an arbitrary output).

This strategy is no-signalling and has winning probability of at least $\frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^{2|\mathscr{X}||\mathscr{A}|}}$, by (45). We also have that $\tilde{Q}(x^n|a^n)$ is uniform over C_{a^n} when $C_{a^n} \neq \varnothing$, since $Q(x^n|a^n)$ only depends on the joint type of (x^n, a^n). Similarly, $\tilde{Q}(y^n|b^n)$ is uniform over D_{b^n} when $D_{b^n} \neq \varnothing$.

Note that for any a^n and a'^n, if C_{a^n} and $C_{a'^n}$ are non-empty, then $|C_{a^n}| = |C_{a'^n}|$. We define $L_A := |C_{a^n}|$ for a non-empty C_{a^n} and similarly define $L_B := |D_{b^n}|$ for a non-empty D_{b^n}. We find

$$\begin{aligned}\frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^{2|\mathscr{X}||\mathscr{A}|}} &\leq \sum_{x^n,a^n,b^n} P_{\mathsf{XAB}}^{\times n}(x^n,a^n,b^n)\tilde{Q}(x^n,x^n|a^n,b^n)\\ &\leq \sum_{x^n,a^n,b^n} P_{\mathsf{XAB}}^{\times n}(x^n,a^n,b^n)\min\{\tilde{Q}(x^n|a^n),\tilde{Q}(x^n|b^n)\}\\ &\leq \frac{1}{\max\{L_A,L_B\}}\sum_{x^n,a^n,b^n} P_{\mathsf{XAB}}^{\times n}(x^n,a^n,b^n)\delta(x^n\in C_{a^n})\delta(x^n\in D_{b^n})\\ &= \frac{1}{\max\{L_A,L_B\}|\mathscr{X}|^n}\sum_{x^n}\left(\sum_{a^n:\,C_{a^n}\ni x^n} P_{\mathsf{A|X}}^{\times n}(a^n|x^n)\right)\left(\sum_{b^n:\,D_{b^n}\ni x^n} P_{\mathsf{A|X}}^{\times n}(b^n|x^n)\right).\end{aligned}$$

Upon defining

$$q_A(x^n) := \sum_{a^n\colon C_{a^n} \ni x^n} P_{\mathsf{A|X}}^{\times n}(a^n|x^n),$$
$$q_B(x^n) := \sum_{b^n\colon D_{b^n} \ni x^n} P_{\mathsf{A|X}}^{\times n}(b^n|x^n),$$

we can write

$$\sum_{x^n} \left(\sum_{a^n\colon C_{a^n} \ni x^n} P_{\mathsf{A|X}}^{\times n}(a^n|x^n) \right) \left(\sum_{b^n\colon D_{b^n} \ni x^n} P_{\mathsf{A|X}}^{\times n}(b^n|x^n) \right) = \sum_{x^n} q_A(x^n) q_B(x^n). \tag{48}$$

By Cauchy-Schwartz inequality, we have

$$\sum_{x^n} q_A(x^n) q_B(x^n) \leq \sqrt{\left(\sum_{x^n} p_A(x^n)^2\right)\left(\sum_{x^n} p_B(x^n)^2\right)}. \tag{49}$$

Therefore, without loss of generality, we can assume that

$$\sum_{x^n} q_A(x^n) q_B(x^n) \leq \sum_{x^n} p_A(x^n)^2. \tag{50}$$

We can upper-bound the winning probability of the strategy as

$$\frac{\omega_{\mathrm{ns}}(\mathsf{X}^n|\mathsf{A}^n;\mathsf{B}^n)_{P^{\times n}}}{(n+1)^{2|\mathscr{X}||\mathscr{A}|}} \leq \frac{1}{\max(L_A, L_B)|\mathscr{X}|^n} \sum_{x^n} q_A(x^n)^2 \leq \frac{1}{L_A|\mathscr{X}|^n} \sum_{x^n} q_A(x^n)^2. \tag{51}$$

Let $\delta > 0$. For each $i \geq 0$, we define

$$\mathcal{R}_i := \{x^n \in \mathscr{X}^n \mid 2^{-n\delta(i+1)} \leq q_A(x^n) < 2^{-n\delta i}\}.$$

We define a list-decoding scheme $(\mathrm{Enc}_i, \mathrm{Dec}_i)$ as follows: $\mathrm{Enc}_i\colon \mathcal{R}_i \to \mathscr{X}^n$ is the identity function and

$$\mathrm{Dec}_i(a^n) = C_{a^n} \cap \mathcal{R}_i.$$

Note that intersecting C_{a^n} with $\mathcal{R}_i$ only decreases the size of the list, making the code weaker. This observation means that we will still be able to use Lemma 5.8 for a list decoding with list size L. For each $x^n \in \mathcal{R}_i$, we have

$$\sum_{a^n\colon \mathrm{Dec}_i(a^n) \ni x^n} P_{\mathsf{A|X}}^{\times n}(a^n|x^n) \geq q_A(x^n) \geq 2^{-n\delta(i+1)},$$

so $(\mathrm{Enc}_i, \mathrm{Dec}_i)$ defines an $(n, |\mathcal{R}_i|, L, 2^{-\delta(i+1)})$-code. By Lemma 5.8, we have

$$\delta(i+1) \geq \min_{Q_{\mathsf{XA}}} D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) + \max\left\{\frac{\log|\mathcal{R}_i|}{n} - \frac{\log(L_A)}{n} - I(X;A)_Q, 0\right\} + O\left(\frac{\log n}{n}\right).$$

We find that if $q_A(x^n) > 0$, then $q_A(x^n) \geq 2^{-n\mu}$, with $\mu := \max_{x,a: P_{\mathsf{A|X}}(a|x)>0} -\log(P_{\mathsf{A|X}}(a|x))$. Thus, if $i \geq t := \lfloor \frac{\mu}{\delta} \rfloor$, then $\mathcal{R}_i$ is empty. Now, we find

$$\frac{1}{L_A} \sum_{x^n \in \mathscr{X}^n} q_A(x^n)^2 = \sum_{i=0}^{t} \sum_{x^n \in \mathcal{R}_i} \frac{1}{L_A} q_A(x^n)^2 \tag{52}$$
$$\leq \sum_{i=0}^{t} \frac{|\mathcal{R}_i|}{L_A} 2^{-2n\delta i} \tag{53}$$
$$\leq \sum_{i=0}^{t} 2^{n\left(\frac{\log|\mathcal{R}_i|}{n} - \frac{\log(L_A)}{n} - 2\min_{Q_{\mathsf{XA}}}\left(D(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}}|Q_{\mathsf{X}}) + \max\left\{\frac{\log|\mathcal{R}_i|}{n} - \frac{\log(L)}{n} - I(X;A)_Q, 0\right\}\right) + O\left(\frac{\log n}{n}\right) + \delta\right)} \tag{54}$$
$$\leq \sum_{i=0}^{t} 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}}|Q_{\mathsf{X}})\right) + O\left(\frac{\log n}{n}\right) + \delta\right)} \tag{55}$$
$$= \left(\lfloor \frac{\mu}{\delta} \rfloor + 1\right) 2^{n\left(\max_{Q_{\mathsf{XA}}}\left(I(X;A)_Q - 2D(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}}|Q_{\mathsf{X}})\right) + O\left(\frac{\log n}{n}\right) + \delta\right)}. \tag{56}$$

Combining (51) and (56), taking logarithm from both sides, and choosing $\delta = 1/n$ yields the desired converse results. □

5.4 Calculating the exponent for BSCs

We calculate, for BSCs, the value of the limit of the exponent in Theorem 5.2: $\max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$. To this extent, let Q_{XA} be a distribution over $\{0,1\} \times \{0,1\}$. Let us calculate the exponent one step at a time. First of all, we have

$$I(X;A)_Q = H(X)_Q + H(A)_Q - H(X,A)_Q$$

and

$$H(X)_Q = -\sum_{x=0}^{1} Q_{\mathsf{X}}(x) \log(Q_{\mathsf{X}}(x)) = -\sum_{x=0}^{1} \left(\sum_{a=0}^{1} Q_{\mathsf{XA}}(x,a) \right) \log \left(\sum_{a=0}^{1} Q_{\mathsf{XA}}(x,a) \right).$$

We can find $H(A)_Q$ in a similar way. We also have

$$H(X,A)_Q = -\sum_{x,a=0}^{1} Q_{\mathsf{XA}}(x,a) \log(Q_{\mathsf{XA}}(x,a)).$$

Now let us find the value of $D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}})$:

$$\begin{aligned} D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) &= \sum_{x=0}^{1} Q_{\mathsf{X}}(x) D(Q_{\mathsf{A|X}=x} \| P_{\mathsf{A|X}=x}) \\ &= \sum_{x=0}^{1} \left(\sum_{a=0}^{1} Q_{\mathsf{XA}}(x,a) \right) \left(\sum_{a=0}^{1} Q_{\mathsf{A|X}}(a|x) \log \left(\frac{Q_{\mathsf{A|X}}(a|x)}{P_{\mathsf{A|X}}(a|x)} \right) \right). \end{aligned}$$

Using numerical analysis we found that the maximum $\max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$ is always achieved by a distribution Q_{XA} for which $Q_{\mathsf{XA}}(0,0) = Q_{\mathsf{XA}}(1,1) =: c$ and $Q_{\mathsf{XA}}(0,1) = Q_{\mathsf{XA}}(1,0) =: d$. Using this property, we have

$$H(X)_Q = H(A)_Q = -2(c+d)\log(c+d),$$

and

$$H(X,Y)_Q = -2c\log(c) - 2d\log(d).$$

We also find

$$D(Q_{\mathsf{A|X}} \| P_{\mathsf{X|A}} \mid Q_{\mathsf{X}}) = 2\left(c \log \left(\frac{c}{(c+d)(1-\alpha)} \right) + d \log \left(\frac{d}{(c+d)\alpha} \right) \right).$$

Combining the expressions above, we find the value $I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|)$. Note that for Q_{XA} to be a distribution, we need $d = \frac{1}{2} - c$. This observation means that we only need to maximize with respect to the variable c (we see α as a constant). We can solve this maximization by calculating the derivative, setting it to 0 and solving for c. Using a computer algebra system, we find

$$\max_{Q_{\mathsf{XA}}} I(X;A)_Q - 2D(Q_{\mathsf{A|X}} \| P_{\mathsf{A|X}} \mid Q_{\mathsf{X}}) - \log(|\mathscr{X}|) = \log(1 - 2(1-\alpha)\alpha). \tag{57}$$

In Fig. 7 we plotted this expression together with the exponent of the optimal winning probability achieved by the strategies Q_n^d for some n (see Section 4.3.2). We can clearly see how this exponent approaches the limit calculated in (57).

Acknowledgements

We thank Christian Majenz for useful discussions. MO was supported by an NWO Vidi grant (Project No. VI.Vidi.192.109). LEF was supported by the Dutch Ministry of Economic Affairs and Climate Policy (EZK), as part of the Quantum Delta NL programme.

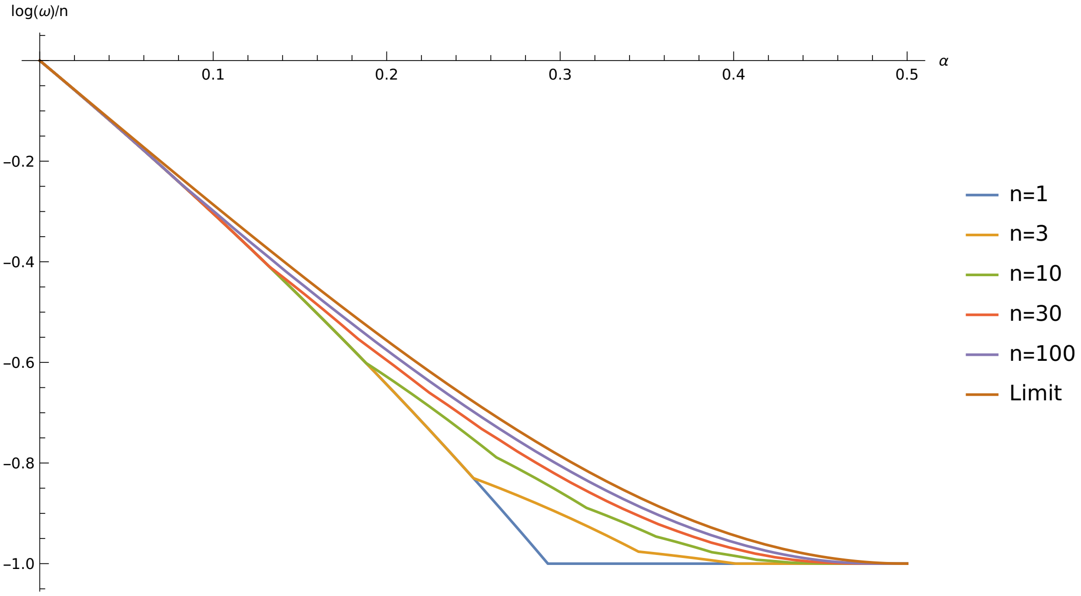

Figure 7: Plot of $\log(\omega)/n$ for different values of n and the limit of this expression, given by Eq. (57), against α. We calculated ω as the optimal winning probability achieved by the strategies Q_n^d (see Section 4.3.2).

References

[1] Richard E. Blahut. "Hypothesis testing and information theory". IEEE Transactions on Information Theory **20**, 405–417 (1974).

[2] Yury Polyanskiy, H. Vincent Poor, and Sergio Verdu. "Channel coding rate in the finite blocklength regime". IEEE Transactions on Information Theory **56**, 2307–2359 (2010).

[3] Ueli Maurer. "Authentication theory and hypothesis testing". IEEE Transactions on Information Theory **46**, 1350–1356 (2000).

[4] Larry Wasserman. "All of statistics: A concise course in statistical inference". Springer Texts in Statistics. Springer. New York (2004). 1st edition.

[5] Carl W. Helstrom. "Quantum detection and estimation theory". Journal of Statistical Physics **1**, 231–252 (1969).

[6] Joonwoo Bae and Leong-Chuan Kwek. "Quantum state discrimination and its applications". Journal of Physics A: Mathematical and Theoretical **48**, 083001 (2015).

[7] Charles H. Bennett, David P. DiVincenzo, Christopher A. Fuchs, Tal Mor, Eric Rains, Peter W. Shor, John A. Smolin, and William K. Wootters. "Quantum nonlocality without entanglement". Physical Review A **59**, 1070–1091 (1999).

[8] Andrew M. Childs, Debbie Leung, Laura Mančinska, and Maris Ozols. "A framework for bounding nonlocality of state discrimination". Communications in Mathematical Physics **323**, 1121–1153 (2013).

[9] Christian Majenz, Maris Ozols, Christian Schaffner, and Mehrdad Tahmasbi. "Local simultaneous state discrimination". Physical Review A**109** (2024).

[10] Anne Broadbent and Sébastien Lord. "Uncloneable quantum encryption via oracles". In Steven T. Flammia, editor, 15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020). Volume 158 of Leibniz International Proceedings in Informatics (LIPIcs), pages 4:1–4:22. Dagstuhl, Germany (2020). Schloss Dagstuhl–Leibniz-Zentrum für Informatik. arXiv:1903.00130.

[11] Christian Majenz, Christian Schaffner, and Mehrdad Tahmasbi. "Limitations on uncloneable encryption and simultaneous one-way-to-hiding" (2021). arXiv:2103.14510.

[12] Prabhanjan Ananth, Fatih Kaleoglu, Xingjian Li, Qipeng Liu, and Mark Zhandry. "On the feasibility of unclonable encryption, and more". In Yevgeniy Dodis and Thomas Shrimpton, editors, Advances in Cryptology – CRYPTO 2022. Pages 212–241. Springer Nature (2022).

[13] Andrea Coladangelo, Jiahui Liu, Qipeng Liu, and Mark Zhandry. "Hidden cosets and applications to unclonable cryptography". In Tal Malkin and Chris Peikert, editors, Advances in Cryptology – Crypto 2021. Pages 556–584. Springer (2021). arXiv:2107.05692.

[14] Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. "A monogamy-of-entanglement game with applications to device-independent quantum cryptography". New Journal of Physics **15**, 103002 (2013). arXiv:1210.4359.

[15] Abbas El Gamal and Young-Han Kim. "Network information theory". Cambridge University Press. (2011).

[16] Alexander S. Holevo. "Quantum systems, channels, information". De Gruyter. (2019).

[17] Omar Fawzi and Paul Fermé. "Beating the sum-rate capacity of the binary adder channel with non-signaling correlations". In 2022 IEEE International Symposium on Information Theory (ISIT). Pages 2750–2755. (2022).

[18] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. "Bell nonlocality". Rev. Mod. Phys. **86**, 419–478 (2014). arXiv:1303.2849.

[19] Jaron Has, Llorenç Escolà Farràs, and Maris Ozols. "Parallel repetition of LSSD (GitHub repository)". `https://github.com/JaronHas/ParallelRepetitionOfLSSD` (2024).

[20] Miguel Navascués, Stefano Pironio, and Antonio Acín. "A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations". New Journal of Physics **10**, 073013 (2008). arXiv:0803.4290.

[21] Richard W. Hamming. "Error detecting and error correcting codes". Bell System Technical Journal **29**, 147–160 (1950).

[22] Harry Buhrman, Serge Fehr, and Christian Schaffner. "On the parallel repetition of multi-player games: The no-signaling case". In Steven T. Flammia and Aram W. Harrow, editors, 9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2014). Volume 27 of Leibniz International Proceedings in Informatics (LIPIcs), pages 24–35. Dagstuhl, Germany (2014). Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. arXiv:1312.7455.

[23] Thomas M. Cover and Joy A. Thomas. "Elements of information theory". Wiley. (2005).

[24] Imre Csiszár and János Körner. "Information theory: coding theorems for discrete memoryless systems". Cambridge University Press. (2011).

[25] Claude E. Shannon. "A mathematical theory of communication". The Bell System Technical Journal **27**, 379–423 (1948).

[26] Suguru Arimoto. "On the converse to the coding theorem for discrete memoryless channels (corresp.)". IEEE Transactions on Information Theory **19**, 357–359 (1973).

[27] Gunter Dueck and János Körner. "Reliability function of a discrete memoryless channel at rates above capacity (corresp.)". IEEE Transactions on Information Theory **25**, 82–85 (1979).

[28] John Watrous. "Lecture 8: The hierarchy of Navascués, Pironio, and Acín". Lecture notes of "Advanced topics in quantum information theory", `https://johnwatrous.com/wp-content/uploads/2023/08/QIT-notes.08.pdf` (2021).

[29] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu, and David Roberts. "Nonlocal correlations as an information-theoretic resource". Physical Review A **71**, 022101 (2005). arXiv:quant-ph/0404097.

[30] Komei Fukuda. `https://people.inf.ethz.ch/fukudak/cdd_home/` (2022).

[31] Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. "Extremal correlations of the tripartite no-signaling polytope". Journal of Physics A: Mathematical and Theoretical **44**, 065303 (2011). arXiv:1101.2477.

A Proofs

A.1 Proof of Theorem 3.1

Let two functions $f\colon \mathscr{A} \to \mathscr{X}$ and $g\colon \mathscr{B} \to \mathscr{X}$ define a deterministic strategy. We prove that either Alice and Bob both performing f or both performing g can only increase the winning probability. Note that Alice and Bob can perform the same strategy, since $\mathscr{A} = \mathscr{B}$. The winning probability of the strategy defined by f and g is given by

$$\begin{aligned}
&\sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{XAB}}(x,a,b)\delta[f(a)=g(b)=x]\\
=&\frac{1}{|\mathscr{X}|}\sum_{\substack{x\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_{\mathsf{AB|X}}(a,b|x)\delta[f(a)=g(b)=x]\\
=&\frac{1}{|\mathscr{X}|}\sum_{x\in\mathscr{X}}\left(\sum_{a\in\mathscr{A}} P_{\mathsf{A|X}}(a|x)\delta[f(a)=x]\right)\left(\sum_{b\in\mathscr{B}} P_{\mathsf{B|X}}(b|x)\delta[g(b)=x]\right)\\
=&\frac{1}{|\mathscr{X}|}\sum_{x\in\mathscr{X}} P_{\mathsf{A|X}}(f^{-1}(x)|x)P_{\mathsf{A|X}}(g^{-1}(x)|x),
\end{aligned}$$

where in the first, second and third equalities we have used hypotheses (i), (ii) and (iii) of Theorem 3.1, respectively and notice that $f^{-1}(x)$ and $g^{-1}(x)$ might be sets. Write $q_f(x) := P_{\mathsf{A|X}}(f^{-1}(x)|x)$ and $q_g(x) := P_{\mathsf{A|X}}(g^{-1}(x)|x)$. Notice that q_f and q_g are vectors indexed by $x \in \mathscr{X}$, so we can write the winning probability as an inner product of these vectors:

$$\frac{1}{|\mathscr{X}|}\langle q_f, q_g\rangle\,. \tag{58}$$

Using the Cauchy–Schwarz inequality,

$$|\langle q_f, q_g\rangle|^2 \le \langle q_f, q_f\rangle\langle q_g, q_g\rangle\,,$$

and thus we cannot have $\langle q_f, q_g\rangle > \langle q_f, q_f\rangle$ and $\langle q_f, q_g\rangle > \langle q_g, q_g\rangle$. Therefore, we can conclude that Alice and Bob either both performing f or both performing g does not decrease the winning probability given in Eq. (58). Now suppose we picked f and g to form an optimal strategy, then by the previous statement, we immediately find a symmetric deterministic strategy that is also optimal.

A.2 Proof of Lemma 5.6

The proof of Lemma 5.6 relies on concepts and theorems from the book by Csiszár and Körner [24]. We will not be discussing these concepts here. We repeat the statement of Lemma 5.6 here for the reader's convenience.

Lemma 5.6. *Let $P_{\mathsf{A|X}}$ be a channel, Q_{XA} a probability distribution over $\mathscr{X}\times\mathscr{A}$ and $\delta > 0$. For $n \ge n_0(|\mathscr{X}|, |\mathscr{A}|, \delta)$, there exists an*

$$\left(n, 2^{n(I(X;A)_Q-\delta)}, \frac{2^{-nD(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}}|Q_{\mathsf{X}})}}{p(n)}\right)$$

code for the channel $P_{\mathsf{A|X}}$ where $p(n)$ is a polynomial depending only on $|\mathscr{A}|$.

Proof. Let $R = I(X;A)_Q - \delta$. By the packing lemma (Lemma 10.1 in [24]), there exists a function $\mathrm{Enc}\colon [2^{nR}] \to \mathscr{X}^n$ such that

- $\mathrm{Enc}(m)$ is of type Q_X for all $m \in [2^{nR}]$;
- $|\mathcal{T}_{Q_{A|X}}(\mathrm{Enc}(m)) \cap \bigcup_{m'\neq m}\mathcal{T}_{Q_{A|X}}(\mathrm{Enc}(m'))| \le |\mathcal{T}_{Q_{A|X}}(\mathrm{Enc}(m))|2^{-n\frac{\delta}{2}}$

(Note that the conditions of the packing lemma are satisfied, because $H(X)_Q \geq I(X;A)_Q$).

Now define Dec: $\mathscr{A}^n \to [2^{nR}]$ by $\mathrm{Dec}(a^n) = m$ if m is the unique message such that $a^n \in \mathcal{T}_{Q_{A|X}}(\mathrm{Enc}(m))$, otherwise we set $\mathrm{Dec}(a^n) = 0$. For all $m \in [2^{nR}]$, we have

$$\sum_{a^n\colon \mathrm{Dec}(a^n)=m} P_{\mathsf{A}|\mathsf{X}}^{\times n}(a^n|\,\mathrm{Enc}(m)) = |\,\mathrm{Dec}^{-1}(m)|2^{-n(D(Q_{\mathsf{A}|\mathsf{X}}\|P_{\mathsf{A}|\mathsf{X}}|Q_{\mathsf{X}})+H(A|X)_Q)} \tag{59}$$

by Lemma 2.6 in [24] (using that $\mathrm{Enc}(m)$ are all of type Q_{X}). By definition of the decoder, we also have

$$\begin{aligned} |\,\mathrm{Dec}^{-1}(m)| &\geq |\mathcal{T}_{Q_{A|X}}(\mathrm{Enc}(m)) \setminus \bigcup_{m'\neq m} \mathcal{T}_{Q_{A|X}}(\mathrm{Enc}(m'))| &(60)\\ &\geq |\mathcal{T}_{Q_{\mathsf{A}|\mathsf{X}}}(\mathrm{Enc}(m))|(1-2^{-n\frac{\delta}{2}}) &(61)\\ &\geq (n+1)^{-|\mathscr{A}|}(1-2^{-n\frac{\delta}{2}})2^{nH(A|X)_Q} &(62)\end{aligned}$$

where (61) follows from the second property of Enc and (62) follows from (32). By combining (62) with (59) we conclude that (Enc, Dec) is a

$$\left(n, 2^{n(I(X;A)_Q-\delta)}, (n+1)^{-|\mathscr{A}|}(1-2^{-n\frac{\delta}{2}})2^{-nD(Q_{\mathsf{A}|\mathsf{X}}\|P_{\mathsf{A}|\mathsf{X}}|Q_{\mathsf{X}})}\right)$$

code. We finally choose $p(n) = 2(n+1)^{|\mathscr{A}|}$ which is a polynomial in n depending only on $|\mathscr{A}|$ and for $n \geq \frac{2}{\delta}$ we have $p(n)^{-1} \leq (n+1)^{-|\mathscr{A}|}(1-2^{-n\frac{\delta}{2}})$. This concludes the existence of an

$$\left(n, 2^{n(I(X;A)_Q-\delta)}, \frac{2^{-nD(Q_{\mathsf{A}|\mathsf{X}}\|P_{\mathsf{A}|\mathsf{X}}|Q_{\mathsf{X}})}}{p(n)}\right)$$

code. □

A.3 Proof of Lemma 5.8

We first repeat the statement of Lemma 5.8 for the reader's convenience.

Lemma 5.8. *For any list-decoding* $(n, 2^{nR}, 2^{nR_L}, 2^{-n\zeta_n})$ *code for* $P_{\mathsf{A}|\mathsf{X}}$*, we have*

$$\zeta_n \geq \min_{Q_{\mathsf{XA}}} \left[D(Q_{\mathsf{A}|\mathsf{X}}\|P_{\mathsf{A}|\mathsf{X}} \mid Q_{\mathsf{X}}) + \max\{R - R_L - I(X;A)_Q, 0\}\right] + O\left(\frac{\log n}{n}\right),$$

where the constant hidden in $O(\cdot)$ *depends only on* $|\mathscr{X}|$ *and* $|\mathscr{A}|$.

Proof. Let (Enc, Dec) be an $(n, 2^{nR}, 2^{nR_L}, 2^{-n\zeta})$ list-decoding code, i.e., $\mathrm{Enc} : [2^{nR}] \to \mathscr{X}^n$, $\mathrm{Dec}(a)$ is a subset of size 2^{nR_L} for all $a \in \mathscr{A}^n$, and for all $m \in [2^{nR}]$,

$$P_{\mathsf{A}|\mathsf{X}}^{\times n}(\mathrm{Dec}^{-1}(m)|\,\mathrm{Enc}(m)) \geq 2^{-n\zeta}. \tag{63}$$

By pigeon hole principle, there exist a type $Q \in \mathcal{P}_n(\mathscr{X})$ and a subset S of size $\frac{2^{nR}}{(n+1)^{|\mathscr{X}|}}$ of $[2^{nR}]$ such that $\mathrm{Enc}(m) \in \mathcal{T}_Q$ for all $m \in S$. Furthermore, for any $m \in S$, we have

$$\begin{aligned} 2^{-n\zeta} &\leq P_{\mathsf{A}|\mathsf{X}}^{\times n}(\mathrm{Dec}^{-1}(m)|\,\mathrm{Enc}(m)) &(64)\\ &= \sum_{Q_{\mathsf{A}|\mathsf{X}}} P_{\mathsf{A}|\mathsf{X}}^{\times n}(\mathcal{T}_Q(\mathrm{Enc}(m)) \cap \mathrm{Dec}^{-1}(m)|\,\mathrm{Enc}(m)) &(65)\\ &= \sum_{Q_{\mathsf{A}|\mathsf{X}}} 2^{-nD(Q_{\mathsf{A}|\mathsf{X}}\|P_{\mathsf{A}|\mathsf{X}}|Q)} \frac{|\mathcal{T}_Q(\mathrm{Enc}(m)) \cap \mathrm{Dec}^{-1}(m)|}{|\mathcal{T}_Q(\mathrm{Enc}(m))|}. &(66)\end{aligned}$$

Averaging the above inequality over all $m \in S$, we obtain that

$$2^{-n\zeta} \leq \frac{1}{|S|}\sum_{m\in S}\sum_{Q_{\mathsf{A}|\mathsf{X}}} 2^{-nD(Q_{\mathsf{A}|\mathsf{X}}\|P_{\mathsf{A}|\mathsf{X}}|Q)} \frac{|\mathcal{T}_Q(\mathrm{Enc}(m)) \cap \mathrm{Dec}^{-1}(m)|}{|\mathcal{T}_Q(\mathrm{Enc}(m))|} \tag{67}$$

Since there are at most $(n+1)^{|\mathscr{A}||\mathscr{X}|}$ conditional types $Q_{\mathsf{A|X}}$, by applying the pigeon hole principle once again, we derive that there exists $Q_{\mathsf{A|X}}$ such that

$$\frac{2^{-n\zeta}}{(n+1)^{|\mathscr{A}||\mathscr{X}|}} \leq 2^{-nD(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}}|Q)} \frac{1}{|S|}\sum_{m\in S}\frac{|\mathcal{T}_Q(\mathrm{Enc}(m))\cap \mathrm{Dec}^{-1}(m)|}{|\mathcal{T}_Q(\mathrm{Enc}(m))|} \tag{68}$$

We now provide two upper bounds on the right hand side of the above inequality. Note that

$$\frac{1}{|S|}\sum_{m\in S}\frac{|\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))\cap \mathrm{Dec}^{-1}(m)|}{|\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))|} \leq 1 \tag{69}$$

We also have

$$\begin{aligned}
\frac{1}{|S|}\sum_{m\in S}\frac{|\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))\cap \mathrm{Dec}^{-1}(m)|}{|\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))|} &\overset{(a)}{\leq} \frac{(n+1)^{|\mathscr{A}|}}{|S|2^{nH(Y|X)_Q}}\sum_{m\in S}|\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))\cap \mathrm{Dec}^{-1}(m)| && (70)\\
&\overset{(b)}{\leq} \frac{(n+1)^{|\mathscr{A}|+|\mathscr{X}|}}{2^{nR}2^{nH(Y|X)_Q}}\sum_{m\in S}|\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))\cap \mathrm{Dec}^{-1}(m)| && (71)\\
&\overset{(c)}{\leq} \frac{(n+1)^{|\mathscr{A}|+|\mathscr{X}|}}{2^{nR}2^{nH(Y|X)_Q}}|\mathcal{T}_{Q_{\mathsf{A}}}|2^{nR_L} && (72)\\
&\overset{(d)}{\leq} (n+1)^{|\mathscr{A}|+|\mathscr{X}|}2^{n(-R+R_L-H(A|X)_Q+H(A)_Q)} && (73)\\
&= (n+1)^{|\mathscr{A}|+|\mathscr{X}|}2^{n(-R+R_L+I(A;X)_Q)} && (74)
\end{aligned}$$

where (a) follows from (32), (b) follows since $|S|\geq \frac{2^{nR}}{(n+1)^{|\mathscr{X}|}}$, (c) follows since $\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))\cap\mathrm{Dec}^{-1}(m)\subset \mathcal{T}_{Q_{\mathsf{A}}}$ and every element of $\mathcal{T}_{Q_{\mathsf{A}}}$ appears in $\mathcal{T}_{Q_{\mathsf{A|X}}}(\mathrm{Enc}(m))\cap\mathrm{Dec}^{-1}(m)$ for at most 2^{nR_L} m, and (d) follows from (31). Combining above inequalities, we obtain that

$$\frac{2^{-n\zeta}}{(n+1)^{|\mathscr{A}||\mathscr{X}|}} \leq 2^{-nD(Q_{\mathsf{A|X}}\|P_{\mathsf{A|X}}|Q_{\mathsf{X}})}\min\left(1,(n+1)^{|\mathscr{A}|+|\mathscr{X}|}2^{n(-R+R_L+I(A;X)_Q)}\right) \tag{75}$$

Taking log from both sides of the above inequality results in the desired bound. □

B NPA hierarchy for LSSD

In this appendix, we describe the NPA hierarchy adapted to the LSSD setting. For more details on the original NPA hierarchy, see [28, 20].

Recall from Section 3 that LSSD game is played by two collaborating players, Alice and Bob, who receive inputs $a\in\mathscr{A}, b\in\mathscr{B}$ and must produce outputs $x_A, x_B\in\mathscr{X}$, respectively (see Fig. 2). Here we will consider the case when the game is defined by a joint probability distribution[3] P_{XAB} that describes how their inputs a,b are correlated with an external variable $x\in\mathscr{X}$ which they need to guess. The LSSD task is to produce outputs x_A and x_B such that $x_A = x_B = x$. We can equivalently describe this by the predicate $V(x,x_A,x_B) := \delta[x_A = x_B = x]$.

Depending on the physical scenario considered, Alice and Bob might share some resource that allows them to correlate their outputs. For the sake of generality, let $\mathcal{C}\subset\mathbb{R}^{\mathscr{X}\times\mathscr{X}\times\mathscr{A}\times\mathscr{B}}$ denote an arbitrary set of correlations they can utilize. We will treat each element $Q\in\mathcal{C}$ as a vector $\mathbb{R}^{\mathscr{X}\times\mathscr{X}\times\mathscr{A}\times\mathscr{B}}$ and write its entries as $Q(x_A,x_B|a,b)$ where $x_A,x_B\in\mathscr{X}, a\in\mathscr{A}, b\in\mathscr{B}$. This notation emphasizes that Q can also be interpreted as a stochastic matrix. Indeed, it will always be the case that $Q(x_A,x_B|a,b)\geq 0$ and $\sum_{x_A,x_B\in\mathscr{X}} Q(x_A,x_B|a,b) = 1$ for all $a\in\mathscr{A}$ and $b\in\mathscr{B}$. For example, when dealing with quantum strategies, Q has the following form, see Eq. (4):

$$Q(x_A,x_B|a,b) = \langle\psi|\big(M_{x_A}(a)\otimes N_{x_B}(b)\big)|\psi\rangle, \qquad \forall a\in\mathscr{A}, b\in\mathscr{B}, x_A,x_B\in\mathscr{X} \tag{76}$$

[3] More generally, P_{XAB} can be replaced by a quantum state ρ_{XAB}, see Section 3.2.

for some finite-dimensional bipartite Hilbert space $\mathcal{H} = \mathcal{H}_\mathsf{A} \otimes \mathcal{H}_\mathsf{B}$, pure state $|\psi\rangle \in \mathcal{H}$, and collections of projective measurements $\{M_x(a) : x \in \mathscr{X}\} \in \mathrm{PM}(\mathcal{H}_\mathsf{A})$ and $\{N_x(b) : x \in \mathscr{X}\} \in \mathrm{PM}(\mathcal{H}_\mathsf{B})$ on $\mathcal{H}_\mathsf{A}$ and $\mathcal{H}_\mathsf{B}$, respectively.[4] We will denote the set of all *quantum correlations* by $\mathcal{Q} \subset \mathbb{R}^{\mathscr{X}\times\mathscr{X}\times\mathscr{A}\times\mathscr{B}}$.

The winning probability of the LSSD game defined by a distribution P_XAB and played with assistance of correlations $\mathcal{C}$ is given by

$$\omega_\mathcal{C}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_\mathsf{XAB}} = \sup_{Q\in\mathcal{C}} \sum_{\substack{x,x_A,x_B\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} P_\mathsf{XAB}(x,a,b)V(x,x_A,x_B)Q(x_A,x_B|a,b). \tag{77}$$

If we let

$$K(x_A,x_B,a,b) := \sum_{x\in\mathscr{X}} P_\mathsf{XAB}(x,a,b)V(x,x_A,x_B), \tag{78}$$

we can rewrite Eq. (77) as

$$\omega_\mathcal{C}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_\mathsf{XAB}} = \sup_{Q\in\mathcal{C}} \langle K, Q\rangle \tag{79}$$

where we treat both K and Q as vectors in $\mathbb{R}^{\mathscr{X}\times\mathscr{X}\times\mathscr{A}\times\mathscr{B}}$ and

$$\langle K, Q\rangle := \sum_{\substack{x_A,x_B\in\mathscr{X}\\ a\in\mathscr{A},b\in\mathscr{B}}} K(x_A,x_B,a,b)Q(x_A,x_B|a,b). \tag{80}$$

To define a *commuting measurement strategy*, we relax the requirement that the finite-dimensional Hilbert space $\mathcal{H}$ has a tensor product structure. We consider a pure[5] state $|\psi\rangle \in \mathcal{H}$ and two collections of measurements on the whole of $\mathcal{H}$: one measurement $\{M_x(a) : x \in \mathscr{X}\}$ for each of Alice's inputs $a \in \mathscr{A}$ and one measurement $\{N_x(b) : x \in \mathscr{X}\}$ for each of Bob's inputs $b \in \mathscr{B}$. We require that these are orthogonal projective measurements, and that Alice's and Bob's operators pair-wise commute. Namely, for all $a \in \mathscr{A}$, $b \in \mathscr{B}$, $x_A \neq x'_A \in \mathscr{X}$ and $x_B \neq x'_B \in \mathscr{X}$ the measurement operators should satisfy

1. $M_{x_A}(a)^\dagger = M_{x_A}(a)$ and $N_{x_B}(b)^\dagger = N_{x_B}(b)$,
2. $M_{x_A}(a)M_{x'_A}(a) = 0$ and $N_{x_B}(b)N_{x'_B}(b) = 0$,
3. $\sum_{x_A\in\mathscr{X}} M_{x_A}(a) = \mathbb{I}$ and $\sum_{x_B\in\mathscr{X}} N_{x_B}(b) = \mathbb{I}$,
4. $[M_{x_A}(a), N_{x_B}(b)] = 0$.

The resulting correlation Q is then defined as

$$Q(x_A,x_B|a,b) := \langle\psi|M_{x_A}(a)N_{x_B}(b)|\psi\rangle, \qquad \forall a\in\mathscr{A}, b\in\mathscr{B}, x_A,x_B\in\mathscr{X}. \tag{81}$$

Compared to Eq. (76), here the two commuting sets of measurements are global since the underlying space $\mathcal{H}$ has no tensor product structure.

We will now construct a positive semidefinite matrix G whose entries contain the values $Q(x_A,x_B|a,b)$ from Eq. (81), and then impose linear constraints on G that capture the above conditions on the measurement operators $M_{x_A}(a)$ and $N_{x_B}(b)$. The rows and columns of G will be indexed by[6]

$$\Sigma_1 := \{\varepsilon\} \sqcup \Sigma_A \sqcup \Sigma_B \qquad \text{where} \qquad \Sigma_A := \mathscr{X}\times\mathscr{A}, \quad \Sigma_B := \mathscr{X}\times\mathscr{B}. \tag{82}$$

For each $s \in \Sigma_1$, define a vector in $\mathcal{H}$ as follows:

$$|\psi(s)\rangle := \begin{cases} |\psi\rangle & \text{if } s = \varepsilon, \\ M_x(a)|\psi\rangle & \text{if } s = (x,a) \in \Sigma_A, \\ N_x(b)|\psi\rangle & \text{if } s = (x,b) \in \Sigma_B, \end{cases} \tag{83}$$

[4] One can assume without loss of generality that the shared quantum state is pure and both measurements are orthogonal.

[5] The assumption that the state is pure and that the measurements are projective is without loss of generality.

[6] We assume the sets $\mathscr{A}$ and $\mathscr{B}$ are disjoint so that the disjoint union makes sense here.

and let $G \in \mathbb{R}^{\Sigma_1 \times \Sigma_1}$ be the Gram matrix of these vectors:

$$G_{s,t} := \langle \psi(s) | \psi(t) \rangle, \qquad \forall s, t \in \Sigma_1. \tag{84}$$

Since G is a Gram matrix, it is clearly positive semidefinite:

$$G \succeq 0. \tag{85}$$

Notice that G contains all of the values $Q(x_A, x_B | a, b)$ from Eq. (81), as well as some additional values such as $\langle \psi | \psi \rangle$, $\langle \psi | M_x(a) | \psi \rangle$, and others.

Because of the various relations among the measurement operators $M_{x_A}(a)$ and $N_{x_B}(b)$ listed earlier, the Gram matrix G is subject to the following linear constraints:

1. Since $|\psi\rangle$ is a normalized state, $\langle \psi | \psi \rangle = 1$ and thus
$$G_{\varepsilon,\varepsilon} = 1. \tag{86}$$

2. Due to the completeness relations $\sum_{x \in \mathscr{X}} M_x(a) = \mathbb{I} = \sum_{x \in \mathscr{X}} N_x(b)$, we have that for any vector $|v\rangle \in \mathcal{H}$, $\sum_{x \in \mathscr{X}} \langle \psi | M_x(a) | v \rangle = \langle \psi | v \rangle$ and $\sum_{x \in \mathscr{X}} \langle v | M_x(a) | \psi \rangle = \langle v | \psi \rangle$, and similarly for $N_x(b)$. Letting $|v\rangle = |\psi(s)\rangle$ for some $s \in \Sigma_1$, this translates to
$$\sum_{x \in \mathscr{X}} G_{(x,a),s} = G_{\varepsilon,s}, \qquad \sum_{x \in \mathscr{X}} G_{s,(x,a)} = G_{s,\varepsilon}, \qquad \forall a \in \mathscr{A}, s \in \Sigma_1, \tag{87}$$
$$\sum_{x \in \mathscr{X}} G_{(x,b),s} = G_{\varepsilon,s}, \qquad \sum_{x \in \mathscr{X}} G_{s,(x,b)} = G_{s,\varepsilon}, \qquad \forall b \in \mathscr{B}, s \in \Sigma_1. \tag{88}$$

3. Since within each measurement the projectors are orthogonal, we also have $\langle \psi | M_x(a) M_{x'}(a) | \psi \rangle = 0 = \langle \psi | N_x(b) N_{x'}(b) | \psi \rangle$ and thus
$$G_{(x,a),(x',a)} = 0, \qquad \forall x \neq x' \in \mathscr{X}, a \in \mathscr{A}, \tag{89}$$
$$G_{(x,b),(x',b)} = 0, \qquad \forall x \neq x' \in \mathscr{X}, b \in \mathscr{B}. \tag{90}$$

4. Since $M_x(a)$ are projectors, $\langle \psi | M_x(a) M_x(a) | \psi \rangle = \langle \psi | M_x(a) | \psi \rangle$ and likewise for $N_x(b)$, so
$$G_{(x,a),(x,a)} = G_{(x,a),\varepsilon} = G_{\varepsilon,(x,a)}, \qquad \forall x \in \mathscr{X}, a \in \mathscr{A}, \tag{91}$$
$$G_{(x,b),(x,b)} = G_{(x,b),\varepsilon} = G_{\varepsilon,(x,b)}, \qquad \forall x \in \mathscr{X}, b \in \mathscr{B}. \tag{92}$$

5. Since the two sets of projectors commute, $\langle \psi | M_{x_A}(a) N_{x_B}(b) | \psi \rangle = \langle \psi | N_{x_B}(b) M_{x_A}(a) | \psi \rangle$, we have
$$G_{(x_A,a),(x_B,b)} = G_{(x_B,b),(x_A,a)}, \qquad \forall x_A, x_B \in \mathscr{X}, a \in \mathscr{A}, b \in \mathscr{B}. \tag{93}$$

Let $\mathcal{Q}_1 \subset \mathbb{R}^{\mathscr{X} \times \mathscr{X} \times \mathscr{A} \times \mathscr{B}}$ denote the set of all correlations Q such that there exists a matrix $G \in \mathbb{R}^{\Sigma_1 \times \Sigma_1}$ which satisfies

$$G_{(x_A,a),(x_B,b)} = Q(x_A, x_B | a, b), \qquad \forall x_A, x_B \in \mathscr{X}, a \in \mathscr{A}, b \in \mathscr{B}, \tag{94}$$

as well as $G \succeq 0$ and the linear constraints in Eqs. (86) to (93). Note that deciding the membership of Q in $\mathcal{Q}_1$ is a semidefinite feasibility problem – it requires finding a positive semidefinite matrix $G \succeq 0$ subject to linear constraints.

Since the local measurement operators $M_{x_A}(a) \otimes \mathbb{I}$ and $\mathbb{I} \otimes N_{x_B}(b)$ commute, the original set of quantum correlations $\mathcal{Q}$ defined by Eq. (76) satisfies $\mathcal{Q} \subseteq \mathcal{Q}_1$. Therefore, based on Eq. (79),

$$\omega_{\mathrm{q}}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}} := \sup_{Q \in \mathcal{Q}} \langle K, Q \rangle \leq \sup_{Q \in \mathcal{Q}_1} \langle K, Q \rangle =: \omega_{\mathrm{q}_1}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}}, \tag{95}$$

where the vector $K \in \mathbb{R}^{\mathscr{X} \times \mathscr{X} \times \mathscr{A} \times \mathscr{B}}$ defined in Eq. (78) specifies the LSSD game in question. The value $\omega_{\mathrm{q}_1}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}}$ corresponds to the *first level* of the NPA hierarchy. We can compute it by a semidefinite program as follows. Define a symmetric matrix $H \in \mathbb{R}^{\Sigma_1 \times \Sigma_1}$ with entries

$$H_{(x_A,a),(x_B,b)} := H_{(x_B,b),(x_A,a)} := \frac{1}{2} K(x_A, x_B, a, b), \qquad \forall x_A, x_B \in \mathscr{X}, a \in \mathscr{A}, b \in \mathscr{B} \tag{96}$$

and 0 otherwise. Then $\langle K, Q\rangle = \operatorname{tr}(HG)$ is a linear function of G, so we can compute the value of $\omega_{\mathrm{q}_1}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}}$ via a semidefinite program that maximizes $\operatorname{tr}(HG)$ over all positive semidefinite matrices G satisfying the conditions listed above.

The *second level* of the NPA hierarchy is obtained by a similar SDP that involves a larger *extended* Gram matrix G whose rows and columns are indexed by[7]

$$\Sigma_2 := \Sigma_1 \sqcup (\Sigma_A \times \Sigma_A) \sqcup (\Sigma_A \times \Sigma_B) \sqcup (\Sigma_B \times \Sigma_B). \tag{97}$$

We extend the original set of vectors $|\psi(s)\rangle$ from Eq. (83) by defining new vectors for the remaining elements $s \in \Sigma_2 \setminus \Sigma_1$ as follows:

$$|\psi(s)\rangle := \begin{cases} M_x(a)M_{x'}(a')|\psi\rangle & \text{if } s = ((x,a),(x',a')) \in \Sigma_A \times \Sigma_A, \\ M_x(a)N_{x'}(b')|\psi\rangle & \text{if } s = ((x,a),(x',b')) \in \Sigma_A \times \Sigma_B, \\ N_x(b)N_{x'}(b')|\psi\rangle & \text{if } s = ((x,b),(x',b')) \in \Sigma_B \times \Sigma_B. \end{cases} \tag{98}$$

As before in Eq. (84), the entries of the extended G are also given by inner products $\langle\psi(s)|\psi(t)\rangle$ for all $s,t \in \Sigma_2$, and we impose additional linear constraints on them similar to those in Eqs. (86) to (93) to capture the fact that Alice and Bob's operators describe mutually commuting projective measurements.

We denote by $\mathcal{Q}_2 \subset \mathbb{R}^{\mathscr{X}\times\mathscr{X}\times\mathscr{A}\times\mathscr{B}}$ the set of all correlations Q for which there exists an extended Gram matrix $G \in \mathbb{R}^{\Sigma_2\times\Sigma_2}$ that agrees with Q on Σ_1, see Eq. (94), and which satisfies the linear constraints for the second level of the NPA hierarchy. Note that $\mathcal{Q}_2 \subseteq \mathcal{Q}_1$ since the second level imposes additional constraints compared to the first level. Intuitively, the *ℓ-th* level of the NPA hierarchy is obtained by considering the Gram matrix of the vectors of the level $\ell - 1$ plus new vectors obtained from products of ℓ projectors, see [28, 20] for a more formal description.

For our analysis in Section 4.1, we consider the SDP for an intermediate level of the NPA hierarchy between $\mathcal{Q}_1$ and $\mathcal{Q}_2$, where G is the Gram matrix for the set of vectors labelled by

$$\Sigma_{1+MN} := \Sigma_1 \sqcup (\Sigma_A \times \Sigma_B). \tag{99}$$

We define $\mathcal{Q}_{1+MN}$ analogously to $\mathcal{Q}_1$ and $\mathcal{Q}_2$. Since $\Sigma_1 \subset \Sigma_{1+MN} \subset \Sigma_2$, we have $\mathcal{Q}_1 \supseteq \mathcal{Q}_{1+MN} \supseteq \mathcal{Q}_2 \supseteq \mathcal{Q}$ and therefore

$$\sup_{Q\in\mathcal{Q}_{1+MN}} \langle K, Q\rangle =: \omega_{\mathrm{q}_{1+MN}}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}} \geq \omega_{\mathrm{q}_2}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}} \geq \omega_{\mathrm{q}}(\mathsf{X}|\mathsf{A};\mathsf{B})_{P_{\mathsf{XAB}}}. \tag{100}$$

C Three-party binary LSSD

In this appendix, we show (partially numerically) that there exist no probability distribution P_{XABC}, where x, a, b and c are all binary, such that the corresponding LSSD game can be won with higher probability using no-signalling strategies than with classical strategies. We get to this conclusion by showing that none of the no-signalling correlations at the extreme points of the no-signalling polytope can ever perform better than classical strategies.

In the next subsection we discuss some results on optimal classical and no-signalling strategies. These results allow us to discard some no-signalling strategies of which we know that they cannot perform better than classical strategies. For the strategies that are left, we turn to linear programming to numerically show that they also cannot perform better than classical.

C.1 Some results on optimal strategies

Multi-partite no-signalling correlations Up until now, we have only looked at correlations between two parties. However, the concepts of locality and no-signalling can be extended to any finite number of parties. We show how to do this extension for no-signalling correlations.

In the case of more than two parties, a correlation is no-signalling if no subset of parties J can collectively signal to the rest of the parties I. So the output of the parties indexed by I cannot depend on the input to the parties indexed by J.

[7] We omit $\Sigma_B \times \Sigma_A$ since Alice and Bob's operators commute.

Definition C.1 (Definition 11 in [22])**.** *An m-partite correlation $Q_{\mathsf{X}_1\cdots\mathsf{X}_m|\mathsf{A}_1\cdots\mathsf{A}_m}$ on $\mathscr{X}_1\times\cdots\times\mathscr{X}_m\times\mathscr{A}_1\times\cdots\times\mathscr{A}_m$ is called no-signalling if for any index set $I\subset\{1,\dots,m\}$ and its complement $J=\{1,\dots,m\}\setminus I$ it holds that*

$$\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a_J)=\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a'_J), \tag{101}$$

for all $x_I\in\mathscr{X}_I, a_I\in\mathscr{A}_I$ and $a_J,a'_J\in\mathscr{A}_J$.

The next lemma states that we can loosen the constraints a little and still be left with an equivalent definition of no-signalling. Specifically, it states that it is sufficient to require that any single party cannot signal to the rest.

Lemma C.2. *Suppose Q is a m-partite correlation satisfying Eq. (101) for all index sets I such that their complements J have cardinality 1 and for all $x_I\in\mathscr{X}_I, a_I\in\mathscr{A}_I$ and $a_J,a'_J\in\mathscr{A}_J$. Then Q is a no-signalling correlation.*

Proof. We prove this lemma by induction on the cardinality of the complement J of an index set I. If $|J|=1$, condition (101) holds by assumption. Now suppose $|J|=n$, and let $x_I\in\mathscr{X}_I, a_I\in\mathscr{A}_I$ and $a_J,a'_J\in\mathscr{A}_J$. Take $j\in J$ and let $J'=J\setminus\{j\}$. We now find

$$\begin{aligned}\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a_J)&=\sum_{x_{J'}\in\mathscr{X}_{J'}}\sum_{x_j\in\mathscr{X}_j} Q(x_I,x_{J'},x_j\big|a_I,a_{J'},a_j)\\&\overset{(i)}{=}\sum_{x_{J'}\in\mathscr{X}_{J'}}\sum_{x_j\in\mathscr{X}_j} Q(x_I,x_{J'},x_j\big|a_I,a_{J'},a'_j)\\&\overset{(ii)}{=}\sum_{x_{J'}\in\mathscr{X}_{J'}}\sum_{x_j\in\mathscr{X}_j} Q(x_I,x_{J'},x_j\big|a_I,a'_{J'},a'_j)\\&=\sum_{x_J\in\mathscr{X}_J} Q(x_I,x_J\big|a_I,a'_J),\end{aligned}$$

where (i) follows by assumption on Q and (ii) by induction (we are free to exchange the sums). □

This first lemma is an extension of the classical part of Lemma 3.2 in the paper by Majenz et al. [9]. It gives a list of all deterministic strategies (or more accurately: winning probability thereof) we need to consider in finding the optimal classical winning probability. The proof of this lemma relies on the relatively simple observation that the players should have equal output sets (sets consisting of all values they could possibly output according to their strategy).

Lemma C.3. *Let P_{XABC} be a probability distribution over $\mathscr{X}\times\mathscr{A}\times\mathscr{B}\times\mathscr{C}$ with $\mathscr{A}=\mathscr{B}=\mathscr{C}=\{0,1\}$ and $\mathscr{X}=[d]$, $d\geq 2$. The classical winning probability for P_{XABC} is given by*

$$\omega_c(\mathsf{X}\big|\mathsf{A};\mathsf{B};\mathsf{C})_P=\max_{\substack{s,t\\ s\neq t}}\max\left\{\begin{matrix}P_{\mathsf{X}}(s),\\ P_{\mathsf{XABC}}(s,0,0,0)+P_{\mathsf{XABC}}(t,1,1,1),\\ P_{\mathsf{XABC}}(s,1,0,0)+P_{\mathsf{XABC}}(t,0,1,1),\\ P_{\mathsf{XABC}}(s,0,1,0)+P_{\mathsf{XABC}}(t,1,0,1),\\ P_{\mathsf{XABC}}(s,0,0,1)+P_{\mathsf{XABC}}(t,1,1,0)\end{matrix}\right\}. \tag{102}$$

Proof. First, remember that we only have to consider deterministic strategies (see Section 3.1). Any deterministic strategy can be represented by three functions $f,g,h\colon\{0,1\}\to\mathscr{X}$. Given such a strategy, the probability of winning is given by

$$\sum_{x,a,b,c} P_{\mathsf{XABC}}(x,a,b,c)\delta[f(a)=g(b)=h(c)=x]=\sum_{a,b,c} P_{\mathsf{XABC}}(f(a),a,b,c)\delta[f(a)=g(b)=h(c)]. \tag{103}$$

Notice that there is always an optimal strategy such that $\{f(0),f(1)\}=\{g(0),g(1)\}=\{h(0),h(1)\}$. Suppose, for example, that for some a^*, we have that $f(a^*)\notin\{g(0),g(1)\}$. It follows that $\delta[f(a^*)=g(b)=h(c)]=0$ for all b,c. Changing Alice's output on input a^*, such that $f(a^*)\in\{g(0),g(1)\}$, causes

$\delta[f(a^*) = g(b) = h(c)]$ to possibly be equal to 1 for some b, c. This change introduces non-negative terms in the sum of Eq. (103), while not losing any others, thereby increasing the winning probability.

There are 5 possible ways in which we have $\{f(0), f(1)\} = \{g(0), g(1)\} = \{h(0), h(1)\}$. The first is that all players ignore their input and always output some fixed s. In this case, the probability of winning is given by

$$\sum_{a,b,c} P_{\mathsf{XABC}}(s, a, b, c) = P_X(s)\,,$$

yielding the first term in Eq. (102). The other 4 possibilities are when they all take their input into account:

- $f(0) = g(0) = h(0)$ and $f(1) = g(1) = h(1)$ or,
- $f(1) = g(0) = h(0)$ and $f(0) = g(1) = h(1)$ or,
- $f(0) = g(1) = h(0)$ and $f(1) = g(0) = h(1)$ or,
- $f(0) = g(0) = h(1)$ and $f(1) = g(1) = h(0)$.

defining $f(0) =: s$ and $f(1) =: t$, the winning probability in each of these cases is equal to a term in Eq. (102). □

Whereas the previous lemma reduced the number of interesting deterministic strategies, the next lemma and its corollary will do so for no-signalling strategies.

Lemma C.4. *Let P be a probability distribution over $\mathscr{X} \times \mathscr{A}_1 \times \cdots \times \mathscr{A}_m$ with $|\mathscr{X}| = d$ and $d \geq 2$. Let Q be a no-signalling strategy for which*

$$Q(x, \ldots, x \big| a_1, \ldots, a_m) \leq \frac{1}{d},$$

holds for all $x \in \mathscr{X}$ and $a_1 \in \mathscr{A}_1, \ldots, a_m \in \mathscr{A}_m$. Then its winning probability in the LSSD game defined by P is at most the best classical winning probability:

$$\sum_{\substack{x\in\mathscr{X}\\ a_1\in\mathscr{A}_1,\ldots,a_m\in\mathscr{A}_m}} P(x, a_1, \ldots, a_m) Q(x, \ldots, x \big| a_1, \ldots, a_m) \leq \omega_{\mathrm{c}}(\mathsf{X}\big|\mathsf{A}_1; \ldots; \mathsf{A}_m)_P.$$

Proof. The proof relies on the simple fact that the m players can always use deterministic strategies to win with at least probability $1/d$ by ignoring their inputs and guessing the value of x to be the one most likely in P. The probability that the referee picks a certain value x is given by $P(x) = \sum_{a\in\mathscr{A}_1\times\cdots\times\mathscr{A}_m} P(x, a)$ and since $\sum_x P(x) = 1$, there exists an $x^* \in \mathscr{X}$ such that $P(x^*) \geq 1/d$. We conclude that $\omega_{\mathrm{c}}(\mathsf{X}\big|\mathsf{A}_1; \ldots; \mathsf{A}_m)_P \geq 1/d$.

We use the previous argument to finish the proof:

$$\begin{aligned}&\sum_{\substack{x\in\mathscr{X}\\ a_1\in\mathscr{A}_1,\ldots,a_m\in\mathscr{A}_m}} P(x, a_1, \ldots, a_m) Q(x, \ldots, x \big| a_1, \ldots, a_m)\\ \leq &\frac{1}{d} \sum_{\substack{x\in\mathscr{X}\\ a_1\in\mathscr{A}_1,\ldots,a_m\in\mathscr{A}_m}} P(x, a_1, \ldots, a_m) = \frac{1}{d} \leq \omega_{\mathrm{c}}(\mathsf{X}\big|\mathsf{A}_1; \ldots; \mathsf{A}_m)_P.\end{aligned}$$

□

Corollary C.5. *Consider an LSSD problem with m players defined by a distribution P for which $\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}_1; \ldots; \mathsf{A}_m)_P < \omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A}_1; \ldots; \mathsf{A}_m)_P$. There is an optimal no-signalling strategy Q at one of the vertices of the no-signalling polytope, such that there exist $x \in \mathscr{X}$, with $|\mathscr{X}| = d$, and $a_1 \in \mathscr{A}_1, \ldots, a_m \in \mathscr{A}_m$ for which $Q(x, \ldots, x\big|a_1, \ldots, a_m) > 1/d$.*

Proof. Since the set of all no-signalling strategies is a convex polytope, and the winning probability of a no-signalling strategy is a linear function, we know that the optimal winning probability is achieved by a strategy Q at one of the vertices of the polytope (see Section 2.2). We also know that there exist $x \in \mathscr{X}$ and $a_1 \in \mathscr{A}_1, \ldots, a_m \in \mathscr{A}_m$ such that $Q(x, \ldots, x|a_1, \ldots, a_m) > 1/d$, because otherwise this strategy would not achieve winning probability higher than $\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}_1; \ldots; \mathsf{A}_m)_P$ by Lemma C.4. □

In the case of two players, we would now be done in showing that there is no binary LSSD game with a gap between no-signalling and classical winning probabilities, since all no-signalling correlations at the extreme points of the no-signalling polytope satisfy the conditions of Lemma C.3 [29, Theorem 1]. We will see in the next section that for three players, this is not the case. However, Corollary C.5 is still very useful as it eliminates many of the no-signalling strategies.

C.2 No gap between classical and no-signalling

Theorem C.6. *$\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A}; \mathsf{B}; \mathsf{C})_P = \omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}; \mathsf{B}; \mathsf{C})_P$ for all probability distributions P_{XABC} over binary inputs and outputs.*

Proof. Thanks to Eq. (9), we can equivalently show that

$$\sup_P \big(\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A}; \mathsf{B}; \mathsf{C})_P - \omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}; \mathsf{B}; \mathsf{C})_P\big) = 0.$$

Now we have turned the problem into an optimization problem. It is, however, not possible to solve this problem using a single linear program, since the target function is not linear: the target function is the maximum of the difference between two sets. Luckily, using Corollary C.5 and some additional tricks, we can solve this problem using multiple linear programs.

First of all, we note that the set of all probability distributions P_{XABC} forms a convex polytope in $\mathbb{R}^n$. The polytope is defined by the following linear constraints:

$$\forall x, a, b, c \;\; P_{\mathsf{XABC}}(x, a, b, c) \geq 0,$$

and

$$\sum_{x,a,b,c} P_{\mathsf{XABC}}(x, a, b, c) = 1.$$

Apart from the variables that describe a probability distribution, we also add two variables c_{d} and c_{ns} to the linear program, which represent $\omega_{\mathrm{c}}(\mathsf{X}|\mathsf{A}; \mathsf{B}; \mathsf{C})_P$ and $\omega_{\mathrm{ns}}(\mathsf{X}|\mathsf{A}; \mathsf{B}; \mathsf{C})_P$ respectively. These two variables should satisfy the following constraints:

$$c_{\mathrm{d}} \geq \sum_{x,a,b,c} P_{\mathsf{XABC}}(x, a, b, c) Q_{\mathrm{d}}(x, x, x|a, b, c),$$

for all deterministic strategies Q_{d} and

$$c_{\mathrm{ns}} \geq \sum_{x,a,b,c} P_{\mathsf{XABC}}(x, a, b, c) Q_{\mathrm{ns}}(x, x, x|a, b, c), \tag{104}$$

for all no-signalling strategies Q_{ns} at the vertices of the no-signalling polytope.

Now, the problem is to maximize $c_{\mathrm{ns}} - c_{\mathrm{d}}$, which is a linear function in two variables, so we can use a linear program. However, since we have not put an upper bound on c_{ns}, this problem is obviously unbounded. We can work around this issue by changing one of the constraints in Eq. (104) to an equality. Solving the linear program with one of these constraints set to an equality constraint gives us the maximum gap under the assumption that the corresponding no-signalling strategy is the best strategy. By considering all no-signalling strategies in this way we can find the maximum gap between classical and no-signalling winning probabilities.

All that is left is to find the no-signalling strategies at the extreme points of the no-signalling polytope. We can find them using a *Python* package called *cddlib*, which is based on a C package under the same name [30]. Similar to linear programs, this package can provide all vertices of the polytope corresponding to a given set of linear constraints. In our case the constraints say that the strategy Q_{ns} is a conditional probability distribution on $\mathscr{X}^3 \times \mathscr{A} \times \mathscr{B} \times \mathscr{C}$ and it is no-signalling (where we can use Lemma C.2

to omit redundant constraints). We find with "`three_player_polytope_extrema.py`" [19] that this no-signalling polytope has 53856 extreme points, which is in line with the findings of the paper by Pironio et al. [31, Section 2.2].

Since the above number of extremal no-signalling strategies is quite large, we would like to reduce it so that we need to solve fewer linear programs. Using Corollary C.5, there must be an optimal strategy of a specific form, which reduces the number of relevant no-signalling strategies from 53 856 to 174. In addition, we can also use Lemma C.3 to reduce the number of relevant deterministic strategies from $2^6 = 64$ to 10. This calculation is performed by "`filter_three_player_strategies.py`" [19].

Now that we have everything needed to find the maximum gap between binary three-party classical and no-signalling strategies, we use the *Mathematica* notebook "`Three-party binary LSSD.nb`" [19] to exactly solve the above 174 linear programs. In each case the optimal value is 0, meaning that there is no binary LSSD game for three players such that no-signalling resources improve its winning probability. □