Informationally restricted correlations: a general framework for classical and quantum systems

Armin Tavakoli[1,2], Emmanuel Zambrini Cruzeiro [3], Erik Woodhead[3], and Stefano Pironio[3]

[1]Département de Physique Appliquée, Université de Genève, CH-1211 Genève, Switzerland

[2]Institute for Quantum Optics and Quantum Information – IQOQI Vienna, Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria

[3]Laboratoire d'Information Quantique, CP 225, Université libre de Bruxelles (ULB), Av. F. D. Roosevelt 50, 1050 Bruxelles, Belgium

We introduce new methods and tools to study and characterise classical and quantum correlations emerging from prepare-and-measure experiments with informationally restricted communication. We consider the most general kind of informationally restricted correlations, namely the ones formed when the sender is allowed to prepare statistical mixtures of mixed states, showing that contrary to what happens in Bell nonlocality, mixed states can outperform pure ones. We then leverage these tools to derive device-independent witnesses of the information content of quantum communication, witnesses for different quantum information resources, and demonstrate that these methods can be used to develop a new avenue for semi-device independent random number generators.

1 Introduction

Consider an experiment of the kind illustrated in Fig. 1, where a sender, Alice, selects an input $x \in \{1, \dots, n_X\}$, encodes it into some physical system and transmits it to a receiver, Bob. Bob performs on the incoming system some measurement, represented by an input $y \in \{1, \dots, n_Y\}$, and gets an outcome $b \in \{1, \dots, n_B\}$. This *prepare-and-measure* experiment is ubiquitous in physics and forms the basis of many communication systems.

The transmission of physical messages between Alice and Bob serves to establish certain correlations between them. These correlations can be fully characterised by the set of probabilities $p(b|x, y)$ which represent how, for a given measurement y performed by Bob, his outcome b depends on Alice's input x. In full generality, we can associate to each input x selected by Alice a quantum state ρ_x and to each measurement y selected by Bob a positive operator-valued measure (POVM) $\{M_{b|y}\}_b$, so that we can write

$$p(b|x, y) = \mathrm{Tr}\big[\rho_x M_{b|y}\big]. \tag{1}$$

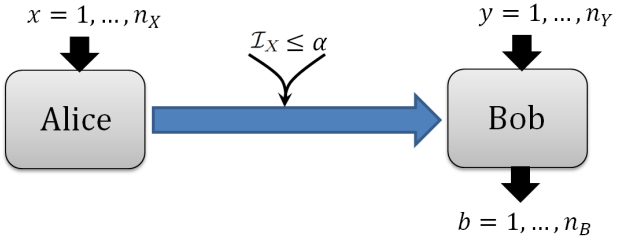

Figure 1: Illustration of prepare-and-measure experiment in which the communication is restricted to carry at most α bits of information about X

The special case where Alice and Bob are manipulating classical systems, instead of quantum ones, can be treated analogously by taking the states and measurements to be diagonal in the same basis:

$$\rho_x = \sum_m p(m|x)|m\rangle\langle m|, \tag{2}$$

$$M_{b|y} = \sum_m p(b|y, m)|m\rangle\langle m|, \tag{3}$$

where the variable m denotes the possible values of Alice's classical message.

In this work, we are interested in characterising what kind of correlations between Alice and Bob, i.e., which set of probabilities $p(b|x, y)$, are possible under the sole restriction of some constraint on the communication capabilities of the classical or quantum systems ρ_x emitted by Alice.

To date, the most commonly considered communication constraint in this setting has been a bound on the Hilbert-space dimension d of the emitted quantum systems (corresponding to the number of different possible messages m in the classical case). In the last two decades, a large body of works has investigated the interplay between correlations and dimension in this setting [1–11]. This line of work led, e.g., to the notion of dimension witnesses [5, 12] and to semi-device-independent protocols [13], such as randomness generation [14], quantum key distribution [15], and self-testing [16]. Evidently, a quantum or clas-

sical d-dimensional system can carry at most $\log_2 d$ bits of information and thus a bound on the dimension represents an information constraint. However, the physical dimension does not provide a complete picture of the concept of information. For instance, there are many systems of dimension $d' > d$ that do not carry more than $\log_2 d$ bits of information. Furthermore, in practical semi-device-independent protocols, assuming an exact bound on the dimension may be problematic to justify (a fact that has partly motivated other recent approaches [17–20]). A more satisfying and practically relevant approach may be to constrain the communication in terms of a continuous information measure.

Following [21], we specify here the communication constraint on the physical systems ρ_x received by Bob as an upper bound

$$P_{\mathrm{g}}(X|B) \leq G \tag{4}$$

on the guessing probability of the input X[1],

$$P_{\mathrm{g}}(X|B) = \max_{\{N_x\}_x} \sum_x q_x \operatorname{Tr}\left[\rho_x N_x\right], \tag{5}$$

where the maximisation is taken over all possible POVMs $\{N_x\}_x$ on the physical system that Bob receives. This guessing probability represents the optimum average probability with which Bob would correctly guess Alice's input x if he were to perform an ideal POVM on the incoming messages ρ_x, assuming that Alice selects each input with prior probability q_x [22]. The guessing probability $P_{\mathrm{g}}(X|B)$ can take any value from $P_{\mathrm{g}}(X|B) = \max_x q_x$ when the states are the same and hence carry no information about Alice's input (in which case Bob's best guessing strategy is to output the most probable input x according to q_x), up to $P_{\mathrm{g}}(X|B) = 1$ when they are perfectly distinguishable. Different communication restrictions on the messages can be specified by the choice of the bound G, as well as the input probabilities q_x.

Equivalently, one can express the communication restriction (4) as an upper bound $\mathcal{I}(X|B) \leq \alpha$ on the information measure

$$\mathcal{I}(X|B) = H_{\min}(X) - H_{\min}(X|B), \tag{6}$$

defined in term of the min-entropies $H_{\min}(X) = -\log_2\big(\max_x\{q_x\}\big)$ and $H_{\min}(X|B) = -\log_2\big(P_{\mathrm{g}}(X|B)\big)$. This quantity, expressed in bits, ranges from $\mathcal{I}(X|B) = 0$ when the states carry no information about Alice's input, up to $\mathcal{I}(X|B) = \log_2(n_X)$ bits, when they are perfectly distinguishable and chosen equiprobably, i.e., $q_x = 1/n_X$. There exist in principle a number of different other information measures that we could consider (see e.g. [23]) but the one we choose has a clear operational meaning and is convenient to work with.

[1] X and B are random variables.

We emphasise that q_x does not represent the actual prior from which Alice selects her input. Instead, it is a part of the assumption on Alice's source. Indeed, we are interested here in constraining *conditional* probabilities $p(b|x,y)$ which therefore do not depend on any prior probabilities with which Alice's input x and Bob's inputs y are selected. To constrain these conditional probabilities $p(b|x,y)$ we make a certain assumption about the source, specifically about the information-capacity of the ensemble of states $\{\rho_x\}$ it prepares. This information-capacity can be defined in various ways. The definition we chose here can be thought of as a fictious game: how well the classical variable x could correctly be identified by Bob if it were encoded by Alice in the state ρ_x and chosen with probability q_x. In the same way that the optimal measurement performed to guess x in this fictious game is not necessarily the same as the actual measurements taking place in Bob's measurement apparatuses and leading to the conditional probabilities $p(b|x,y)$, the prior probabilities q_x need not be the same as the prior probabilities p_x used by Alice to select her input in any actual scenario or protocol involving the conditional probabilities $p(b|x,y)$. In particular, a given scenario, say a DIRNG protocol where Alice's select her input with some fixed probabilities p_x, can be analyzed using different choices of q_x, this simply correspond to different assumptions about the source.

Note that one can also completely eliminate q_x from the analysis by choosing the uniform prior $q_x = 1/n_X$ (where n_X denotes the number of inputs of Alice). For a bound of the form $\mathcal{I}(X|B) \leq \alpha$, this corresponds to the strongest assumption on the source in the sense that $\mathcal{I}(X|B)_{\mathrm{uni}} \leq \alpha$ implies $\mathcal{I}(X|B)_{\mathrm{bias}} \leq \alpha$ for any choice of biased distribution q_x, as shown in [24].

Finally, we remark that instead of viewing the bound (4) as characterizing the preparations of Alice, we can alternatively view it as a constraint on the channel relating Alice to Bob. Indeed, $\epsilon = 1 - G$ can be understood as an upper-bound on the average[2] error through which a classical message of size n_X can be communicated in one shot through the channel for whatever encoding Alice may choose [25].

We develop here a versatile toolbox for characterising the set of probabilities $p(b|x,y)$ that are possible given arbitrary information constraints $P_{\mathrm{g}}(X|B) \leq G$ (or, equivalently, $\mathcal{I}(X|B) \leq \alpha$). Our approach is fully general and does not make any assumptions about the states and measurements beyond the information constraint, and in particular no assumptions about their dimension. In the classical case, we provide a characterisation of the set of informationally restricted correlations in terms of linear programming and in the quantum case through a hierarchy of semidefinite programming relaxations. We also show, in analogy

[2] The reference [25] defines the error ϵ for uniform prior, but one can also generalize this concept for abitrary priors q_x.

with the dimension bounded case, how to apply our methods to construct device-independent witnesses of communication (quantified in terms of our information measure), resource inequalities for classical and quantum systems carrying one bit of information, and semi-device-independent random number generation (RNG) protocols. In particular, we will show concrete examples of high-rate RNG and also demonstrate that data obtained in RNG experiments assuming a 1-qubit bound can be recycled to certify the same amount of randomness under the strictly weaker assumption of a 1-bit information bound.

Our work can be seen as a follow-up to Ref. [21], which originally proposed to replace the dimension bound in semi-device-independent scenarios by the information bound $P_{\mathrm{g}}(X|B) \leq G$ (or $\mathcal{I}(X|B) \leq \alpha$) considered here. However, [21] implicitly modelled the correlations established between Alice and Bob as statistical mixtures

$$p(b|x,y) = \sum_{\lambda} p(\lambda) p_{\lambda}(b|x,y) \tag{7}$$

of correlations $p_{\lambda}(b|x,y)$ obtained by measuring *pure* states:

$$p_{\lambda}(b|x,y) = \langle \psi_x^{(\lambda)} | M_{b|y}^{(\lambda)} | \psi_x^{(\lambda)} \rangle. \tag{8}$$

The guessing probability constraining the communication was then defined as the following averaged quantity over the classical shared variable λ:

$$P_{\mathrm{g}}(X|B) = \sum_{\lambda} p(\lambda) \max_{\{N_x^{(\lambda)}\}_x} \sum_x q_x \langle \psi_x^{(\lambda)} | N_x^{(\lambda)} | \psi_x^{(\lambda)} \rangle. \tag{9}$$

Similarly, in the classical case, the correlations between Alice and Bob were modelled as statistical mixtures of correlations

$$p_{\lambda}(b|x,y) = \sum_m \delta(m, m_x^{(\lambda)})\, p_{\lambda}(b|y,m)\,, \tag{10}$$

established by sending *deterministic* messages $m_x^{(\lambda)}$ for given x and λ.

The sets of such pure state correlations, in the quantum case, or deterministic correlations, in the classical case, compatible with a given communication constraint $P_{\mathrm{g}}(X|B) \leq G$ are easily seen to be particular subcases of the more general correlations that we consider here. Indeed, they can be obtained by assuming the states and measurements in (1) to take the following specific forms

$$\rho_x = \sum_{\lambda} p(\lambda)\, |\lambda\rangle\langle\lambda| \otimes |\psi_x^{(\lambda)}\rangle\langle\psi_x^{(\lambda)}|, \tag{11}$$

$$M_{b|y} = \sum_{\lambda} |\lambda\rangle\langle\lambda| \otimes M_{b|y}^{(\lambda)} \tag{12}$$

in the quantum case, and

$$\rho_x = \sum_{\lambda,m} p(\lambda) p_{\lambda}(m|x)\, |\lambda\rangle\langle\lambda| \otimes |m\rangle\langle m|, \tag{13}$$

$$M_{b|y} = \sum_{\lambda,m} p(\lambda) p_{\lambda}(b|y,m)\, |\lambda\rangle\langle\lambda| \otimes |m\rangle\langle m| \tag{14}$$

in the classical case, which recovers both the convex sum (7) and the average guessing probability (9). Interestingly, while in more traditional works on correlations, such as in the study of Bell nonlocality [26], statistical mixtures of pure states (or of deterministic correlations) generate the full set of correlations, they only represent a proper subset of the possible correlations in our information-restricted setting. This is because given a set of arbitrary states ρ_x satisfying the information constraint $P_{\mathrm{g}}(X|B) \leq G$, one can generally not re-interpret them as a mixtures of pure states without increasing their distinguishability, hence potentially violating the condition $P_{\mathrm{g}}(X|B) \leq G$.

The formulation we consider here is fully general and does not make any implicit assumption on the structures of the states appearing in the definition (1). Throughout the paper, we will compare our results to those that would be obtained under the pure-state approach of [21] in order to illustrate the differences in the two formulations.

2 Basic properties and simple scenarios

In the following, we refer to the prepare-and-measure scenario of Fig. 1, with n_X inputs for Alice, n_Y inputs for Bob, and n_B outputs, as a (n_X, n_Y, n_B)-scenario. Given an information bound specified by a probability distribution q_x and a number $G \in [\max_x\{q_x\}, 1]$, we denote by $\mathcal{Q}$ the set of quantum correlations compatible with that information bound, i.e., the set of probability distributions $p(b|x,y)$ for which there exist states ρ_x and measurement operators $M_{b|y}$ defined on some Hilbert space of arbitrary dimension d that satisfy the Born rule (1) and the constraint (4). Similarly, $\mathcal{C}$ denotes the set of classical correlations, i.e., satisfying in addition (2)–(3). By plugging this specific form for the states and measurements in (1) and (4), classical correlations can also be defined as those that can be written as

$$p(b|x,y) = \sum_m p(m|x) p(b|y,m) \tag{15}$$

and satisfying the information constraint

$$P_{\mathrm{g}}(X|B) = \sum_m \max_x q_x p(m|x) \leq G, \tag{16}$$

since the optimal POVM $\{N_x\}$ in this case is the one that reads the classical message m and outputs the value x that maximises $q_x p(m|x)$.

The sets $\mathcal{Q}$ and $\mathcal{C}$ are easily seen to be convex, using a construction akin to (11) and (12). That is, we can without loss of generality assume that the states sent by Alice and the measurements performed by Bob depend on some shared randomness λ (independent of x).

As a consequence, when writing the correlations explicitly as a convex sum (7), we can without loss of generality assume Bob's measurements to be extremal conditioned on λ: if the measurements of Bob depend on some local randomness, we can always incorporate it instead in the shared randomness λ. In the classical case $\mathcal{C}$, this means that we can without loss of generality assume Bob's classical response $p_\lambda(b|y,m)$ to be deterministic, i.e., such that $p_\lambda(b|y,m) \in \{0,1\}$. However, as noted earlier, we cannot without loss of generality assume the states to be pure (or deterministic in the classical case) when conditioned on λ as rewriting a mixed-state as a convex combination of pure states could violate the original guessing probability bound.

The sets $\mathcal{Q}$ and $\mathcal{C}$ satisfy certain basic inequalities. Obviously, since the $p(b|x,y)$ are probabilities, they must by definition satisfy the positivity and normalisation conditions

$$p(b|x,y) \geq 0, \qquad \forall b,x,y \tag{17}$$

and

$$\sum_b p(b|x,y) = 1, \qquad \forall x,y. \tag{18}$$

In addition, since post-processing cannot improve the distinguishability between messages and since all measurements $\{M_{b|y}\}_b$ of Bob can be viewed as (typically suboptimal) information-extraction POVMs, it holds that

$$\sum_b \max_x q_x p(b|x,y) \leq G, \qquad \forall y, \tag{19}$$

since, as in (16), when Bob gets the result b when he performs the measurement corresponding to input y, his best guess of x is the value that maximises $q_x p(b|x,y)$. This last constraint can explicitly be rewritten as a series of linear inequalities

$$\sum_b q_{x_b} p(b|x_b,y) \leq G, \forall y, \forall \boldsymbol{x} = (x_1, x_2, \ldots, x_{n_B}) \tag{20}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n_B}) \in \{1, \ldots, n_X\}^{\times n_B}$ assigns to each output b a value x_b.

We remark that, though it is harmless to specify them, not all of the inequalities (20) are always relevant as they may already be implied by normalisation and positivity of the probabilities alone (as well potentially as constraints specific to $\mathcal{C}$ and $\mathcal{Q}$). Precisely which ones are redundant depends on the upper bound G chosen. The instances with all the components of $\boldsymbol{x}$ equal ($x_1 = x_2 = \cdots = x_{n_B}$) in particular are always redundant as the left side of (20) is in these cases always upper bounded by the smallest possible value, $\max_x\{q_x\}$, of the guessing probability. At the opposite extreme, (19) always becomes redundant entirely for sufficiently high G when Alice's device has more inputs than Bob's has outcomes. This, supposing we label Alice's inputs so that $q_1 \geq q_2 \geq \cdots \geq q_{n_X}$, is because the left side of (19) is also always bounded by

$$\sum_b \max_x q_x p(b|x,y) \leq \sum_{x=1}^{n_B} q_x, \tag{21}$$

which is strictly less than one if Alice has more than n_B inputs that are used with nonzero probability.

The set of correlations satisfying Eqs. (17), (18), and (19) is a polytope $\mathcal{G}$. The polytope $\mathcal{G}$ can be interpreted as the set of correlations attainable under informational restrictions when no assumption is made on the underlying physical theory. Therefore, recalling also that the classical set is contained in the quantum set, we have the inclusions $\mathcal{C} \subseteq \mathcal{Q} \subseteq \mathcal{G}$.

An important first step in semi-device-independent approaches is to establish that one can distinguish between classical and quantum correlations, i.e., that $\mathcal{C} \subset \mathcal{Q}$. We show here below that in the simplest case of communication experiments with only two inputs on Alice ($n_X = 2$), the classical, quantum and theory-independent sets are identical ($\mathcal{C} = \mathcal{Q} = \mathcal{G}$). Notably, this stands in contrast to other established approaches to semi-device-independence [17, 18]. Later, we will find that $\mathcal{C} \subset \mathcal{Q}$ indeed is possible when Alice has more than two inputs. In sections 3 and 4 we describe how to characterise the classical set and quantum set, respectively, in a general and systematic manner.

2.1 $\mathcal{C} = \mathcal{Q} = \mathcal{G}$ when Alice has $n_X = 2$ inputs

We show that for $n_X = 2$ it holds that $\mathcal{C} = \mathcal{Q} = \mathcal{G}$ by proving that every $p(b|x,y) \in \mathcal{G}$ admits a classical model. To this end, note that the constraints Eqs. (17)–(19) are decoupled with respect to y. In other words, for each individual value of y, we obtain a separate polytope and the full set of probabilities is just the Cartesian product of the n_Y identical polytopes corresponding to the individual values of y. We derive the vertices of these polytopes in Appendix A. For $n_B = 3$ (which is representative), up to permutations of Bob's outputs they are

$$\boldsymbol{v}_1(y) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \tag{22}$$

$$\boldsymbol{v}_2(y) = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1-G}{q_2} & 1-\frac{1-G}{q_2} & 0 \end{pmatrix}, \tag{23}$$

$$\boldsymbol{v}_3(y) = \begin{pmatrix} \frac{1-G}{q_1} & 1-\frac{1-G}{q_1} & 0 \\ 1 & 0 & 0 \end{pmatrix}, \tag{24}$$

$$\boldsymbol{v}_4(y) = \begin{pmatrix} \frac{1-G}{q_1} & 1-\frac{1-G}{q_1} & 0 \\ \frac{1-G}{q_2} & 0 & 1-\frac{1-G}{q_2} \end{pmatrix}, \tag{25}$$

where we use a matrix notation

$$\boldsymbol{v}_j(y) = \begin{pmatrix} p(1|1,y) & p(2|1,y) & \cdots \\ p(1|2,y) & p(2|2,y) & \cdots \end{pmatrix} \tag{26}$$

to summarise the probabilities $p(b|x,y)$ defining each vertex $\boldsymbol{v}_j$. The vertices for $n_B \neq 3$ are trivial variations of those above: for $n_B > 3$ the vertices are the

same except with additional columns of zeros while for $n_B < 3$ we simply discard the vertices that have more than n_B columns with nonzero entries in them.

Crucially, all the vertices $\boldsymbol{v}_1(y)$–$\boldsymbol{v}_4(y)$, including all their permutations, can be generated by performing different measurements on the same two commuting (classical) states

$$\rho_1 = \frac{1-G}{q_1}|0\rangle\langle 0| + \left(1 - \frac{1-G}{q_1}\right)|1\rangle\langle 1|, \quad (27)$$

$$\rho_2 = \frac{1-G}{q_2}|0\rangle\langle 0| + \left(1 - \frac{1-G}{q_2}\right)|2\rangle\langle 2|. \quad (28)$$

For example, the vertex $\boldsymbol{v}_3(y)$ is obtained by measuring $\{M_{b|y}\}$ with

$$M_{1|y} = |0\rangle\langle 0| + |2\rangle\langle 2|, \quad (29)$$

$$M_{2|y} = |1\rangle\langle 1|, \quad (30)$$

$$M_{3|y} = 0. \quad (31)$$

Furthermore, any convex mixtures of vertices of the kind above, which is to say, any probability $p(b|x,y)$ satisfying the conditions (17)–(19) above, can be generated by performing the corresponding convex mixtures of POVMs on Bob's side. We conclude that Eqs. (17), (18), and (19) completely characterise both $\mathcal{C}$, $\mathcal{Q}$ and $\mathcal{G}$.

2.2 Inequivalence of general correlations and pure-state correlations

Following [21], we denote by $\mathcal{Q}_{\text{pure}} \subseteq \mathcal{Q}$ the subset of $\mathcal{Q}$ consisting of convex combination of pure-state correlations (8) and $\mathcal{C}_{\text{det}} \subseteq \mathcal{C}$ the subset of $\mathcal{C}$ consisting of convex combinations of deterministic classical correlations (10). As we show below, already for the simplest communication scenario ($n_X = 2$), we can distinguish between $\mathcal{Q}$ and $\mathcal{Q}_{\text{pure}}$ as well as between $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$, i.e., $\mathcal{Q}_{\text{pure}} \subset \mathcal{Q}$ and $\mathcal{C}_{\text{det}} \subset \mathcal{C}$. Note, though, that the relation between $\mathcal{Q}_{\text{pure}}$ and $\mathcal{C}$ is more complex. We will see that in the simple scenario below that $\mathcal{Q}_{\text{pure}} \subset \mathcal{C}$. But in other scenarios one can have correlations in $\mathcal{Q}_{\text{pure}}$ that are outside $\mathcal{C}$ so that the two sets intersect, but none is strictly contained in the other. This justifies looking at the larger quantum set $\mathcal{Q}$, which by definition always satisfies $\mathcal{C} \subseteq \mathcal{Q}$ and thus can never be outperformed using classical correlations.

Before looking at the general $n_X = 2$ case, let us first consider the exceptional situation that Alice's inputs are equiprobable ($q_1 = q_2 = 1/2$). The states (27) and (28) become

$$\rho_1 = 2(1-G)|0\rangle\langle 0| + (2G-1)|1\rangle\langle 1|, \quad (32)$$

$$\rho_2 = 2(1-G)|0\rangle\langle 0| + (2G-1)|2\rangle\langle 2|. \quad (33)$$

Consider now a deterministic classical strategy with one bit of shared randomness. Specifically, Alice receives either $\lambda = 1$ with probability $p(1) = 2(1-G)$ or $\lambda = 2$ with probability $p(2) = 2G - 1$. If $\lambda = 1$ Alice prepares the state $|0\rangle\langle 0|$, while if $\lambda = 2$ Alice prepares the state $|x\rangle\langle x|$ depending on her input $x \in \{1, 2\}$. This strategy generates the same states as (32) and (33) on average and the average guessing probability is still G. Thus, all correlations in $\mathcal{G}$ can be obtained and one finds no difference between the various sets: $\mathcal{C}_{\text{det}} = \mathcal{C} = \mathcal{Q}_{\text{pure}} = \mathcal{Q} = \mathcal{G}$.

In contrast, whenever the prior is biased ($q_1 \neq q_2$), we find that the pure-state correlations and the general correlations are inequivalent (see Fig. 2). Considering the scenario $(n_X, n_Y, n_B) = (2, 1, 2)$, the correlations can be characterised in terms of the expectation values

$$E_x = p(1|x) - p(2|x) \quad (34)$$

for $x = 1, 2$, where we have omitted y due to its fixed value. In Appendix B, we show that the nontrivial facets of $\mathcal{C}$ and $\mathcal{Q}$ are

$$|q_1 E_1 - q_2 E_2| \leq 2G - 1, \quad (35)$$

in terms of the guessing probability bound G. The facets of $\mathcal{C}_{\text{det}}$ are likewise straightforward to derive due to the small number of possible deterministic strategies. We do this in Appendix B and find that the nontrivial facets are

$$|E_1 - E_2| \leq 2\frac{G - q_{\max}}{q_{\min}}, \quad (36)$$

where $q_{\max} = \max(q_1, q_2)$ and $q_{\min} = \min(q_1, q_2)$. Whenever $q_1 \neq q_2$ this bounds a strictly smaller set than (35).

Finally, we derive the exact boundaries of the set $\mathcal{Q}_{\text{pure}}$ in Appendix B. Unlike the classical sets and $\mathcal{Q}$, this set is not a polytope. Aside from the trivial constraints $|E_x| \leq 1$, it is bounded by an infinite family,

$$|c_1 E_1 - c_2 E_2| \leq \sqrt{1 - \frac{4c_1 c_2}{q_1 q_2} G(1-G)}, \quad (37)$$

of linear inequalities, for parameters c_1 and c_2 satisfying $c_1 + c_2 = 1$ in the range $q_{\min} \leq c_1, c_2 \leq q_{\max}$. This set is larger than $\mathcal{C}_{\text{det}}$ but smaller than $\mathcal{C}$ and $\mathcal{Q}$. Note that at the extreme $c_1 = q_1$, (37) reduces to (35). Hence, two flat parts of the boundary of $\mathcal{Q}_{\text{pure}}$ (see Fig. 2) coincide with the nontrivial facets of $\mathcal{Q}$.

3 Characterising classical correlations

In this section, we explain how one can systematically determine the boundaries of the classical set $\mathcal{C}$, which is a polytope; the characterisation of the deterministic set $\mathcal{C}_{\text{det}}$ was already addressed in [21]. We then apply our method to explicitly derive the boundaries of $\mathcal{C}$ in the $(3, 2, 2)$ scenario assuming Alice's inputs are chosen equiprobably, finding that $\mathcal{C}$ is strictly larger than $\mathcal{C}_{\text{det}}$ in this case. This differs from the case with two inputs considered earlier, where $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$ were

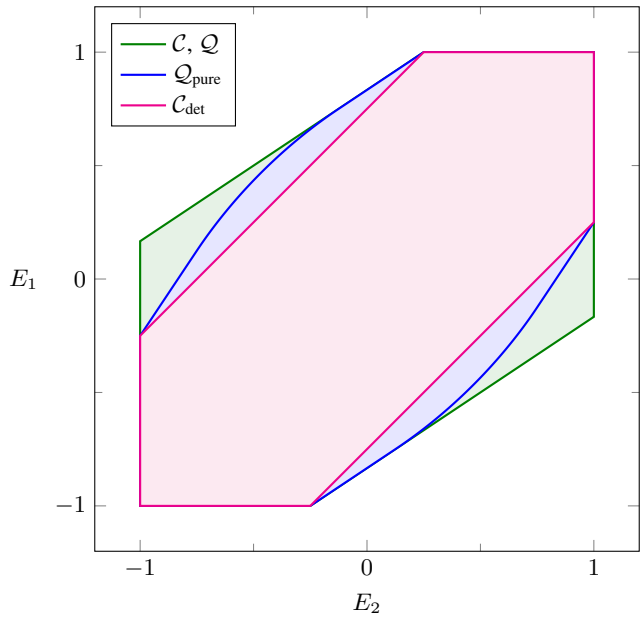

Figure 2: Informationally restricted correlations in the $(2, 1, 2)$ scenario with prior probabilities $(q_1, q_2) = (0.6, 0.4)$ and a bound $G = 3/4$ on the guessing probability (corresponding to $\mathcal{I}(X|B) \leq \log_2(5) - 2$ bits of information). The possible correlations are illustrated for deterministic classical strategies (magenta), deterministic quantum strategies (blue) and for classical and quantum stochastic strategies (green), which are the same in this case.

only found to be different when Alice's inputs are not equiprobable. Finally we also point out how one can, alternatively, generally test by linear programming whether a correlation is in $\mathcal{C}$ or not without explicitly needing to determine its boundaries.

3.1 Identifying the boundaries of $\mathcal{C}$

3.1.1 General method

The classical set $\mathcal{C}$ is, as mentioned above and as we have seen explicitly for $n_X = 2$ in the previous section, a polytope and we could in principle characterise it by determining its facets for any given upper bound G on the guessing probability. This direct approach would, however, require us to rederive the facets of $\mathcal{C}$ for each value of G that we may be interested in. To avoid this we instead consider a related but different set, which we call $\mathcal{C}^+$, of possible pairs $\big(p(b|x,y), G\big)$ of probability distributions $p(b|x,y)$ and guessing probability bounds G that are compatible with (15) and (16), which we repeat here for convenience:

$$p(b|x,y) = \sum_m p(m|x)p(b|y,m), \tag{38}$$

$$G \geq \sum_m \max_x q_x p(m|x). \tag{39}$$

Casting the problem in this way allows us to derive the boundaries of the classical set while leaving G as a free variable.

The set $\mathcal{C}^+$ is clearly convex, as it is easily seen that $\big(q\,p_1(b|x,y) + (1-q)\,p_2(b|x,y),\ q\,G_1 + (1-q)\,G_2\big)$ belongs to it if $(p_1(b|x,y), G_1)$ and $(p_2(b|x,y), G_2)$ do. To characterise it, it is thus sufficient to characterise its extreme points and take their convex hull.

As explained at the beginning of Section 2, remember that the extremal points of $\mathcal{C}$ have deterministic response probabilities for Bob: $p(b|y,m) \in \{0, 1\}$. If we fix such a deterministic response for Bob, the probabilities $p(b|x,y)$ are then entirely determined by the probability distribution $p(m|x)$ of Alice's messages. Those are simply constrained by

$$G - \sum_m \max_x q_x p(m|x) \geq 0, \tag{40}$$

$$p(m|x) \geq 0, \tag{41}$$

$$\sum_m p(m|x) = 1, \tag{42}$$

which represents a finite set of linear inequalities for the set $\mathcal{M}^+$ of possible pairs $\big(p(m|x), G\big)$ of message probabilities and guessing probability bounds. The set $\mathcal{M}^+$ is thus a polyhedron, i.e., an object like a polytope except that it is not necessarily bounded[3]. Explicitly, this is a set $\mathcal{P} = \{\boldsymbol{p}\}$ of points that can be generated from a finite number of vertices $\boldsymbol{v}_i \in \mathcal{V}$ and conic generators $\boldsymbol{w}_j \in \mathcal{W}$, i.e,

$$\mathcal{P} = \text{Conv}(\mathcal{V}) + \text{Cone}(\mathcal{W}) \tag{43}$$

or, more explicitly, the set of points $\{\boldsymbol{p}\}$ that can be expressed as

$$\boldsymbol{p} = \sum_i \lambda_i \boldsymbol{v}_i + \sum_j \mu_j \boldsymbol{w}_j \tag{44}$$

$$\text{with } \lambda_i, \mu_j \geq 0, \quad \sum_i \lambda_i = 1. \tag{45}$$

Provided that the number of possible messages m is limited to a finite number, the vertices and conic generators of $\mathcal{M}^+$ can be determined using software such as PORTA or PANDA [27]. In Appendix C we prove that every pair $\big(p_\lambda(b|x,y), G_\lambda\big)$ can be constructed with a message of size $2^{n_X - 1}$ without loss of generality. In practice, however, the number of necessary messages may be considerably less than this in general: in cases with two or three inputs where we explicitly determined the vertices we never found that the number of necessary different messages exceeded the number of inputs n_X.

Once the vertices and conic generators $\big(p(m|x), G\big)$ of $\mathcal{M}^+$ have been obtained, one can generate all extreme points $\big(p(b|x,y), G\big)$ of $\mathcal{C}^+$ using (38) for each of the finite number of possible deterministic distributions $p(b|y,m)$ for Bob. We thus find that $\mathcal{C}^+$ is

[3] The set is not closed because the number G that we impose as an upper bound on the guessing probability is in principle unbounded. We could, of course, simply choose to impose a bound on it, such as $G \leq 1$. In that case, the set would become a (closed) polytope.

described by a finite number of vertices and conic generators, i.e., it is a polyhedron. Solving the facet enumeration problem, which again can be done in software provided that the problem is not too large, yields a finite number of inequalities that completely characterises the set of points $\big(p(b|x,y),G\big)$ compatible with classical stochastic communication.

3.1.2 Boundaries of $\mathcal{C}$ in the (3,2,2) scenario

We found earlier, in Section 2.2, that the classical stochastic and deterministic sets $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$ are always the same if Alice has two equiprobable inputs. The $(3,2,2)$ setting is therefore the smallest in which we could hope to find that $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$ are different even if Alice's inputs are chosen with the same probabilities ($q_x = 1/3$). This is indeed what we find for certain values of the upper bound G that we impose on the guessing probability.

We applied the method we described in the previous subsection to find the facets of $\mathcal{C}^+$ in the $(3,2,2)$ setting with $q_x = 1/3$. In terms of the correlators $E_{xy} = p(1|x,y) - p(2|x,y)$, in addition to the trivial conditions $\pm E_{xy} \leq 1$ and $G \geq 1/3$ its facets, up to relabellings of the inputs and outputs, are

$$-E_{11} - E_{12} - E_{21} + E_{22} + E_{31} \leq 6G - 1, \quad (46)$$
$$-E_{11} + E_{31} \leq 6G - 2. \quad (47)$$

For comparison, the facets of the deterministic version of the set, which we could call $\mathcal{C}^+_{\text{det}}$, are[4]

$$-E_{11} - E_{12} - E_{21} + E_{22} + E_{31} \leq 6G - 1, \quad (48)$$
$$-E_{11} - E_{12} - E_{21} + E_{22} + 2E_{31} \leq 12G - 4, \quad (49)$$
$$-E_{11} + E_{31} \leq 6G - 2. \quad (50)$$

We see here that $\mathcal{C}^+$ and $\mathcal{C}^+_{\text{det}}$ share two nontrivial classes of facets. Of these, (47) and (50), which we can rewrite as

$$\tfrac{1}{3}p(1|31) + \tfrac{1}{3}p(2|11) \leq G, \quad (51)$$

are instances of the constraints (20) that we pointed out apply regardless of the underlying physical theory in Section 2. They are not always facets of the sets $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$ with G fixed due to Alice having more inputs than Bob has outputs in this setting: in particular they become redundant if G is larger than $2/3$. The other boundary (46) and (48) common to the two sets, by contrast, is a nontrivial facet of both $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$ for all $1/3 \leq G \leq 1$.

The only difference between the stochastic and deterministic classical sets is the class of boundaries (49) unique to $\mathcal{C}_{\text{det}}$. Eq. (49) nontrivially constrains the correlations for any $G < 5/6$ but becomes redundant for $G \geq 5/6$. This tells us that $\mathcal{C}$ and $\mathcal{C}_{\text{det}}$ coincide if $G \geq 5/6$ but that $\mathcal{C}_{\text{det}}$ is a strictly smaller set than $\mathcal{C}$ for $G < 5/6$ in the $(3,2,2)$ setting with equiprobable priors.

[4] Ref. [21] originally inferred these boundaries by deriving the facets of $\mathcal{C}_{\text{det}}$ for multiple fixed values of G between $1/3$ and 1. We rederived them here following the analogue of the method of the previous subsection applied to deterministic communication, treating G as a free variable. This confirms conclusively that the boundaries hold for all values of G.

3.2 Membership testing by linear programming

While knowing the boundaries of $\mathcal{C}$ is useful for certain purposes, it is possible to solve the basic problem of testing for membership in $\mathcal{C}$ without explicitly needing to derive its boundaries. Given a bound G on the guessing probability, determining whether or not a given behaviour $p(b|x,y)$ is contained in the corresponding classical set $\mathcal{C}$ is equivalent to determining whether the pair $\big(p(b|x,y),G\big)$ is contained in the set $\mathcal{C}^+$ that we introduced in the previous subsection. This amounts to determining whether $p(b|x,y)$ and G can respectively be expressed as and bounded[5] by averages

$$p(b|x,y) = \sum_\lambda p(\lambda)p_\lambda(b|x,y), \quad (52)$$
$$G \geq \sum_\lambda p(\lambda)G_\lambda \quad (53)$$

of the respective components of vertices $\big(p_\lambda(b|x,y),G_\lambda\big)$ of $\mathcal{C}^+$.

Recalling how we generate the vertices of $\mathcal{C}^+$ from those of $\mathcal{M}^+$ in the previous subsection, we may substitute every vertex probability $p_\lambda(b|x,y)$ in (52) by

$$p_\lambda(b|x,y) = \sum_m p_\lambda(m|x)p_\lambda(b|y,m) \quad (54)$$

where $p_\lambda(m|x)$ is a vertex probability of $\mathcal{M}^+$ and $p_\lambda(b|y,m)$ is a deterministic response function. Furthermore, we may limit the number of messages to an alphabet of size $n_M = 2^{n_X-1}$ without loss of generality. This allows us to express the problem above as

$$p(b|x,y) = \sum_{\lambda,m} p(\lambda)p_\lambda(m|x)p_\lambda(b|y,m), \quad (55)$$
$$G \geq \sum_\lambda p(\lambda)G_\lambda, \quad (56)$$

with $p_\lambda(b|y,x) \in \{0,1\}$ and where $\big(p_\lambda(m|x),G_\lambda\big)$ is a vertex of $\mathcal{M}^+$, for all λ.

There are a finite number $n_K = n_B^{n_M \cdot n_Y}$ of possible deterministic response functions on Bob's side. Let us denote these $p_k(b|y,m)$, identified by an index k taking one of $n_B^{n_M \cdot n_Y}$ distinct values, and group the

[5] If we follow the exact formulation in the previous subsection then, as we point out in Appendix C, $\mathcal{C}^+$ has one conic generator $\big(p(b|x,y),G\big) = (0,1)$ in addition to its vertices which can be added to any point in $\mathcal{C}^+$ to increase its guessing probability bound component. Eliminating this conic generator results in (53) being an inequality.

remaining terms by k. Defining Λ_k as the set of λs appearing in the problem above for which

$$p_\lambda(b|y,m) = p_k(b|y,m), \tag{57}$$

we can rewrite our problem as

$$p(b|x,y) = \sum_{k,m} p(k) p_k(m|x) p_k(b|y,m), \tag{58}$$

$$G \geq \sum_k p(k) G_k \tag{59}$$

where

$$p(k) = \sum_{\lambda\in\Lambda_k} p(\lambda), \tag{60}$$

$$p_k(m|x) = \sum_{\lambda\in\Lambda_k} p(\lambda|k) p_\lambda(m|x), \tag{61}$$

$$G_k = \sum_{\lambda\in\Lambda_k} p(\lambda|k) G_\lambda, \tag{62}$$

and $p(\lambda|k)$ is defined in such a way that $p(k)p(\lambda|k) = p(\lambda)$.

The reexpression (58) and (59) of our problem is superficially the same as (55) and (56) except that now there is a known finite number of the indices k and m, while the pairs $\big(p_k(m|x), G_k\big)$ are no longer necessarily vertices of $\mathcal{M}^+$. The $\big(p_k(m|x), G_k\big)$s are still necessarily *contained* in $\mathcal{M}^+$, however, since $\mathcal{M}^+$ is convex, and thus by definition satisfy

$$G_k \geq \sum_m \max_x q_x p_k(m|x) \tag{63}$$

together with

$$p_k(m|x) \geq 0, \qquad \sum_k p_k(m|x) = 1. \tag{64}$$

Using these constraints in place of (61) and (62) and then eliminating the G_ks simplifies the problem to

$$p(b|x,y) = \sum_{k,m} p(k) p_k(m|x) p_k(b|y,m), \tag{65}$$

$$G \geq \sum_{k,m} p(k) \max_x q_x p_k(m|x), \tag{66}$$

where $p(k)$ and $p_k(m|x)$ are probability distributions.

To turn this into a linear programming problem we combine $p(k)$ and $p_k(m|x)$ into a joint distribution,

$$p(k,m|x) = p(k) p_k(m|x), \tag{67}$$

which satisfies the marginal condition that $\sum_m p(k,m|x) = p_k$ is independent of x for all k. With this last replacement the full problem becomes

$$p(b|x,y) = \sum_{k,m} p(k,m|x) p_k(b|y,m), \tag{68}$$

$$G \geq \sum_{k,m} \max_x q_x p(k,m|x), \tag{69}$$

$$p(k,m|x) \geq 0, \tag{70}$$

$$\sum_{k,m} p(k,m|x) = 1, \tag{71}$$

$$\sum_m p(k,m|x) = \sum_m p(k,m|x'), \qquad \forall x \neq x'. \tag{72}$$

Recalling that (69) is a shorthand for $n_M^{n_B}$ linear inequalities, determining whether there exist $n_K \cdot n_M \cdot n_X$ weights $p(k,m|x)$ that satisfy Eqs. (68)–(72) for a given behaviour $p(b|x,y)$ is a linear programming feasibility problem.

We remark, finally, that if we drop the marginal constraint (72) and combine (k,m) into a new variable which we rename m, we recover the definition of the classical set $\mathcal{C}$ that we started with in Section 2. This confirms that we did not inadvertently relax the problem when we replaced the vertices $\big(p_\lambda(m|x), G_\lambda\big)$ of $\mathcal{M}^+$ with the conditions (63) and (64) on $p_k(m|x)$. Deriving the linear programming feasibility problem following our characterisation of $\mathcal{C}^+$, however, allows us to put a finite upper limit $n_K \cdot n_M \cdot n_X$, with $n_K = n_B^{n_M \cdot n_Y}$ and $n_M = 2^{n_X - 1}$, on the number of weights $p(k,m|x)$ that we need to consider.

4 Characterising quantum correlations

In this section, we develop tools for the characterisation of informationally restricted quantum correlations. In Section 4.1, we develop an efficient method for optimising any given linear witness from inside the set of informationally restricted quantum correlations $\mathcal{Q}$. Hence, this method enables lower bounds on quantum correlations. In Section 4.2, we present a hierarchy of semidefinite relaxations of $\mathcal{Q}$ (and of $\mathcal{Q}_{\text{pure}}$). This allows us to establish increasingly precise necessary criteria of a given correlation admitting a quantum model. In Section 4.3, we apply these methods to the simplest relevant communication experiment and use it to device-independently quantify the information content of a quantum ensemble. In Section 4.4, we focus on the case of one bit of information and prove several strict resource inequalities involving two-dimensional systems, pure-state informationally restricted systems and general informationally restricted systems, in both the quantum and classical setting.

4.1 Lower bounds: alternating convex search method

In many situations arising in the study of quantum correlations, it is possible to use alternating convex

searches in order to optimise a linear functional of the quantum correlations (a linear "witness"), such as in the case of Bell inequalities [28, 29] or quantum dimension witnesses [30]. Such a search amounts to attempting to solve the full optimisation problem (over both states and measurements) by repeatedly optimising over the states and measurements separately in an alternating manner. The advantage of such an approach is that often each separate optimisation, over states (measurements) for fixed measurements (states), is convex and can be solved by standard methods. While alternating convex search often works well in practice, it is not guaranteed to converge and therefore only offers lower bounds on the optimal quantum correlations.

In order to optimise a linear witness over the set of informationally restricted quantum correlations, one encounters a less straightforward situation. For a fixed set of states, it is clear that the optimisation over the set of measurements can be evaluated as a semidefinite program (SDP). In contrast, for a fixed set of measurements, the optimisation over the set of states is less obvious due to the relevance of the informational restriction. Evidently, the optimisation must be performed under the constraint $P_{\mathrm{g}} \leq G$ which itself involves a maximisation over the extraction POVM $\{N_x\}_x$. We show how this difficulty can be overcome so that lower bounds on informationally restricted quantum correlations can be efficiently computed through alternating implementations of SDPs.

Consider that we are given a linear witness $\mathcal{A}$, in general written as

$$\mathcal{A} = \sum_{x,y,b} c_{xyb} p(b|x,y) = \sum_{x,y,b} c_{xyb} \operatorname{Tr}\left[\rho_x M_{b|y}\right] \quad (73)$$

for some real coefficients c_{xyb}, and asked to maximise it over the set of informationally restricted states ρ_x and measurements $M_{b|y}$. For this purpose, let us define an auxiliary positive semidefinite operator σ with the property that

$$\forall x: \quad \sigma \geq q_x \rho_x. \quad (74)$$

This allows us to place the following upper bound on the guessing probability:

$$\begin{aligned} P_{\mathrm{g}}(X|B) &= \max_{\{N_x\}_x} \sum_x q_x \operatorname{Tr}\left[\rho_x N_x\right] \\ &\leq \max_{\{N_x\}_x} \sum_x \operatorname{Tr}\left[\sigma N_x\right] = \operatorname{Tr}[\sigma], \end{aligned} \quad (75)$$

where we have used that $\sum_x N_x = \mathbb{1}$. The introduction of σ stems from considering the semidefinite dual of the guessing probability and does therefore not constitute a relaxation of the problem [31]. Its advantage is that it allows us to treat the informational restriction as a tracial constraint enforced through the additional semidefinite constraints in (74). We may therefore cast the maximisation of the linear witness $\mathcal{A}$, for a given bound G on the guessing probability, as the following optimisation problem:

$$\begin{aligned} \max_{\rho_x,\sigma,M_{b|y}} \quad & \sum_{x,y,b} c_{xyb} \operatorname{Tr}\left[\rho_x M_{b|y}\right] \\ \text{such that} \quad & \rho_x \geq 0, \qquad \operatorname{Tr}[\rho_x] = 1, \\ & \sigma \geq q_x \rho_x, \qquad \operatorname{Tr}[\sigma] \leq G, \\ & M_{b|y} \geq 0, \qquad \sum_b M_{b|y} = \mathbb{1}. \end{aligned} \quad (76)$$

If we fix the measurement operators $\{M_{b|y}\}$, this problem becomes an SDP for the states $\{\rho_x\}$ and σ. Conversely, if we fix the states $\{\rho_x\}$ and σ, it is an SDP for the measurement operators $\{M_{b|y}\}$. We can thus alternate these SDPs to obtain a lower bound on the optimal value. Note that it is implicit that these SDPs must be performed in a given Hilbert space dimension, but one may find successively better lower bounds by increasing the dimension. The usefulness of this method is exemplified in Section 4.3.

Note that the above approach cannot be applied to the pure-state set $\mathcal{Q}_{\mathrm{pure}}$, since the condition $\rho_x \geq 0$ would have to be replaced by $\rho_x^2 = \rho_x$, which is nonlinear. The existence of a practical algorithm lower-bounding the general quantum set $\mathcal{Q}$ is another advantage of our general formulation.

4.2 Upper bounds: hierarchy of semidefinite relaxations

The idea used in the previous section, of introducing the auxiliary operator σ, can be further leveraged to systematically obtain increasingly precise upper bounds on the informationally restricted set of quantum correlations. We now present a hierarchy of semidefinite relaxations for the set $\mathcal{Q}$, which is based on the tracial variant [32, 33] of the NPA non-commutative polynomial optimisation hierarchy [34, 35].

Let us first slightly rewrite the problem (76) as

$$\begin{aligned} \max_{\rho_x,\sigma,M_{b|y}} \quad & \sum_{x,y,b} c_{xyb} \operatorname{Tr}\left[\rho_x M_{b|y}\right] \\ \text{such that} \quad & \rho_x - \rho_x^2 \geq 0, \quad \operatorname{Tr}[\rho_x] = 1, \\ & \sigma - q_x \rho_x \geq 0, \quad G\mathbb{1} - \sigma \geq 0, \quad \operatorname{Tr}[\sigma] \leq G, \\ & M_{b|y} M_{b'|y} = \delta_{bb'} M_{b|y}, \quad \sum_b M_{b|y} = \mathbb{1}, \end{aligned} \quad (77)$$

where compared with (76), we have replaced the constraint $\rho_x \geq 0$ by $\rho_x - \rho_x^2 \geq 0$, added the redundant constraint $G\mathbb{1} - \sigma \geq 0$, and assumed, without loss of generality if we do not bound the dimension d of the Hilbert space, that the measurements $\{M_{b|y}\}_b$ are projective. The optimization problems (76) and (77) are entirely equivalent, but the second formulation is better suited for the tracial non-commutative opti-

mization method of [32][6], which we now explain how to apply.

Let w denote a monomial, i.e. a product, of the $n_X + 1 + n_B n_Y$ basic operators $\rho_1, \rho_2, \ldots, \rho_{n_X}$, σ, and $M_{1,1}, \ldots, M_{n_B|1}, M_{1|2}, \ldots, M_{n_B|n_Y}$. We refer to the number k of such basic operators in the product w as the degree k of this monomial. By convention, the identity operator $\mathbb{1}$ is the monomial of degree 0. Let $\mathcal{W}_k$ denote the set of all monomials of degree at most k and let $n(k)$ denote the number of such monomials. Linear combinations $p = \sum_{w\in\mathcal{W}_k} \alpha_w w$ of the monomials then correspond to polynomials of degree k in the basic operators.

Let L be a linear functional that assigns to each monomial w in $\mathcal{W}_{2k}$ of degree $2k$ the real number $L(w)$, and thus which assigns to each polynomial $p = \sum_{W\in\mathcal{W}_{2k}} \alpha_w w$ of degree $2k$ the real number $L(p) = \sum_{w\in\mathcal{W}_k} \alpha_w L(w)$. Given such a functional L, we define

- the *moment* matrix $\Gamma_k(L)$, as the matrix of size $n(k)$ whose entries are indexed by monomials $u, v \in \mathcal{W}_k$ and are equal to $[\Gamma_k(L)]_{u,v} = L(u^\dagger v)$;
- the *localizing* matrix $\Gamma_k(L; p)$ associated to a polynomial p of degree two or less, as the matrix of size $n(k) - 1$ whose entries are indexed by monomials $u, v \in \mathcal{W}_{k-1}$ and are equal to $[\Gamma_k(L; p)]_{u,v} = L(u^\dagger p\, v)$.

Consider now the following problem for $k \geq 1$,

$$\begin{aligned}
\max_{L} \quad & \sum_{x,y,b} c_{xyb}\, L\left(\rho_x M_{b|y}\right) \\
\text{such that } & \Gamma_k(L) \geq 0, \\
& \Gamma_k(L; \rho_x - \rho_x^2) \geq 0, \quad L(\rho_x) = 1, \\
& \Gamma_k(L; \sigma - q_x\rho_x) \geq 0, \quad \Gamma_k(L; G\mathbb{1} - \sigma) \geq 0, \\
& L(\sigma) \leq G, \\
& L(p) = L(p'), \text{ if } \mathrm{Tr}[p] = \mathrm{Tr}[p'] \\
& \qquad \text{for any polynomials } p, p' \text{ of degree } 2k,
\end{aligned} \tag{78}$$

where in the last condition the identity $\mathrm{Tr}[p] = \mathrm{Tr}[p']$ is evaluated by taking into account the polynomial identities $M_{b|y} M_{b'|y} = \delta_{bb'} M_{b|y}$ and $\sum_b M_{b|y} = \mathbb{1}$ satisfied by the measurement operators. This optimization problem is an SDP (since it amounts to optimize $n(2k)$ variables, the values $L(w)$ of the monomials w of degree less than $2k$, subject to linear constraints and to the positivity of matrices whose entries are linearly related to these variables).

[6]Specifically, the constraints $\rho_x - \rho_x^2$ imply not only that $\rho_x \geq 0$ but also that $\rho_x \leq \mathbb{1}$. Together with the constraint $\sigma \leq G\mathbb{1}$ this guarantees that the feasible set of (77) satisfies the archimedean assumption and that the entries of the moment matrices stay bounded. Assuming that the measurements are projectives instead of general POVMs dispenses us from introducing localizing matrices associated with them.

Clearly any solution of (77) defines a solution of (78) through $L(w) = \mathrm{Tr}[w]$ [7]. Thus the problem (78) represents an SDP relaxation of (77) approximating the set $\mathcal{Q}$ from the outside. By increasing the relaxation level k, one obtains a hierarchy of increasingly constraining conditions on $\mathcal{Q}$.

Note that the above method can also be used to characterise the set of pure-state quantum correlations $\mathcal{Q}_{\text{pure}}$ by replacing in (76) the positivity constraints $\rho_x - \rho_x^2 \geq 0$ by the polynomial constraints $\rho_x = \rho_x^2$, resulting in the simpler relaxation

$$\begin{aligned}
\max_{L} \quad & \sum_{x,y,b} c_{xyb}\, L\left(\rho_x M_{b|y}\right) \\
\text{such that } & \Gamma_k(L) \geq 0, \\
& \Gamma_L(\rho_x) = 1, \\
& \Gamma_k(L; \sigma - q_x\rho_x) \geq 0, \quad \Gamma_k(L; G\mathbb{1} - \sigma) \geq 0, \\
& L(\sigma) \leq G, \\
& L(p) = L(p'), \text{ if } \mathrm{Tr}[p] = \mathrm{Tr}[p'] \\
& \qquad \text{for any polynomials } p, p' \text{ of degree } 2k,
\end{aligned} \tag{79}$$

where the last condition is evaluated using, in addition to the polynomial constraints on the measurement operators, the conditions $\rho_x = \rho_x^2$.

We remark that by additionally imposing that all operators commute, we can also bound classical correlations via the above SDPs. This can be useful in scenarios that are too large to be efficiently treated with the methods developed in Section 3.

Finally, let us stress that the series of SDP relaxations that we introduced are relaxations. Convergence to the exact quantum set is not guaranteed in the limit $k \to \infty$, see [32, 33, 36] for more details about the general properties of the SDP hiearchy for non-commutative tracial optimization.

4.3 Device-independent witnessing of the information content of quantum communication

Consider a quantum communication experiment in which we do not know the amount of information communicated from Alice to Bob. Is it possible to determine a lower bound on the amount of information that Alice must send to Bob given only the observed correlations $p(b|x, y)$? This amounts to the task of device-independently testing the information content of quantum communication. Using the tools of the previous sections, we exemplify such device-independent certification in the simplest relevant communication experiment.

As we have seen in Section 2.1, there can be no quantum advantage when the scenario only features two states. Moreover, no advantage is possible when

[7]This can be seen by following the same lines as in [34, 35], where the linear functional was instead defined as $L(w) = \langle\Psi|w|\Psi\rangle$ for some reference state $|\Psi\rangle$.

Bob only has a single input, because his measurement could then be performed already in Alice's lab and the outcomes simply relayed to Bob as classical communication (since performing a measurement cannot increase the guessing probability). Therefore, the simplest relevant scenario in which we expect a quantum advantages is that in which Alice has three states ($n_X = 3$) and Bob has two binary-outcome measurements ($n_Y = n_B = 2$). In this scenario, we focus on the linear witness (46) (here labelled $\mathcal{A}_{322}$) corresponding to a facet of the classical polytope under uniform priors ($q_x = 1/3$).

Firstly, we apply the SDP hierarchy for the set $\mathcal{Q}_{\text{pure}}$ to find upper bounds on $\mathcal{A}_{322}$ as a function of the guessing probability (information). We implemented the SDP relaxation (79) with $k = 3$ but to simplify the numerical optimisation considered a subset of all SDP and linear constraints. Specifically, we only imposed the positivity of the submatrix of $\Gamma_3(L)$ whose rows and columns are indexed by the monomials

$$\{\mathbb{1},\, \sigma,\, \rho,\, M,\, \rho M,\, \rho\rho,\, MM,\, \rho\sigma,\, M\sigma,\, \rho\rho\rho,\\ MM\sigma,\, \rho MM,\, \rho M\sigma,\, \rho M\rho\}, \quad (80)$$

the positivity of the submatrices $\Sigma_2^x(L)$ whose rows and columns are indexed by the monomials

$$\{\mathbb{1},\, \rho,\, M,\, \rho\rho,\, MM,\, \rho M\}, \quad (81)$$

and the linear constraints $L(P) = L(P')$ involving the entries of such matrices. This corresponds to a 98×98 moment submatrix Γ and three 25×25 localising submatrices Σ^x. Evaluating the corresponding SDPs for different informational restrictions, we obtain the red curve illustrated in Fig. 3. Notably, this upper bound is in fact tight, since it coincides with the explicit pure-state quantum strategy reported in Ref. [21] (thus proving its optimality).

Similarly, we have also implemented the SDP hierarchy for the general quantum set $\mathcal{Q}$ using submatrices of the localising matrices $P_3^x(L)$ based on the same monomial list (81) as for $\Sigma_3^x(L)$. The obtained bounds on the witness are given by the blue curve in Fig. 3. We observe that for every guessing probability $P_{\text{g}} \in (\frac{1}{3}, 1) \setminus \{\frac{2}{3}\}$ we find a larger bound in the general setting as compared to the pure-state setting. In order to show that this gap is not an artefact of the bounds in the general setting not being tight, we have employed the alternating convex search described in Section 4.1 to construct explicit quantum models. The obtained values of the witness are illustrated by the black curve in Fig. 3. We find that for $P_{\text{g}} \in [\frac{1}{3}, \frac{2}{3}]$, the upper and lower bounds in the general setting accurately coincide. In the interval $P_{\text{g}} \in (\frac{2}{3}, 1)$ a small gap between the upper and lower bound remains. Nevertheless, our lower bounds exceed the upper bounds for the pure-state setting, thus proving that informationally restricted quantum correlations outperform their pure-state counterparts. It is interesting to note that for the special case of $P_{\text{g}} = \frac{2}{3}$, which corresponds precisely to $\mathcal{I}(X|B) = 1$ bit of information, there is no discrepancy between $\mathcal{Q}$ and $\mathcal{Q}_{\text{pure}}$.

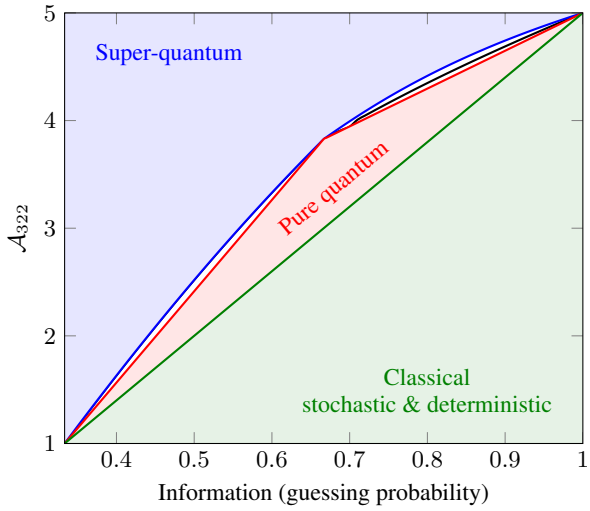

Figure 3: The witness $\mathcal{A}_{322}$ versus the information (in terms of the guessing probability). The plot displays an upper bound (blue) and lower bound (black) on general quantum models, a tight upper bound on pure-state quantum models (red) and a tight upper bound on classical models (green). As the first two curves coincide in the interval $P_{\text{g}} \in [\frac{1}{3}, \frac{2}{3}]$, this part of the quantum boundary is fully characterised. However, in the interval $P_{\text{g}} \in (\frac{2}{3}, 1]$, the quantum boundary is not fully characterised but delimited by the blue and black curves.

We can interpret these results in the context of device-independent tests of information. If the information content of the quantum communication is not known, then we may use the upper bound on the quantum correlations (blue curve) to determine a bound on the minimal amount of information required to explain the observed correlations in a quantum model. For example, Ref. [37] experimentally implemented this communication experiment using both qubit and qutrit ensembles and reported a witness value of $\mathcal{A}_{322}^{\text{qubit}} = 3.7815 \pm 0.0782$ and $\mathcal{A}_{322}^{\text{qutrit}} = 4.9303 \pm 0.1032$ respectively. In order to determine the information content of these ensembles (without assuming their respective dimensions), we use our upper bounds on the quantum correlations. Specifically, when the experimental errors are taken into consideration, we certify a quantum information content of at least $\mathcal{I}(X|B) = 0.98 \pm 0.02$ bits for the first ensemble and $\mathcal{I}(X|B) = 1.54 \pm 0.05$ bits for the second ensemble. Both these results nearly saturate the maximal possible information content of qubit and qutrit ensembles, namely 1 bit and $\log_2 3$ bits respectively.

4.4 Resource inequalities for one bit of information

Consider the information restriction $\mathcal{I}(X|B) \leq \alpha$ with $\alpha = \log_2 d$ for some integer $d \geq 2$. This is a particularly interesting case since it enables a meaningful comparison of classical and quantum correlations to those that can be obtained from d-dimensional classical and quantum communication. Here, we focus on the simplest case of $d = 2$ ($\mathcal{I}_X \leq 1$ bit) and consider the comparative relation between classical and quantum correlations respectively when obtained from i) communication of two-level systems, ii) one bit of communication in pure-state models and iii) one bit of communication in general models. Let us denote the set of classical and quantum correlations achievable with two-dimensional communication by $\mathcal{C}_{\text{dim}}$ and $\mathcal{Q}_{\text{dim}}$. It is clear that the following two chains of inclusions must be true:

$$\mathcal{C}_{\text{dim}} \subseteq \mathcal{C}_{\text{det}} \subseteq \mathcal{C} \quad \text{and} \quad \mathcal{Q}_{\text{dim}} \subseteq \mathcal{Q}_{\text{pure}} \subseteq \mathcal{Q}. \tag{82}$$

The first inclusion in each case follows from the fact that every ensemble of classical or quantum two-level systems can be simulated by classical or quantum ensembles of pure two-level systems under shared randomness[8]. The second inclusion on each line follows trivially from the fact that general classical and quantum models admit deterministic and pure-state models respectively as special cases.

It is interesting to determine which of the inclusions (82) are strict, i.e., which classical and quantum resources are fundamentally different. We first focus on the quantum case and prove that all three resources are inequivalent. Notably, Ref. [21] proved that $\mathcal{Q}_{\text{dim}} \subset \mathcal{Q}$ using a construction that involved 16 states. The proofs presented here are simpler, as they only require three states, but inherently different as they are based on biasing the prior probabilities.

Consider again the input/output scenario $(n_X, n_Y, n_B) = (3, 2, 2)$ and once again the witness $\mathcal{A}_{322}$. In the previous section, we saw that for $\mathcal{I}_X \leq 1$ bit ($P_{\text{g}} \geq \frac{2}{3}$), there was no discrepancy between the general quantum model and the pure-state quantum model. In addition, if we restrict to qubits, the witness $\mathcal{A}_{322}$ reduces to that introduced in Ref. [5], whose maximum is known again to give the same result. However, consider now that we change the prior distribution of Alice's inputs: instead of being uniform, let us choose it as $q_1 = q_2 = \frac{2}{5}$ and $q_3 = \frac{1}{5}$. Since $H_{\min}(X) = \log_2(5) - 1$, the guessing probability corresponding to one bit of information is $P_{\text{g}} = \frac{4}{5}$. What now are the largest possible values of $\mathcal{A}_{322}$ under qubits, pure-state models with $P_{\text{g}} \leq \frac{4}{5}$, and general models with $P_{\text{g}} \leq \frac{4}{5}$?

[8] Recall that since every ensemble of two-level systems carries no more than one bit of information, then also their mixture under shared randomness does not lead to more than one bit of information.

Since biasing the prior affects the information constraint but not the dimension of the physical system, it follows that the largest value of $\mathcal{A}_{322}$ remains unaffected when evaluated over qubits. We have

$$\mathcal{A}_{322} \overset{\mathcal{Q}_{\text{dim}}}{\leq} 1 + 2\sqrt{2} \approx 3.8284, \tag{83}$$

which is a tight bound. However, in the case of pure-state models and general quantum models, biasing the prior means that Bob already has some knowledge of Alice's input. Thus, we would intuitively expect that the correlations improve as compared to the unbiased case. This intuition can be proven using the tools from the previous sections. Evaluating the respective semidefinite relaxations of the set of quantum correlations, we find that

$$\mathcal{A}_{322} \overset{\mathcal{Q}_{\text{pure}}}{\leq} 4.3184, \qquad \mathcal{A}_{322} \overset{\mathcal{Q}}{\leq} 4.4641. \tag{84}$$

We use alternating convex search to place a lower bound on the witness in the stochastic case: for qubits we achieve $\mathcal{A}_{322} = 3.8284$ (saturating (83)), for qutrits we achieve $\mathcal{A}_{322} = 4.2641$ and for ququarts we achieve $\mathcal{A}_{322} = 4.4142$. The ququart strategy uses one pure state and two mixed states each with spectra $(1/2, 1/2, 0, 0)$. The lower bound obtained with ququarts is sufficient to outperform pure-state quantum models and conclude that $\mathcal{Q}_{\text{pure}} \subset \mathcal{Q}$. Moreover, in order to also show that $\mathcal{Q}_{\text{dim}} \subset \mathcal{Q}_{\text{pure}}$, it is sufficient to note that the following strategy based on pure-state quantum communication outperforms the qubit bound. Let Alice prepare the qutrit states

$$|\psi_1\rangle = \frac{1}{2}\begin{pmatrix}\sqrt{3}\\1\\0\end{pmatrix}, |\psi_2\rangle = \frac{1}{2}\begin{pmatrix}1\\\sqrt{3}\\0\end{pmatrix}, |\psi_3\rangle = \begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{85}$$

It is easily checked (e.g. via an SDP) that the guessing probability is $P_{\text{g}} = 4/5$. Then, let Bob perform compatible measurements $\{|3\rangle, |1+2\rangle\}$ and $\{|2\rangle, |1+3\rangle\}$. Then, one finds $\mathcal{A}_{322} = 4$ which exceeds the qubit bound.

Let us now consider the same problem with classical resources. Using the tools from Section 3, we can straightforwardly show the tight inequalities

$$\mathcal{A} \overset{\mathcal{C}_{\text{dim}}}{\leq} 3, \qquad \mathcal{A} \overset{\mathcal{C}_{\text{det}}}{\leq} 4, \qquad \mathcal{A} \overset{\mathcal{C}}{\leq} 4, \tag{86}$$

which immediately assert that informationally restricted classical correlations are more powerful than dimensionally restricted classical correlations; specifically $\mathcal{C}_{\text{dim}} \subset \mathcal{C}_{\text{det}}$. However, it still does not determine whether $\mathcal{C}_{\text{det}}$ is a strict subset of $\mathcal{C}$ for one bit of information. This is left as an open problem.

5 Semi-device-independent random number generation

In the previous section, we have seen how quantum correlations can be bounded in communication experiments in which the only assumption is a bound on

the amount of information that the communication carries. Here, we leverage these methods towards application in semi-device-independent RNG. In a first example, we focus on the facet-defining witness $\mathcal{A}_{322}$ and compute the certified randomness as a function of the information. This allows us to obtain a nearly optimal RNG rate. In a second example, we consider the case of one bit of information and consider the amount of randomness that can be robustly certified under a conventional qubit assumption as compared to that certified under an information assumption. We show that the correlations used in a standard qubit experiment can be recycled to certify the same amount of randomness when the assumption is relaxed to the strictly weaker information assumption.

5.1 Randomness versus information

Let us again consider the witness $\mathcal{A}_{322}$ in (a general) quantum model. In Section 4.3 we obtained the maximal quantum witness value for any information between zero and one bit, corresponding to a guessing probability $P_g \in [\frac{1}{3}, \frac{2}{3}]$. Here, we evaluate the extractable randomness in the output of Bob associated to such maximal quantum witness values. Specifically, we consider that Alice and Bob decide to extract randomness from the event corresponding to Alice's third input ($x = 3$) and Bob's first input ($y = 1$). Then, the certified randomness is given by the min-entropy $H_{\min} = -\log_2 p_*$, where $p_* = \max\{p(1|3,1), p(2|3,1)\}$, compatible with the observed maximal value of $\mathcal{A}_{322}$[9]. Using the introduced semidefinite relaxations, we can place an upper bound on p_* which translates into a lower bound on the certified randomness. The results are illustrated in Fig. 4. These results can also be accurately matched with upper bounds on the randomness obtained via the alternating convex search method (see Section 4.1). Hence, the bound on the certified randomness is tight (up to solver precision). In Fig. 4, we see that by suitably tuning the information in Alice's communication, one can obtain nearly one bit of randomness (which is algebraically maximal for binary-outcome measurements). Specifically, at $P_g \approx 0.522$ we certify approximately 0.995 bits of randomness. Hence, we conclude that nearly optimal randomness can be certified under the information assumption. Notably, for $P_g \approx \frac{2}{3}$, the randomness vanishes. This is due to our choice of setting ($x = 3$, $y = 1$). A substantial amount of randomness can be certified for $P_g \approx \frac{2}{3}$ by instead considering the event $(x, y) = (1, 1)$. However, the rate is significantly lower than that obtained at the optimal choice of information for $(x, y) = (3, 1)$.

[9] To enhance numerical feasibility, we only impose the optimal value of $\mathcal{A}_{322}$ up to four decimals.

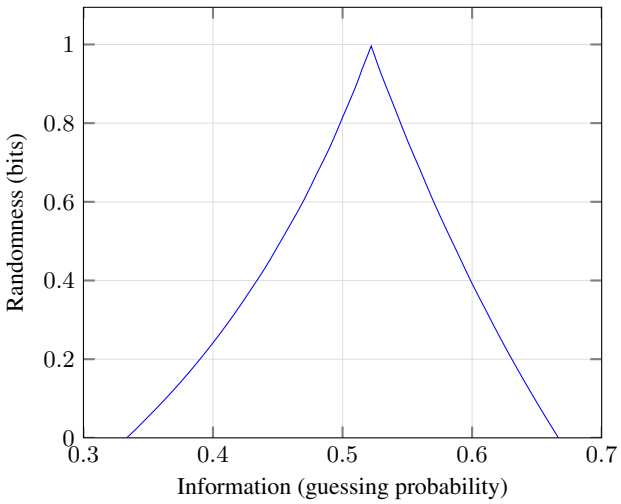

Figure 4: Randomness versus the information (quantified via the guessing probability) of Alice's communication. The results are obtained for the maximal quantum value of the witness $\mathcal{A}_{322}$ in general quantum communication models.

5.2 Qubits versus one bit of information

We investigate the comparison between certified randomness under the conventional assumption of qubits and our assumption of informational restriction. This comparison is only meaningful for one bit of information; to which we therefore restrict ourselves. To this end, we focus on a witness that has previously been employed for RNG in dimension bounded systems [38, 39], namely a quantum random access code.

In a quantum random access code, Alice receives one of four possible inputs labelled by two bits $x = x_1x_2 \in \{1, 2\}^{\times 2}$ while Bob has two possible inputs $y \in \{1, 2\}$ and two possible outputs $b \in \{1, 2\}$. The correlation witness is defined as

$$\mathcal{A}_{\text{RAC}} = \frac{1}{8} \sum_{x,y} (-1)^{x_y} E_{xy}. \tag{87}$$

We analyse this witness in two scenarios, i) Alice sends qubits to Bob (dimension assumption) and ii) Alice sends at most one bit of information to Bob (information assumption). Naturally, since all qubit ensembles carry at most one bit of information, while many higher dimensional ensembles also carry no more than one bit of information, the information assumption is less restrictive than the dimension assumption. It is well known that the optimal value of $\mathcal{A}_{\text{RAC}}$ using qubits is $\frac{1}{\sqrt{2}}$ [16]. Using the tools from Section 4, we find that $\mathcal{A}_{\text{RAC}} = \frac{1}{\sqrt{2}}$ also is the largest possible value under one bit of information.

Due to the symmetries of the witness $\mathcal{A}_{\text{RAC}}$, the choice of event from which randomness is extracted does not influence the amount of randomness certified. We therefore choose the event $(x, y) = (1, 1)$ and employ semidefinite relaxations for informationally restricted quantum correlations to place a lower

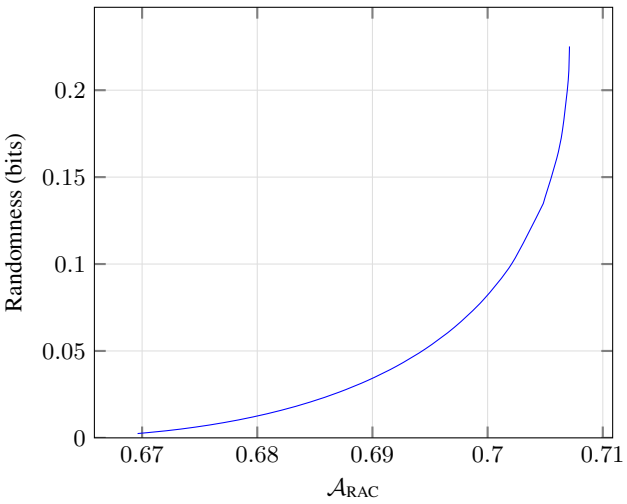

Figure 5: The randomness certified in a quantum random access code. Up to numerical precision, the amount of randomness certified under the qubit assumption and one bit of information assumption is identical.

bound on the randomness as a function of the witness. The results are illustrated in Fig. 5. A nearly optimal value of the witness certifies over 0.2 bits of randomness while also significantly sub-optimal witness values permit a non-zero amount of certified randomness. Then, we consider the same problem under the assumption of qubit communication. To this end, we have used the symmetrised semidefinite relaxation hierarchy of Refs. [8, 40]. Up to solver precision, we certify the same amount of randomness as is obtained under the information assumption, i.e. the curve is identical to that displayed in Fig. 5. Moreover, the obtained lower bounds on the randomness are optimal since we can saturate them with an explicit family of quantum models based on qubits. Hence, we conclude that the quantum random access code allows us to certify the same amount of randomness under the strictly weaker assumption of informational restriction as compared to the dimension bounded scenario, while only requiring the experimental realisation of standard qubit strategies.

6 Conclusion

In this article, we have investigated classical and quantum correlations limited only by the information content of the corresponding classical and quantum communication. This constitutes a departure from conventional dimension bounded communication in favour of an analysis based on entropic quantities. We have presented a complete characterisation of informationally restricted classical correlations in terms of linear programming, thereby generalising the results of [21] based on deterministic communication models. For the set of informationally restricted quantum correlations, we have both developed efficient interior-point search methods and hierarchies of semidefinite relaxations for placing upper bounds on the set. We have applied these tools to device-independently witness the amount of information carried by a classical and quantum ensemble as well as to establish strict resource inequalities for different information resources. Furthermore, we have outlined a new avenue for semi-device-independent quantum information processing based on the information assumption. This was exemplified through the investigation of semi-device-independent random number generation for which we both reported nearly optimal rates and advantages over dimension bounded systems. The results presented in this work provide important tools for analysing informationally restricted classical and quantum correlations.

Our work leaves a number of open problems, some of which we list here. 1) How tight are the bounds obtained through our semidefinite hierarchy for informationally restricted quantum correlations? Can one introduce a semidefinite hierarchy that provably converges to the quantum set? 2) Is there a strict resource inequality for informationally restricted classical correlations for the deterministic versus general communication models? 3) It would be interesting to consider the experimental implementation of semi-device-independent random number generation based on the information assumption. 4) Are there other semi-device-independent protocols that are practical to base on the information assumption? Two obvious candidates to consider for this purpose are quantum key distribution and self-testing.

Finally, we note that the information-restricted approach also can be used as a relaxation method to bound correlations in prepare-and-measure experiments subject to other assumptions, for which methods to bound the set of quantum correlations are not known.

Acknowledgments

We thank Nicolas Brunner and Marie Ioannou for discussions. This work was supported by the Swiss National Science Foundation via the NCCR-SwissMap and the EU Quantum Flagship project QRANGE. E.Z.C. and A.T. acknowledge support by the Swiss National Science Foundation via the Early PostDoc Mobility fellowships P2GEP2 188276 and P2GEP2 194800. S.P. is a Senior Research Associate of the Fonds de la Recherche Scientifique – FNRS.

References

[1] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, in *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Comput-*

ing, STOC '99 (Association for Computing Machinery, New York, NY, USA, 1999) pp. 376–383.
[2] E. F. Galvão, Phys. Rev. A **65**, 012318 (2001).
[3] P. Trojek, C. Schmid, M. Bourennane, C. Brukner, M. Zukowski, and H. Weinfurter, Phys. Rev. A **72**, 050305 (2005).
[4] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols, Quantum random access codes with shared randomness (2008), arXiv:0810.2937 [quant-ph] .
[5] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Phys. Rev. Lett. **105**, 230501 (2010).
[6] N. Brunner, M. Navascués, and T. Vértesi, Phys. Rev. Lett. **110**, 150501 (2013).
[7] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Phys. Rev. Lett. **114**, 170502 (2015).
[8] M. Navascués and T. Vértesi, Phys. Rev. Lett. **115**, 020501 (2015).
[9] M. Smania, A. M. Elhassan, A. Tavakoli, and M. Bourennane, npj Quantum Inf. **2**, 16010 (2016).
[10] A. Tavakoli and M. Żukowski, Phys. Rev. A **95**, 042305 (2017).
[11] D. Martínez, A. Tavakoli, M. Casanova, G. Cañas, B. Marques, and G. Lima, Phys. Rev. Lett. **121**, 150504 (2018).
[12] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Méthot, and V. Scarani, Phys. Rev. Lett. **100**, 210503 (2008).
[13] M. Pawłowski and N. Brunner, Phys. Rev. A **84**, 010302 (2011).
[14] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **84**, 034301 (2011).
[15] E. Woodhead and S. Pironio, Phys. Rev. Lett. **115**, 150501 (2015).
[16] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, Phys. Rev. A **98**, 062307 (2018).
[17] T. Van Himbeeck, E. Woodhead, N. J. Cerf, R. García-Patrón, and S. Pironio, Quantum **1**, 33 (2017).
[18] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner, Phys. Rev. Applied **7**, 054018 (2017).
[19] R. Chaves, J. B. Brask, and N. Brunner, Phys. Rev. Lett. **115**, 110501 (2015).
[20] A. Tavakoli, Phys. Rev. Lett. **126**, 210503 (2021).
[21] A. Tavakoli, E. Zambrini Cruzeiro, J. B. Brask, N. Gisin, and N. Brunner, Quantum **4**, 332 (2020).
[22] R. Konig, R. Renner, and C. Schaffner, IEEE Trans. Inf. Th. **55**, 4337 (2009).
[23] N. Ciganović, N. J. Beaudry, and R. Renner, IEEE Trans. Inf. Th. **60**, 1573 (2013).
[24] A. Tavakoli, J. Pauwels, E. Woodhead, and S. Pironio, Correlations in entanglement-assisted prepare-and-measure scenarios (2021), arXiv:2103.10748v2, 2103.10748 .
[25] L. Wang and R. Renner, Phys. Rev. Lett. **108**, 200501 (2012).
[26] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[27] S. Lörwald and G. Reinelt, EURO J. Comput. Optim. **3**, 297 (2015).
[28] R. F. Werner and M. M. Wolf, Quantum Inf. Comput. **1**, 1 (2001).
[29] K. F. Pál and T. Vértesi, Phys. Rev. A **82**, 022116 (2010).
[30] A. Tavakoli, M. Pawłowski, M. Żukowski, and M. Bourennane, Phys. Rev. A **95**, 020302 (2017).
[31] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, 2018).
[32] S. Burgdorf, K. Cafuta, I. Klep, and J. Povh, Mathematical Programming **137**, 557 (2013).
[33] I. Klep and J. Povh, Journal of Global Optimization **64**, 325 (2016).
[34] M. Navascués, S. Pironio, and A. Acín, New J. Phys. **10**, 073013 (2008), publisher: IOP Publishing.
[35] S. Pironio, M. Navascués, and A. Acín, SIAM J. Opt. **20**, 2157 (2010), publisher: Society for Industrial and Applied Mathematics.
[36] S. Gribling, D. de Laat, and M. Laurent, Foundations of Computational Mathematics **19**, 1013 (2019).
[37] J. Ahrens, P. Badziag, A. Cabello, and M. Bourennane, Nat. Phys. **8**, 592 (2012).
[38] H.-W. Li, Z.-Q. Yin, Y.-C. Wu, X.-B. Zou, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **84**, 034301 (2011).
[39] H.-W. Li, M. Pawłowski, Z.-Q. Yin, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **85**, 052308 (2012).
[40] A. Tavakoli, D. Rosset, and M.-O. Renou, Phys. Rev. Lett. **122**, 070501 (2019).

A Derivation of vertex probabilities

Here we derive the vertex probabilities (22)–(25) in Section 2.1 using the Fourier-Motzkin algorithm. For convenience, we undertake the derivation for joint probabilities

$$p_{bx} = q_x p(b|x), \qquad x = 1, 2, \tag{88}$$

in which we absorb the prior probabilities q_x with which the inputs $x = 1$ and $x = 2$ are chosen. We also drop Bob's input y, since the constraints on these probabilities are just the same repeated for each y.

These probabilities are characterised by

$$\sum_b p_{bx} = q_x, \tag{89}$$
$$p_{bx} \geq 0, \tag{90}$$
$$\sum_b \max(p_{b1}, p_{b2}) \leq G. \tag{91}$$

Setting $p_{b1} = (v_b + \delta_b)/2$ and $p_{b2} = (v_b - \delta_b)/2$ we can reexpress the same problem as

$$1 - \sum_b v_b = 0, \tag{92}$$
$$\delta - \sum_b \delta_b = 0, \tag{93}$$
$$v_b + \delta_b \geq 0, \tag{94}$$
$$v_b - \delta_b \geq 0, \tag{95}$$
$$\Delta - \sum_b |\delta_b| \geq 0, \tag{96}$$

with $\Delta = 2G - 1$ and $\delta = q_1 - q_2$. The last inequality should be read as 2^{n_B} different linear inequalities, corresponding to the 2^{n_B} different combinations of substitutions $|\delta_b| = \pm\delta_b$.

The most general valid inequality can be obtained by taking linear combinations of the above constraints with nonnegative coefficients for the inequalities and arbitrary coefficients for the equalities, i.e.,

$$\begin{aligned} &\lambda\Big(1 - \sum_b v_b\Big) + \mu\Big(\delta - \sum_b \delta_b\Big) \\ &+ \sum_x \nu_{b1}(v_b + \delta_b) + \sum_x \nu_{b2}(v_b - \delta_b) \\ &+ \sum_{\boldsymbol{s}\in\{\pm\}^{\times n_B}} \xi_{\boldsymbol{s}}\Big(\Delta - \sum_b s_b\delta_b\Big) \geq 0 \end{aligned} \tag{97}$$

subject to the conditions

$$\nu_{bx} \geq 0, \tag{98}$$
$$\xi_{\boldsymbol{s}} \geq 0, \tag{99}$$

where in the last term the summation index $\boldsymbol{s} = (s_1, \ldots, s_{n_B})$ is a vector of signs to use in front of the δ_bs. By grouping the constant terms and terms in v_b and δ_b together we can write the most general possible constraint as

$$\gamma + \sum_b \alpha_b v_b + \sum_b \beta_b \delta_b \geq 0 \tag{100}$$

with

$$\gamma = \lambda + \delta\mu + \Delta\xi, \tag{101}$$
$$\alpha_b = -\lambda + \sigma_b, \tag{102}$$
$$\beta_b = -\mu + \varepsilon_b - \xi_b, \tag{103}$$

and

$$\sigma_x = \nu_{b1} + \nu_{b2}, \tag{104}$$
$$\varepsilon_x = \nu_{b1} - \nu_{b2}, \tag{105}$$
$$\xi = \sum_{\boldsymbol{s}} \xi_{\boldsymbol{s}}, \tag{106}$$
$$\xi_b = \sum_{\boldsymbol{s}} \xi_{\boldsymbol{s}} s_b, \tag{107}$$
$$\nu_{bx} \geq 0, \tag{108}$$
$$\xi_{\boldsymbol{s}} \geq 0. \tag{109}$$

From here, our goal is to eliminate variables until we are left with linear constraints involving only γ and the α_bs and β_bs. According to (100), we can interpret these as a sufficient set of points (v_b, δ_b) to test to determine if (100) is a valid inequality for given values of γ, α_b, and β_b.

We first eliminate the ν_{bx}s. We take the sum and difference of (104) and (105) to get $\sigma_b + \varepsilon_b = 2\nu_{b1}$ and $\sigma_b - \varepsilon_b = 2\nu_{b2}$; combined with $\nu_{bx} \geq 0$ this gives the constraints

$$-\varepsilon_b \leq \sigma_b \leq \varepsilon_b \tag{110}$$

directly on σ_b and ε_b and we can from this point forget about the ν_{bx}s. Similarly, (107) is just expressing that ξ_b are the coordinates of a point that is a 'convex combination of the corners $\{\pm 1\}^{\times n_B}$ of the n_B-dimensional cube, except that the coefficients are normalised to a number $\sum_{\boldsymbol{s}} \xi_{\boldsymbol{s}} = \xi$ instead of one. Thus (106) and (107) are equivalent to

$$-\xi \leq \xi_b \leq \xi. \tag{111}$$

Our set of inequalities thus simplifies to

$$\gamma = \lambda + \delta\mu + \Delta\xi, \tag{112}$$
$$\alpha_b = -\lambda + \sigma_b, \tag{113}$$
$$\beta_b = -\mu + \varepsilon_b - \xi_b, \tag{114}$$

subject to

$$\sigma_b - \varepsilon_b \geq 0, \tag{115}$$
$$\sigma_b + \varepsilon_b \geq 0, \tag{116}$$
$$\xi - \xi_b \geq 0, \tag{117}$$
$$\xi + \xi_b \geq 0. \tag{118}$$

Let us next eliminate σ_b and ξ. We get

$$c - \lambda - \delta\mu \geq \Delta\xi_b, \tag{119}$$
$$c - \lambda - \delta\mu \geq -\Delta\xi_b, \tag{120}$$
$$\alpha_b + \lambda \geq \varepsilon_b, \tag{121}$$
$$\alpha_b + \lambda \geq -\varepsilon_b, \tag{122}$$
$$\beta_b + \mu = \varepsilon_b - \xi_b. \tag{123}$$

Eliminating ε_b then gives

$$\gamma - \lambda - \delta\mu - \Delta\xi_b \geq 0, \tag{124}$$
$$\gamma - \lambda - \delta\mu + \Delta\xi_b \geq 0, \tag{125}$$
$$\alpha_b + \lambda \geq 0\,, \tag{126}$$
$$\alpha_b - \beta_b + \lambda - \mu - \xi_b \geq 0, \tag{127}$$
$$\alpha_b + \beta_b + \lambda + \mu + \xi_b \geq 0, \tag{128}$$

and eliminating ξ_b gives

$$\gamma - \lambda - \delta\mu \geq 0, \quad (129)$$
$$\alpha_b + \lambda \geq 0, \quad (130)$$
$$\gamma + \Delta\alpha_b + \Delta\beta_b - (1-\Delta)\lambda + (\Delta-\delta)\mu \geq 0, \quad (131)$$
$$\gamma + \Delta\alpha_b - \Delta\beta_b - (1-\Delta)\lambda - (\Delta+\delta)\mu \geq 0. \quad (132)$$

There are at this point only two unwanted variables left, λ and μ. Eliminating λ first gives

$$\gamma + \alpha_b - \delta\mu \geq 0, \quad (133)$$
$$\gamma + (1-\Delta)\alpha_b + \Delta\alpha_{b'} + \Delta\beta_{b'} + (\Delta-\delta)\mu \geq 0, \quad (134)$$
$$\gamma + (1-\Delta)\alpha_b + \Delta\alpha_{b'} - \Delta\beta_{b'} - (\Delta+\delta)\mu \geq 0. \quad (135)$$

Note that, here, there can be two different values of the output, b and b', since when we combine (130) with (131) and (132) we have to do it for all possible values of b in both inequalities. Additionally, the first of the inequalities we are left with above is redundant since it can be derived by summing the second and third constraints with $b' = b$ and then dividing by two. Combining the two remaining inequalities to eliminate the last variable μ gives the family of inequalities

$$\begin{aligned}\gamma &+ \left(1+\frac{\delta}{\Delta}\right)\frac{1-\Delta}{2}\alpha_b + \left(1-\frac{\delta}{\Delta}\right)\frac{1-\Delta}{2}\alpha_{b'} \\ &+ \left(1+\frac{\delta}{\Delta}\right)\frac{\Delta}{2}\alpha_{b''} + \left(1-\frac{\delta}{\Delta}\right)\frac{\Delta}{2}\alpha_{b'''} \\ &+ \left(1+\frac{\delta}{\Delta}\right)\frac{\Delta}{2}\beta_{b''} - \left(1-\frac{\delta}{\Delta}\right)\frac{\Delta}{2}\beta_{b'''} \geq 0 \quad (136)\end{aligned}$$

with up to four different indices, b, b', b'', and b''', but many of these are redundant. To begin with, we don't need the inequalities with $b \neq b'$, so the system reduces to

$$\begin{aligned}\gamma &+ (1-\Delta)\alpha_b + \frac{\Delta+\delta}{2}\alpha_{b'} + \frac{\Delta-\delta}{2}\alpha_{b''} \\ &+ \frac{\Delta+\delta}{2}\beta_{b'} - \frac{\Delta-\delta}{2}\beta_{b''} \geq 0 \quad (137)\end{aligned}$$

since all of the inequalities in (136) can be recovered by summing two instances of (137) with different values of b with weights $(1+\delta/\Delta)/2$ and $(1-\delta/\Delta)/2$.

Having now eliminated all the unwanted variables we express (137) as

$$\gamma + \boldsymbol{\alpha}\cdot\boldsymbol{v} + \boldsymbol{\beta}\cdot\boldsymbol{\delta} \geq 0 \quad (138)$$

with

$$v_b = 1-\Delta, \quad v_{b'} = \frac{\Delta+\delta}{2}, \quad v_{b''} = \frac{\Delta-\delta}{2}, \quad (139)$$
$$\delta_b = 0, \quad \delta_{b'} = \frac{\Delta+\delta}{2}, \quad \delta_{b''} = -\frac{\Delta-\delta}{2} \quad (140)$$

and all other vs and δs equal to zero. These terms are additive and combine if some of the indices coincide; for example, if $b = b' \neq b''$ then we use $v_b = 1 - \Delta/2 + \delta/2$ and $\delta_b = \Delta/2 + \delta/2$. Eqs. (139) and (140) identify give a set of points $(\boldsymbol{v}, \boldsymbol{\delta})$ that it is sufficient to test to find if (138), for some given coefficients γ, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$, is a valid inequality for all points satisfying the system (92)–(96) of constraints for $\boldsymbol{v}$ and $\boldsymbol{\delta}$ above. In terms of $p_{bx} = (v_b + \delta_b)/2$ and $p_{b2} = (v_b - \delta_b)/2$ and reintroducing G and q_1 and q_2 via $\Delta = 2G - 1$, $\delta = q_1 - q_2$, and $1 = q_1 + q_2$, these correspond to

$$p_{b1} = 1 - G, \quad p_{b'1} = q_1 + G - 1, \quad p_{b''1} = 0 \quad (141)$$

and

$$p_{b2} = 1 - G, \quad p_{b'2} = 0, \quad p_{b''2} = q_2 + G - 1. \quad (142)$$

Considering different ways of taking the outputs b, b', and b'' the same as or different from each other gives five different kinds of probability distributions, up to relabelling the output. In matrix notation like we used in Section 2.1 they are

$$\begin{pmatrix} q_1 & 0 & 0 \\ q_2 & 0 & 0 \end{pmatrix}, \quad (143)$$
$$\begin{pmatrix} q_1 & 0 & 0 \\ 1-G & q_2+G-1 & 0 \end{pmatrix}, \quad (144)$$
$$\begin{pmatrix} 1-G & q_1+G-1 & 0 \\ q_2 & 0 & 0 \end{pmatrix}, \quad (145)$$
$$\begin{pmatrix} 1-G & q_1+G-1 & 0 \\ 1-G & q_2+G-1 & 0 \end{pmatrix}, \quad (146)$$
$$\begin{pmatrix} 1-G & q_1+G-1 & 0 \\ 1-G & 0 & q_2+G-1 \end{pmatrix}. \quad (147)$$

With the exception of (146), which is not a vertex, dividing the first and second rows by the prior probabilities q_1 and q_2 gives the vertices asserted in Section 2.1. We can see that (146) is not a vertex by noticing that it can be obtained from some of the other matrices above. Specifically,

$$\begin{aligned}\begin{pmatrix} 1-G & G-q_1 \\ 1-G & G-q_2 \end{pmatrix} &= \theta_1 \begin{pmatrix} 0 & q_1 \\ 0 & q_2 \end{pmatrix} + \theta_2 \begin{pmatrix} q_1 & 0 \\ 1-G & G-q_2 \end{pmatrix} \\ &+ \theta_3 \begin{pmatrix} 1-G & G-q_1 \\ q_2 & 0 \end{pmatrix} \quad (148)\end{aligned}$$

for the convex coefficients

$$\theta_1 = \frac{(G-q_1)(G-q_2)}{q_1q_2 - (1-G)^2}, \quad (149)$$
$$\theta_2 = \frac{(1-G)(G-q_1)}{q_1q_2 - (1-G)^2}, \quad (150)$$
$$\theta_3 = \frac{(1-G)(G-q_2)}{q_1q_2 - (1-G)^2}. \quad (151)$$

B Characterisation of the $(2, 1, 2)$ scenario

B.1 Characterisation of $\mathcal{C}_{\text{det}}$

Let us begin by considering this scenario when there is no shared randomness. In that case, the problem

is trivial, Alice's messages depend only on x, and if the guessing probability bound G is anything strictly less than one then the only possibility is that Alice sends the same message in both cases, in which case the resulting probabilities must be the same. Therefore without shared randomness, the correlations set collapses to a line $E_1 = E_2$.

In the following we suppose that $q_1 > q_2$ without loss of generality. If shared randomness is available, then Alice can sometimes send the same message (with associated guessing probability q_1) and sometimes send different messages (with guessing probability one) as long as the average guessing probability remains smaller than G.

If Alice sends the same message, Bob can generate the following extremal probabilities

$$(E_1, E_2) = (+1, +1) \text{ or } (-1, -1), \tag{152}$$

while if Alice sends different messages Bob can generate the extremal probabilities

$$(E_1, E_2) = (+1, +1), (+1, -1), (-1, +1), \text{ or } (-1, -1). \tag{153}$$

The extremal probabilities that Bob can generate overall are combinations of (152) with some probability θ and (153) with some probability $1 - \theta$. We should use

$$\theta = \frac{1 - G}{q_2} \qquad \text{and} \qquad 1 - \theta = \frac{G - q_1}{q_2} \tag{154}$$

which are chosen such that

$$\theta \cdot q_1 + (1 - \theta) \cdot 1 = G, \tag{155}$$

in order to respect the guessing probability bound of G on average. (We could make these inequalities rather than equalities, but this is unnecessary since (153) includes the two extremal points in (152) and any excess in the value of θ could be absorbed into that.) After eliminating two redundant ones this yields six vertices,

$$(E_1, E_2) = (+\mu, +1), (+1, +1), (+1, +\mu), (-\mu, -1), (-1, -1), (-1, -\mu) \tag{156}$$

with

$$\mu = \frac{1 + q_1 - 2G}{q_2}. \tag{157}$$

All probabilities represented by values (E_1, E_2) in this scenario must be convex combinations of these six vertices. In addition to the trivial conditions $|E_x| \leq 1$, this implies two facet inequalities,

$$|E_1 - E_2| \leq 2\frac{G - q_1}{q_2}. \tag{158}$$

B.2 Characterisation of $\mathcal{Q}_{\mathsf{det}}$

The problem is very similar to a quantum set studied in Section 3.1 in [17]. For pure states, the guessing probability associated to the ensemble $\mathcal{E} = \{q_1, \psi_1; q_2, \psi_2\}$ is

$$P_{\mathrm{g}}(X|\mathcal{E}) = \frac{1}{2} + \frac{1}{2}\sqrt{1 - 4q_1 q_2 \big|\langle\psi_1|\psi_2\rangle\big|^2}. \tag{159}$$

Assuming the guessing probability satisfies $P_{\mathrm{g}}(X|\mathcal{E}) \leq G$ for some bound G and rearranging for the inner product gives

$$\big|\langle\psi_1|\psi_2\rangle\big|^2 \geq \frac{G(1 - G)}{q_1 q_2}. \tag{160}$$

In the following we derive what this implies for a linear combination

$$W = c_1 E_1 - c_2 E_2 \tag{161}$$

of correlation terms $E_x = \mathrm{Tr}[E\psi_x]$ with $-\mathbb{1} \leq E \leq \mathbb{1}$. We remark first that the witness W is trivial if the coefficients c_1 and c_2 are not of the same sign because the positivity constraints $E_x \leq 1$ alone imply

$$c_1 E_1 - c_2 E_2 \leq |c_1| + |c_2|, \tag{162}$$

which is trivially attained with $E_1 = E_2 = \pm 1$ when the coefficients are of opposite signs. We thus concentrate on the case that c_1 and c_2 are both of the same sign. In the rest of this section we suppose without loss of generality that c_1 and c_2 are nonnegative and that $c_1 + c_2 = 1$. We also suppose for simplicity that $q_1 \geq q_2$.

Bounding W with c_1 and c_2 taken to have the same sign gives

$$\begin{aligned} c_1 E_1 - c_2 E_2 &= \mathrm{Tr}\big[E(c_1\psi_1 - c_2\psi_2)\big] \\ &\leq \big\|c_1\psi_1 - c_2\psi_2\big\|_1 \\ &= \sqrt{(c_1 + c_2)^2 - 4c_1 c_2 \big|\langle\psi_1|\psi_2\rangle\big|^2} \\ &= \sqrt{1 - 4c_1 c_2 \big|\langle\psi_1|\psi_2\rangle\big|^2}, \end{aligned} \tag{163}$$

where we substituted $c_1 + c_2 = 1$ in the last line. Combining this with the bound (160) on $|\langle\psi_1|\psi_2\rangle|$ in terms of G gives

$$c_1 E_1 - c_2 E_2 \leq \sqrt{1 - \frac{4c_1 c_2}{q_1 q_2} G(1 - G)}. \tag{164}$$

The inequality (164) gives a tight upper bound on the witness to the left in terms of the guessing probability assuming Alice sends one of two pure states $|\psi_x\rangle$ with probabilities q_x. To generalise to allow shared randomness we need to take the convex hull of the right side of (164). Fortunately this is straightforward. The right side of (164) is convex if

$$\frac{c_1 c_2}{q_1 q_2} \geq 1 \tag{165}$$

and concave otherwise; this can be determined by computing the second derivative of the family of functions $f_Q(x) = \sqrt{1 - 4Qx(1 - x)}$.

The condition identifying convexity is satisfied under two conditions: if $c_1 \geq q_1$ or if $c_1 \leq q_2$ (remember we are supposing $q_1 \geq 1/2 \geq q_2$). In this case we need to interpolate (164) between the extreme values $G = q_1$ and $G = 1$. This gives

$$c_1E_1 - c_2E_2 \leq \frac{1-G}{q_2}|c_1 - c_2| + \frac{G-p_1}{q_2}. \tag{166}$$

Supposing $c_1 \geq q_1 \geq q_2 \geq c_2$ gives

$$c_1E_1 - c_2E_2 \leq \frac{1}{q_2}\Big((1-G)(c_1-c_2) + G - q_1\Big), \tag{167}$$

which simplifies to

$$c_1E_1 - c_2E_2 \leq \frac{1}{q_2}\Big(c_1 - q_1 + (2G-1)c_2\Big). \tag{168}$$

Most of this family of inequalities is redundant, since it is implied by the special case with $c_x = q_x$,

$$q_1E_1 - q_2E_2 \leq 2G - 1, \tag{169}$$

and the trivial inequality $E_1 \leq 1$. This can be seen by rewriting (168) as

$$\frac{c_1-q_1}{q_2}E_1 + \frac{c_2}{q_2}\big(q_1E_1 - q_2E_2\big) \leq \frac{c_1-q_1}{q_2} + \frac{c_2}{q_2}\big(2G-1\big). \tag{170}$$

Similarly, if $c_2 \geq q_1 \geq q_2 \geq c_1$, we get a family of inequalities that is the same as (168) except with c_1 and c_2 interchanged on the right side,

$$c_1E_1 - c_2E_2 \leq \frac{1}{q_2}\Big(c_2 - q_1 + (2G-1)c_1\Big), \tag{171}$$

but only the special case with $c_1 = q_2$ and $c_2 = q_1$, i.e.,

$$q_2E_1 - q_1E_2 \leq 2G - 1, \tag{172}$$

is not implied by other inequalities. This confirms that the only nontrivial linear inequalities satisfied by correlations in the quantum deterministic set are precisely (164) for $q_2 \leq c_1, c_2 \leq q_1$.

Note that part of the boundary of the quantum set coincides with an ellipse, characterised by

$$(1-\gamma)(E_1+E_2)^2 + \gamma(E_1-E_2)^2 = 4\gamma(1-\gamma), \tag{173}$$

for $\gamma = G(1-G)/(q_1q_2)$.

C Finite message dimension in classical scenarios

Similarly to appendix A, we work with joint probabilities

$$p_{mx} = q_x p(m|x) \tag{174}$$

with the priors q_x absorbed. These as well as allowed values of the upper bound G on the guessing probability are characterised by

$$q_x - \sum_m p_{mx} = 0, \qquad \forall m, \tag{175}$$

$$p_{mx} \geq 0, \qquad \forall m, x, \tag{176}$$

$$G - \sum_m p_{mx_m} \geq 0, \qquad \forall \boldsymbol{x} = (x_m). \tag{177}$$

The most general family of inequalities implied by this is

$$\sum_x \xi_x \Big(q_x - \sum_m p_{mx}\Big) + \mu_{mx}\, p_{mx} + \sum_{\boldsymbol{x}} \lambda_{\boldsymbol{x}} \Big(G - \sum_m p_{mx_m}\Big) \geq 0, \tag{178}$$

for any ξ_x and any nonnegative μ_{mx} and $\lambda_{\boldsymbol{x}}$. We can express this as

$$\gamma + \sum_{mx} \alpha_{mx} p_{mx} + \beta G \geq 0. \tag{179}$$

where

$$\gamma = \sum_x \xi_x q_x, \tag{180}$$

$$\alpha_{mx} = -\xi_x + \mu_{mx} - \lambda_{mx}, \tag{181}$$

$$\beta = \lambda, \tag{182}$$

$$\lambda = \sum_{\boldsymbol{x}} \lambda_{\boldsymbol{x}}, \tag{183}$$

$$\lambda_{mx} = \sum_{\boldsymbol{x}} \lambda_{\boldsymbol{x}} \delta_{x x_m}, \tag{184}$$

$$\lambda_{\boldsymbol{x}} \geq 0, \tag{185}$$

$$\mu_{mx} \geq 0, \tag{186}$$

where $\delta_{xx'}$ is the Kronecker delta.

We aim to simplify this system to obtain the most straightforward possible constraints for γ, α_{mx}, and β. First, to eliminate the $\lambda_{\boldsymbol{x}}$s, note that (183), (184), and (185) imply

$$\lambda_{mx} \geq 0 \qquad \text{and} \qquad \sum_x \lambda_{mx} = \lambda. \tag{187}$$

Conversely, any variables λ_{mx} that satisfy these conditions can be written in the form (184) with nonnegative $\lambda_{\boldsymbol{x}}$s, for example with

$$\lambda_{\boldsymbol{x}} = \lambda^{-(n_M-1)} \prod_m \lambda_{mx_m}, \tag{188}$$

where n_M is the number of different values of the variable m. Substituting also $\lambda = \beta$ simplifies the system to

$$\gamma = \sum_x \xi_x q_x, \tag{189}$$

$$\alpha_{mx} = -\xi_x + \mu_{mx} - \lambda_{mx}, \tag{190}$$

$$\beta = \sum_x \lambda_{mx}, \tag{191}$$

$$\lambda_{mx} \geq 0, \tag{192}$$

$$\mu_{mx} \geq 0. \tag{193}$$

Eliminating the λ_{mx}s gives

$$\gamma = \sum_x \xi_x q_x, \tag{194}$$

$$\beta = \sum_x \big(-\alpha_{mx} - \xi_x + \mu_{mx}\big), \tag{195}$$

$$\mu_{mx} \geq \alpha_{mx} + \xi_x, \tag{196}$$

$$\mu_{mx} \geq 0, \tag{197}$$

and eliminating the μ_{mx}s gives

$$\gamma - \sum_x q_x \xi_x = 0, \tag{198}$$

$$\beta + \sum_x u_x \big(\alpha_{mx} + \xi_x\big) \geq 0, \quad \forall m, u_x \in \{0,1\}. \tag{199}$$

The last step would consist of eliminating the ξ_xs. Note first that the instance $u_x = 0$ for all x of (199) gives an inequality

$$\beta \geq 0 \tag{200}$$

which does not involve any of the variables ξ_x. This corresponds to the (unique) conic generator

$$\big(p(m|x), G\big) = (0, 1) \tag{201}$$

of the polyhedron $\mathcal{M}^+$ which, in turn, just expresses a property of $\mathcal{M}^+$ that was already evident from its definition: we can increase the guessing probability bound component G of any point $\big(p(m|x), G\big)$ in $\mathcal{M}^+$, by adding any nonnegative multiple of (201) to it, and the resulting point will still be in $\mathcal{M}^+$.

For the remaining instances of (199), we seek to bound the maximum number of different values of the message index m that can appear in any inequality in the process of eliminating the n_X variables ξ_x. We rewrite the problem as

$$\gamma = \sum_x q_x \xi_x, \tag{202}$$

$$\sum_x u_x \xi_x \geq -\beta - \sum_x u_x \alpha_{mx}, \tag{203}$$

to make it clear that the initial inequalities (203) all give lower bounds on the ξ_xs and the problem can be seen as combining (202) with sums of instances of (203) such that the left side equals $\sum_x q_x \xi_x$. Eliminating first one of the ξ_xs, which consists of combining (202) with all the instances of (203) in which the chosen variable ξ_x appears with a nonzero coefficient u_x, yields a system of inequalities that each involve only one value of m. The process of eliminating the remaining $n_X - 1$ variables ξ_x can then at worst double the number of different values of m appearing in the inequalities at each step. The inequalities we obtain for γ, α_{mx}, and β at the end of this process can thus not involve more than $2^{n_X - 1}$ different values of the index m. With the the exception of (200) we can write all of them in the form

$$\gamma + \sum_{mx} \alpha_{mx} p_{mx} + \beta G, \tag{204}$$

from which we infer that the vertices of $\mathcal{M}^+$ are strategies $\big(p(m|x), G\big)$ in which no more than $2^{n_X - 1}$ different messages m are used in each strategy, i.e., in a matrix notation

$$\big(p(m|x)\big) = \begin{pmatrix} p(1|1) & p(2|1) & \cdots \\ p(1|2) & p(2|2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{205}$$

the components $p(m|x)$ of the vertices of $\mathcal{M}^+$ all have at most $2^{n_X - 1}$ columns containing nonzero entries.

At this point we remark that we have not restricted the number of messages m used overall, which is simply whatever number n_M of different values of m we allow to appear in the problem from the beginning, since the $2^{n_X - 1}$ messages used in each vertex will generally be different for each vertex. Remember, however, that we are not interested in the communication strategies represented by $\mathcal{M}^+$ themselves but the extremal correlations

$$p(b|x, y) = \sum_m p(m|x) p(b|y, m) \tag{206}$$

that can ultimately be generated with them, which also depend on Bob's extremal responses $p(b|y, m)$, and we can use a symmetry of the setting to reduce the communication strategies we need to consider. In particular, both the sets of extremal communication strategies $\big\{\big(p(m|x), G\big)\big\}$ and of Bob's extremal responses $\{p(b|y, m)\}$ are symmetric with respect to relabellings of the messages, under which (206) is also invariant. We can hence limit the number of messages n_M we need to consider to $2^{n_X - 1}$ for the purpose of generating the extremal points $\big(p(b|x, y), G\big)$ of $\mathcal{C}^+$, as allowing more messages will only result in more ways of generating the same correlations $p(b|x, y)$ through (206).