Informationally restricted quantum correlations

Armin Tavakoli[1], Emmanuel Zambrini Cruzeiro[1], Jonatan Bohr Brask[2], Nicolas Gisin[1], and Nicolas Brunner[1]

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

[2]Department of Physics, Technical University of Denmark, Fysikvej, 2800 Kongens Lyngby, Denmark

Quantum communication leads to strong correlations, that can outperform classical ones. Complementary to previous works in this area, we investigate correlations in prepare-and-measure scenarios assuming a bound on the information content of the quantum communication, rather than on its Hilbert-space dimension. Specifically, we explore the extent of classical and quantum correlations given an upper bound on the one-shot accessible information. We provide a characterisation of the set of classical correlations and show that quantum correlations are stronger than classical ones. We also show that limiting information rather than dimension leads to stronger quantum correlations. Moreover, we present device-independent tests for placing lower bounds on the information given observed correlations. Finally, we show that quantum communication carrying $\log d$ bits of information is at least as strong a resource as d-dimensional classical communication assisted by pre-shared entanglement.

1 Introduction

Separated parties, initially independent, can become correlated via communication. Intuitively, more communication enables stronger correlations. Also, the strength of the correlations may vary depending on the nature of the communication; for example if the message is carried by a quantum system rather than a classical one. In general, understanding the relation between communication and correlations is a fundamental question, at the intersection of information theory and physics.

Consider a simple scenario (see Fig. 1) with two separated parties. A first party, Alice, receives an input x and sends a message to a second party, Bob. Upon receiving this message, as well as some input y, Bob produces an output b. When repeated many times (with inputs x and y randomly sampled), this experiment is described by the conditional probability distribution $p(b|x,y)$ which characterises the correlations between Alice and Bob. Clearly, the amount of information about x encoded in Alice's message determines the strength of the possible correlations. If Alice sends no message at all (or if the message is independent of x), then no correlations are generated, i.e. $p(b|x,y) = p(b|y)$. On the other hand, if the message perfectly encodes x, then maximal correlations can be established; any distribution $p(b|x,y)$ is possible. Thus the main question is: how strong correlations can be established provided that the amount of communication from Alice to Bob is quantitatively limited?

Naturally, the answer depends on how exactly communication is quantified. The most common approach consists in measuring communication via the dimension of the message, i.e. the number of bits the message could carry. This is used in the field of communication complexity (see e.g. [1]), where the goal is to find out how the minimum dimension required to solve a problem (i.e. demanding that the output b corresponds to a certain function of the inputs x and y) scales with the problem size. Notably, the use of quantum communication is advantageous since it allows one to solve certain problems with exponentially smaller dimension [2, 3]. In parallel, there has been interest in characterising the set of possible correlations $p(b|x,y)$ for classical and quantum systems of bounded dimension [4–7]. Again, quantum correlations turn out to be stronger than classical ones. This led to a novel framework for quantum information processing termed "semi-device-independent" [8–12], where devices are assumed to process quantum

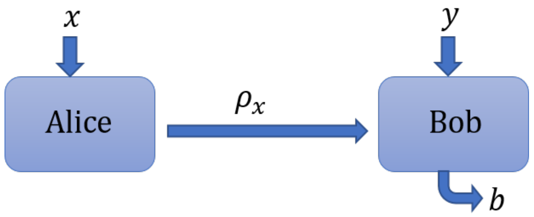

Figure 1: Prepare-and-measure scenario. In this work we investigate the strength of possible correlations $p(b|x,y)$ given a limit on the information carried by the quantum message ρ_x.

systems of bounded dimension, but are otherwise uncharacterised.

However, measuring communication via the dimension provides only a partial characterisation. Information-theoretic concepts are typically better suited to get a complete picture. This raises a natural question, namely to understand the relation between the strength of correlations and the amount of information that the communication contains. But then, information about what? In correlation experiments, the answer is very natural: we are interested in the information that the message contains about Alice's input x.

Here we formalise this problem and investigate classical and quantum correlations for informationally restricted communication. Naturally, however, there are many different ways of quantifying information based on entropies. We quantify the information content of an ensemble of prepared states (classical or quantum) via a one-shot version of accessible information based on min-entropies [13]. This choice of information measure has a two-fold motivation. Firstly, it admits a simple operational interpretation in terms of how well one could determine Alice's input from her message, via the best possible measurement. Secondly, it proves convenient to work with. Our approach is clearly complementary to previous works based on dimension. Firstly, information is a continuous quantity, while dimension is discrete; one can consider ensembles of states carrying only half a bit of information about Alice's input, which would have no analogue using dimension. Secondly, even when considering ensembles of states carrying $\log d$ bits of information (for some dimension d), there exist ensembles of dimension $d' > d$ that carry no more than $\log d$ bits of information, e.g. certain ensembles of non-orthogonal quantum states.

In this work, we develop a framework for characterising informationally restricted correlations. For the case of classical systems, we show that the relevant set of correlations forms a convex polytope, which can be fully characterised. This allows one to find the minimal amount of information required to reproduce a given correlation using classical communication. In turn, we prove that quantum correlations can be stronger than classical ones. Moreover, we derive device-independent lower bounds on the information, given observed correlations. These ideas are illustrated in a simple scenario. We also show that ensembles of higher-dimensional quantum states carrying no more than one bit of information can generate stronger correlations than two-dimensional quantum systems (i.e. qubits). Finally, we show that any correlations achievable with classical communication (of a d-dimensional message) assisted by pre-shared entanglement can also be achieved using quantum communication carrying $\log d$ bits of information.

2 Setting

We start by defining informationally restricted correlations in a quantum prepare-and-measure scenario. The sender, Alice, receives an input $x \in [n]$ sampled from a random variable X (where $[s] = \{1, \ldots, s\}$) which she encodes into a quantum state ρ_x that she relays to the receiver, Bob. Bob also receives a random input $y \in [l]$ and then measures the received state with some generalised measurement (positive operator-valued measure, POVM) $\{M_{b|y}\}$ with outcome $b \in [k]$. The observed correlations are

$$p(b|x,y) = \operatorname{tr}\left(\rho_x M_{b|y}\right). \tag{1}$$

Let us now characterise the information in Alice's message about her input x. Since x is random, sampled from some distribution $p_X(x)$, the ensemble of messages is given by $\mathcal{E} = \{p_X(x), \rho_x\}$. How well could an observer, via any possible POVM $\{N_z\}$, guess x from $\mathcal{E}$? The *guessing probability* is

$$P_g(X|\mathcal{E}) = \max_{\{N_z\}} \sum_{x=1}^{n} p_X(x) \operatorname{tr}\left[\rho_x N_x\right]. \tag{2}$$

Note that the optimal POVM, $\{N_z^*\}$, does not need to be part of set of POVMs $\{M_{b|y}\}$. Hence the statistics obtained from measuring $\{N_z^*\}$ do not necessarily appear in the correlations $p(b|x,y)$.

The observer's minimal uncertainty about X when provided $\mathcal{E}$, i.e. the conditional min-entropy, is $H_{\min}(X|\mathcal{E}) = -\log\left[P_g(X|\mathcal{E})\right]$. The amount of information gained by learning $\mathcal{E}$, i.e. the information carried by $\mathcal{E}$, is then the difference in uncertainty without and with the communication [13];

$$\mathcal{I}_X(\mathcal{E}) = H_{\min}(X) - H_{\min}(X|\mathcal{E}), \tag{3}$$

where $H_{\min}(X) = -\log\left[\max_x p_X(x)\right]$ is the min-entropy. The quantity $\mathcal{I}_X(\mathcal{E})$ can be viewed as a single-shot version of accessible information [14, 15]. Note that for any given ensemble $\mathcal{E}$, the guessing probability (and hence the information) can be computed via a semidefinite program [16].

We can now define the set of possible correlations $p(b|x,y)$ when the information of the message is upper bounded. Importantly, we do not limit the Hilbert-space dimension for representing the set of the quantum states $\{\rho_x\}$. We also allow for shared randomness between Alice's and Bob's devices. This makes the model more general, and at the same time simplifies the characterisation of the sets of correlations (as these sets are now convex). Formally, we define the set $\mathcal{S}_\alpha^{\mathrm{Q}}$ of correlations of the form

$$p(b|x,y) = \sum_\lambda p(\lambda)\,\mathrm{tr}\left(\rho_x^{(\lambda)} M_{b|y}^{(\lambda)}\right), \tag{4}$$

where λ denotes the shared classical variable, distributed according to $p(\lambda)$, and the information is bounded by $\mathcal{I}_X \leq \alpha$. The quantity $\mathcal{I}_X$ is computed via Eq. (3), considering the average guessing probability of the ensemble $\mathcal{E} = \{p(\lambda), \mathcal{E}_\lambda\}$:

$$P_g(X|\mathcal{E}) = \sum_\lambda p(\lambda) P_g(X|\mathcal{E}_\lambda), \tag{5}$$

where $P_g(X|\mathcal{E}_\lambda)$ denotes the guessing probability for the subensemble $\mathcal{E}_\lambda = \{p_X(x), \rho_x^{(\lambda)}\}$.

3 Classical correlations

Similarly to above, we can characterise the set of classical correlations, $\mathcal{S}_\alpha^{\mathrm{C}}$, subject to an information bound. In this setting, Alice encodes x into a classical message $m \in [d]$. Bob then provides an output based on his input y and the message m. Considering again shared randomness, the resulting correlations take the form

$$p(b|x,y) = \sum_\lambda p(\lambda) \sum_{m=1}^{d} p_{\mathrm{A}}(m|x,\lambda) p_{\mathrm{B}}(b|m,y,\lambda). \tag{6}$$

In order to characterise correlations of the above form such that $\mathcal{I}_X \leq \alpha$, we proceed as follows. First, notice that the dimension d of the message may a priori be unbounded. However, it turns out that, without loss of generality, one can restrict to the case $d = n$. Next, notice that each encoding of the message $p_{\mathrm{A}}(m|x,\lambda)$ can be taken to be deterministic, i.e. m is a deterministic function of x and λ. Finally, to each of these deterministic encodings, we can associate a guessing probability $P_g^{(\lambda)}$. A detailed discussion is given in Appendix.

With these in hand, we notice that the constraint $\mathcal{I}_X \leq \alpha$ is equivalent to $\sum_\lambda p(\lambda) P_g^{(\lambda)} \leq 2^{\alpha - H_{\min}(X)}$, which is linear in $p(\lambda)$. Therefore, the set $\mathcal{S}_\alpha^{\mathrm{C}}$ forms a convex polytope. The facets of the polytope correspond to linear inequalities

$$\sum_{x,y,b} r_{xyb}\, p(b|x,y) \leq \beta \tag{7}$$

where r_{xyb} and β are real coefficients, which give a complete characterisation of $\mathcal{S}_\alpha^{\mathrm{C}}$.

We have explicitly characterised $\mathcal{S}_\alpha^{\mathrm{C}}$ for scenarios featuring a small number[1] of inputs and outputs. We find three types of facet inequalities: (i) positivity conditions, e.g. $p(b|x,y) \geq 0$, (ii) inequalities ensuring the information bound on the observed correlations, e.g. $\sum_x p(b = x|x,y) \leq 2^{\alpha - H_{\min}(X)}$ (assuming here $n = k$), and (iii) other inequalities. Inequalities (i) and (ii) are in a sense trivial, as they must be satisfied by all physical correlations (when assuming $\mathcal{I}_X \leq \alpha$). On the contrary, inequalities (iii) are non-trivial, and thus capture limits of classical correlations. These inequalities do not necessarily hold for quantum correlations, as we show below.

Finally, note that the problem of determining whether some observed correlations $p(b|x,y)$ can be obtained classically with $\mathcal{I}_X \leq \alpha$ bits of infor-

[1] Typically, characterising $\mathcal{S}_\alpha^{\mathrm{C}}$ quickly becomes computationally demanding as we increase the number of inputs and outputs (the number of vertices grows rapidly). While we could solve cases with $n = 2, 3$ efficiently and the case of $n = 4$ preparations within reasonable time, evaluating $n = 5$ preparations becomes time-consuming on a standard desktop computer.

mation is a linear program. One can thus determine the minimal amount of information required to produce $p(b|x,y)$ in a classical protocol.

4 Quantum advantage

A critical question is whether informationally restricted quantum correlations can outperform their classical counterparts (and thereby provide a quantum advantage). To answer this question, we have considered simple scenarios – labelled by the number of inputs and outputs, i.e. (n,l,k) – and characterised their classical polytope $\mathcal{S}_\alpha^{\mathrm{C}}$. Alice's input is always chosen to be uniformly distributed, i.e. $p_X(x)=1/n$. The simplest scenario where we could find a non-trivial facet inequality is (3,2,2). We conjecture that $n\geq 3$ is necessesary (we have checked that no quantum advantage is possible for $(2,2,2)$ and $(2,2,3)$).

The scenario $(3,2,2)$ features two non-trivial facets showing a quantum advantage (see Appendix) . Here we focus on one of them:

$$F_1\equiv -E_{11}-E_{12}-E_{21}+E_{22}+E_{31}\leq 2^{\alpha+1}-1 \quad (8)$$

where $E_{xy}=p(0|x,y)-p(1|x,y)$ and $\mathcal{I}_X\leq\alpha\in[0,\log 3]$. Notice that for $\alpha=1$, this inequality is identical to the simplest dimension witness of Ref. [4] for classical bits.

Importantly, the above inequality can be violated in quantum theory whenever[2] $\mathcal{I}_X\in(0,\log 3)$, as illustrated in Fig. 2. Let Alice and Bob share one bit of randomness ($\lambda\in\{0,1\}$) with distribution $q\equiv p(\lambda=0)$. When $\lambda=0$, Alice prepares the qubit ensemble $\mathcal{E}_0=\{\frac{1}{3},|\psi_x\rangle\}$ with $|\psi_1\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$, $|\psi_2\rangle=|0\rangle$ and $|\psi_3\rangle=\sin\frac{\pi}{8}|0\rangle-\cos\frac{\pi}{8}|1\rangle$. Bob measures the observables $-\frac{\sigma_x+\sigma_z}{\sqrt{2}}$ and $\frac{\sigma_z-\sigma_x}{\sqrt{2}}$, where $(\sigma_x,\sigma_y,\sigma_z)$ are the Pauli matrices. When $\lambda=1$, Alice sends no information and Bob outputs $b=1$ regardless of y. This strategy results in the witness value $F_1=1+2\sqrt{2}q$, while the information is $\mathcal{I}_X=\log(1+q)$. Thus, this strategy is relevant in the range $\mathcal{I}_X\in[0,1]$. When $\mathcal{I}_X\in[1,\log(3)]$, we consider another mixed strategy. For $\lambda=0$ we use again the ensemble $\mathcal{E}_0$ and associated measurements, and for $\lambda=1$ a qutrit strategy in which Alice sends x to Bob, thus attaining the maximal value of $F_1=5$. We get $F_1=\left(1+2\sqrt{2}\right)q+5\left(1-q\right)$ and $\mathcal{I}_X=\log(3-q)$.

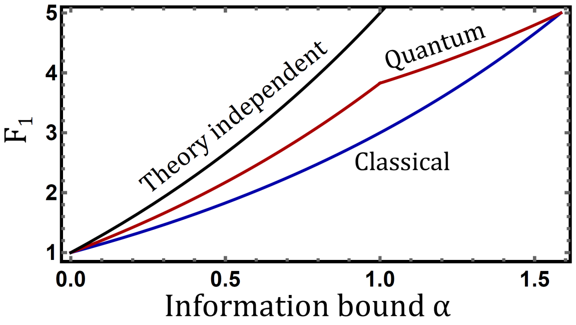

Figure 2: Witness value F_1 as a function of the information bound $\mathcal{I}_X\leq\alpha$. Classical correlations necessarily satisfy the inequality (8) (blue curve). Quantum correlations outperform classical ones for $\alpha\in(0,\log 3)$; the red curve is obtained by a family of quantum protocols and therefore constitutes a lower bound on the optimal quantum correlations. The black curve represents a lower bound on theory-independent correlations, i.e. a lower bound on the information needed for the value F_1.

An interesting question is to find the optimal value of F_1 for any possible quantum strategy with bounded information. This is a non-trivial question as one should consider quantum systems of arbitrarily large Hilbert-space dimension. Based on numerical search, we show in Appendix the existence of slightly better quantum strategies than the above one, but we did not find a simple parameterisation for them.

5 Device-independent bounds on information

While determining the limits of quantum correlations for limited information is challenging, we can nevertheless infer a general, theory-independent, lower bound on information given observed correlations $p(b|x,y)$.

The assumption $\mathcal{I}_X\leq\alpha$ implies that, from any of the distributions $\{p(b|x,1),\ldots,p(b|x,l)\}$, one cannot extract more than α bits of information about x. Allowing for an arbitrary post-processing of the data (Bob creating a new output b' from y and b), i.e. $p(b'|y,b)\geq 0$ with $\sum_{b'}p(b'|y,b)=1$ where $b'\in[n]$, we obtain the constraints

$$\forall y:\quad \sum_{x,b}p_X(x)p(b|x,y)p(b'=x|y,b)\leq 2^{\alpha-H_{\min}(X)}.$$

[2] The extremal cases $\mathcal{I}_X\in\{0,\log 3\}$ are trivial since they correspond to no information and relaying x respectively.

Determining whether a given correlation $p(b|x,y)$ is compatible with the above constraints can be cast as a linear program. If the program admits no feasible solution, then an information $\mathcal{I}_X > \alpha$ is necessary to reproduce $p(b|x,y)$. Note that, while the above constraints are necessary to ensure that $\mathcal{I}_X \leq \alpha$, they are most likely not sufficient in general. How to derive stronger constraints on information is an interesting open problem.

To illustrate the relevance of these ideas, we have derived a lower bound on $\mathcal{I}_X$ given an observed value of the witness F_1. The results are illustrated in Fig. 2 and demonstrate the possibility of certifying a device-independent lower bound on the information. Note that the bound applies to quantum correlations, and more generally to any operational theory.

6 Information versus dimension

Another relevant question is to compare quantum correlations with bounded information to those achievable with bounded dimension. Such comparison makes sense when $\mathcal{I}_X \leq \log d$, where d is the Hilbert-space dimension of the quantum systems. Clearly, any correlation achieved via d-dimensional systems (qudits) requires at most $\mathcal{I}_X = \log d$, as any ensemble of qudits carries no more than $\log d$ bits of information [14]. However, it turns out that there are quantum correlations not achievable via qudits that can nevertheless be obtained with information $\mathcal{I}_X = \log d$.

Specifically, we consider the case $d = 2$ and exhibit quantum correlations achievable with $\mathcal{I}_X = 1$ that cannot be obtained from qubits. Consider a Random Access Code [17–19] in which Alice receives a uniformly random four-bit input $x = (x_1, x_2, x_3, x_4) \in [2]^4$. Bob has settings $y \in [4]$, and returns a binary output b with which he aims to guess x_y. The score is

$$F_{\text{RAC}} = \frac{1}{64}\sum_{x,y} p(b = x_y|x,y). \tag{9}$$

Qubit strategies must satisfy $F_{\text{RAC}} < 3/4$; this follows from the impossibility of having four mutually unbiased bases for qubits [12, 18]. Moreover, numerical optimisation strongly suggests that $F_{\text{RAC}} \leq 0.741$ for qubits [18].

It is nevertheless possible to obtain the score $F_{\text{RAC}} = 3/4$ using quantum ensembles with $\mathcal{I}_X = 1$. The strategy employs 16 four-dimensional quantum states of the form

$$\rho_x = \frac{1}{8}\Big(2\mathbb{1}\otimes\mathbb{1} - (-1)^{x_4}\mathbb{1}\otimes\sigma_y - (-1)^{x_1}\sigma_x\otimes\sigma_x - (-1)^{x_2}\sigma_y\otimes\sigma_x - (-1)^{x_3}\sigma_z\otimes\sigma_x\Big), \tag{10}$$

and Bob measures the observables $B_1 = \sigma_x \otimes \sigma_x$, $B_2 = \sigma_y \otimes \sigma_x$, $B_3 = \sigma_z \otimes \sigma_x$ and $B_4 = \mathbb{1} \otimes \sigma_y$. Note that, despite being four-dimensional, these states are noisy (with purity $\text{tr}\left(\rho_x^2\right) = 1/2\ \forall x$) and carry only one bit of information. Since all states have the same spectrum, $(1/2, 1/2, 0, 0)$, this can be checked analytically as follows. For any quantum ensemble, the information is upper bounded by

$$I_X \leq \log(d) + \log\left(\frac{\max_x p_X(x)\lambda_{\max}(\rho_x)}{\max_x p_X(x)}\right), \tag{11}$$

where $\lambda_{\max}(\rho_x)$ is the largest eigenvalue of ρ_x, and d the Hilbert-space dimension. The bound is obtained from using the relation $\text{tr}[\rho_x N_x] \leq \lambda_{\max}(\rho_x)\,\text{tr}[N_x]$ in Eq. (2) and then $\sum_x N_x = \mathbb{1}_d$. The bound Eq. (11) is tight when (i) for each x, ρ_x only has one non-zero eigenvalue (with possible multiplicity) and (ii) $p_X(x)\lambda_{\max}(\rho_x)$ is constant in x. The ensemble in Eq. (10) satisfies this criteria.

An interesting question is whether larger separation is possible. That is, how much stronger can quantum correlations with $\mathcal{I}_X = \log d$ bits of information become compared to quantum correlations using d-dimensional quantum systems. In Appendix, we show that, in a scenario without shared randomness, this advantage can be made arbitrarily large. Specifically, we construct quantum correlations achievable with $\mathcal{I}_X = 1$ bit of information, that can only be reproduced using an arbitrary large Hilbert-space dimension.

7 Quantum communication versus entanglement-assisted classical communication

Informationally restricted quantum systems also have interesting implications when comparing quantum resources in different communication scenarios. On the one hand, Alice may send an amount of quantum communication to Bob (as in Fig. 1). On the other hand, Alice and Bob may share unlimited entanglement and Alice communicates the same amount classically. These two approaches are generally not equivalent. In fact, for dimensionally restricted classical and quantum messages, there is no strict hierarchy. In some cases, quantum communication outperforms entanglement-assisted classical communication [20–22] and vice versa in others [22–24]. Given this seemingly complicated picture, no generally valid criterion is known for determining which quantum resource is more efficient for a given communication task. Interestingly, we show that every correlation obtained via entanglement-assisted classical communication of a d-dimensional message can also be obtained via quantum communication carrying at most $\log d$ bits of information. That is, in this setting, quantum communication is the stronger resource.

Consider a scenario with classical communication, where Alice and Bob can use a pre-shared entangled state ρ_{AB}. Upon receiving input x, Alice performs a measurement $\{A_{a|x}\}$ with outcome a on her half of ρ_{AB}, which projects Bob's system onto the state $\sigma_{a|x} = \mathrm{tr}_{\mathrm{A}}([A_{a|x} \otimes \mathbb{1}_{\mathrm{B}}]\rho_{\mathrm{AB}})/p(a|x)$, where $p(a|x) = \mathrm{tr}([A_{a|x} \otimes \mathbb{1}_{\mathrm{B}}]\rho_{\mathrm{AB}})$. Alice then sends a classical message to Bob; which we represent as a collection of d-dimensional quantum states $\mu_{a|x}$ diagonal in the same basis. Thus, Bob holds the classical-quantum state $\mu_{a|x} \otimes \sigma_{a|x}$, on which he can perform some measurements in order to establish correlations $p(b|x, y)$. The information cost of this protocol originates only from the classical message, as the entanglement is pre-shared.

Now, we construct a quantum communication protocol to simulate the above correlations using at most $\log d$ bits of information. Upon receiving x, Alice samples from $p(a|x)$, and sends to Bob the classical-quantum state $\mu_{a|x} \otimes \sigma_{a|x}$. Evidently, Bob can now reproduce the same correlations $p(b|x, y)$. The key point is now to show that this protocol does not require more information than above. The ensemble (averaged over a) can be written $\mathcal{E}_{\mathrm{QC}} = \{p_X(x), \tau_x\}$ where $\tau_x = \sum_a p(a|x)\mu_{a|x} \otimes \sigma_{a|x}$. The corresponding guessing probability is

$$P_g^{\mathrm{QC}} = \max_{\{N_z\}} \sum_{a,x} p_X(x)p(a|x)\,\mathrm{tr}\left(\mu_{a|x} \otimes \sigma_{a|x} N_x\right) \tag{12}$$

where the POVM $\{N_z\}$ acts jointly on the classical message space and on the quantum state space. We can place the following upper bound on the guessing probability

$$P_g^{\mathrm{QC}} \leq \max_{\{N_z\}} \sum_x p_X(x)\,\mathrm{tr}\left(\left(\sum_a p(a|x)\sigma_{a|x}\right) N_x^{\mathrm{B}}\right), \tag{13}$$

where we have used that $\mathrm{tr}(\mu_{a|x} \otimes \sigma_{a|x} N_x) \leq \mathrm{tr}(\sigma_{a|x} N_x^{\mathrm{B}})$, where N_x^{B} is the partial trace of N_x over the first system (the classical message space). Importantly, since for every x the ensemble $\{p(a|x), \sigma_{a|x}\}$ is remotely prepared by Alice on Bob's side, it follows that

$$\begin{aligned}\sum_a p(a|x)\sigma_{a|x} &= \sum_a \mathrm{tr}_{\mathrm{A}}\left(A_{a|x} \otimes \mathbb{1}_{\mathrm{B}}\rho_{\mathrm{AB}}\right)\\ &= \mathrm{tr}_{\mathrm{A}}\left(\rho_{\mathrm{AB}}\right) = \rho_{\mathrm{B}}.\end{aligned} \tag{14}$$

Therefore, the guessing probability obeys

$$P_g^{\mathrm{QC}} \leq \max_{\{N_z\}} \sum_x p_X(x)\,\mathrm{tr}\left(N_x^{\mathrm{B}}\rho_{\mathrm{B}}\right) \tag{15}$$

$$\leq \left(\max_x p_X(x)\right) \max_{\{N_z\}} \mathrm{tr}\left(\sum_x N_x^{\mathrm{B}}\rho_{\mathrm{B}}\right). \tag{16}$$

Finally, we use the completeness relation of POVMs to obtain

$$\sum_x N_x^{\mathrm{B}} = \sum_x \mathrm{tr}_1\left(N_x\right) = \mathrm{tr}_1\left(\mathbb{1}_d \otimes \mathbb{1}\right) = d\mathbb{1}, \tag{17}$$

where we have used that the identity operator on the classical message space is d-dimensional. Thus, we conclude that

$$P_g^{\mathrm{QC}} \leq d \max_x p_X(x). \tag{18}$$

Consequently, the information is bounded by

$$\begin{aligned}\mathcal{I}_X &= -\log\left(\max_x p_X(x)\right) + \log\left(P_g^{\mathrm{QC}}\right)\\ &\leq -\log\left(\max_x p_X(x)\right) + \log\left(d \max_x p_X(x)\right) = \log d.\end{aligned} \tag{19}$$

This concludes the proof: quantum communication of $\log d$ bits of information is a stronger resource than classical communication of a d-dimensional message assisted by any amount of entanglement.

Finally, we also note that this proof remains valid also if Alice uses her classical outcome a and her input x to encode a quantum d-dimensional message $\mu_{a|x}$. This is, however, not the most general quantum operation that may be considered.

8 Conclusions

We have investigated correlations in prepare-and-measure scenarios under the assumption of an upper bound on the information. We have shown how to fully characterise correlations in the case of classical systems and proved a quantum advantage. Moreover, we showed that stronger quantum correlations can be obtained when bounding the information rather than the dimension, and devised device-independent tests of information.

An outstanding open question is to characterise quantum correlations when the transmitted information is bounded. Is it sufficient to consider quantum ensembles of finite dimension, as in the classical case? Or are there correlations that require infinite-dimensional quantum systems? Another point is to understand how much stronger quantum correlations can be compared to classical ones. For the case where shared randomness is not allowed, we could show a diverging advantage. Is it also the case in a scenario including shared randomness? In addition, it would be interesting to consider informationally restricted correlations based on other information measures than the one we consider here; for instance based on Shannon entropies. Another possible direction is to explore connections between our approach and other scenarios in information theory, for instance the (quantum) information bottleneck function [25, 26]. Furthermore, it would also be relevant to explore the role of informationally restricted correlations with respect to the line of research focused on operational contextuality [27] in which one considers prepare-and-measure scenarios featuring an assumption of oblivious communication (see e.g. [28, 29]).

Finally, we briefly discuss the prospects of using our approach in experiments, notably towards possible applications in semi-device-independent (SDI) quantum information processing. In this area, protocols so far were mostly based on a dimension assumption, see e.g. [8–12], which is usually justified from the physics of the experiment. For instance, a setup where the relevant degree of freedom is the polarization of a single photon motivates the assumption that the prepared states can be described as qubits. In practice, however, single-photon sources feature imperfections which result in unavoidable multi-photon emissions, which clearly no longer satisfy the qubit assumption. Taking these into account is typically cumbersome and inefficient (see for instance [11]). In comparison, the information approach might be much better adapted here. From a physical model of the source, the rate of multi-photon events can be estimated. For instance, a weak laser source will exhibit Poisson statistics. For each photon number the carried information can be estimated, which in turn results in an overall bound on the carried information. In this way, one could continuously tune the information bound, taking into account the relevant degrees of freedom and photon statistics. Bounding the information rather than the dimension may therefore represent a more natural assumption, better motivated by the physics of the source. It would be interesting to explore these ideas in practice, as well as to understand the relation between the information approach and other SDI approaches recently developed, based on bounding the energy [30], the overlap [31, 32] or the entropy [33] of the quantum communication.

Acknowledgements

We thank Jean-Daniel Bancal, Joseph Renes, Renato Renner, Joseph Bowles, Alastair Abbott, Francesco Buscemi, Michele Dall'Arno, Stefano Pironio and Erik Woodhead for discussions. This work was supported by the Swiss National Science Foundation (Starting grant DIAQ, NCCR-QSIT) and the Independent Research Fund Denmark.

References

[1] H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Nonlocality and communication complexity, Rev. Mod. Phys. **82**, 665 (2010).

[2] H. Buhrman, R. Cleve and A. Wigderson, Quantum vs. classical communication and computation, Proceedings of the 30th Annual ACM Symposium on Theory of Computin, 63 (1998).

[3] R. Raz, Exponential separation of quantum and classical communication complexity, In Proceedings of 31st ACM STOC, 358 (1999).

[4] R. Gallego, N. Brunner, C. Hadley, and A. Acín, Device-Independent Tests of Classical and Quantum Dimensions, Phys. Rev. Lett. **105**, 230501 (2010).

[5] J. Ahrens, P. Badziag, A. Cabello, and M. Bourennane, Experimental Device-independent Tests of Classical and Quantum Dimensions, Nature Physics **8**, 592 (2012).

[6] M. Hendrych, R. Gallego, M. Micŭda, N. Brunner, A. Acín, J. P. Torres, Experimental estimation of the dimension of classical and quantum systems, Nature Physics **8**, 588 (2012).

[7] M. Navascués, and T. Vértesi, Bounding the Set of Finite Dimensional Quantum Correlations, Phys. Rev. Lett. **115**, 020501 (2015).

[8] M. Pawłowski, and N. Brunner, Semi-device-independent security of one-way quantum key distribution, Phys. Rev. A **84**, 010302(R) (2011).

[9] H-W. Li, Z-Q. Yin, Y-C. Wu, X-B. Zou, S. Wang, W. Chen, G-C. Guo, and Z-F. Han, Semi-device-independent random-number expansion without entanglement, Phys. Rev. A **84**, 034301 (2011).

[10] E. Woodhead, S. Pironio, Secrecy in Prepare-and-Measure Clauser-Horne-Shimony-Holt Tests with a Qubit Bound, Phys. Rev. Lett. **115**, 150501 (2015).

[11] T. Lunghi, J. B. Brask, C. C. W. Lim, Q. Lavigne, J. Bowles, A. Martin, H. Zbinden, and N. Brunner, Self-Testing Quantum Random Number Generator, Phys. Rev. Lett. **114**, 150501 (2015).

[12] A. Tavakoli, J. Kaniewski, T. Vértesi, D. Rosset, and N. Brunner, Self-testing quantum states and measurements in the prepare-and-measure scenario, Phys. Rev. A **98**, 062307 (2018).

[13] N. Ciganović, N. J. Beaudry, and Renato Renner, Smooth Max-Information as One-Shot Generalization for Mutual Information, IEEE Transactions on Information Theory **60**, 1573 (2014).

[14] A. S. Holevo, Bounds for the quantity of information transmitted by a quantum communication channel, Problems of Information Transmission. **9**, 177 (1973).

[15] R. Jozsa, D. Robb, and W. K. Wootters, Lower bound for accessible information in quantum mechanics, Phys. Rev. A **49**, 668 (1994).

[16] Convex Optimization, S. Boyd and L. Vandenberghe, Cambridge University Press, 2004.

[17] A. Ambainis, A. Nayak, A. Ta-Shma, U. Vazirani, Dense quantum coding and a lower bound for 1-way quantum automata, Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC'99), 376-383 (1999).

[18] A. Ambainis, D. Leung, L. Mancinska, M. Ozols, Quantum Random Access Codes with Shared Randomness, arXiv:0810.2937.

[19] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane, Quantum random access codes using single d-Level systems, Phys. Rev. Lett. **114**, 170502 (2015).

[20] A. Tavakoli, M. Pawłowski, M. Żukowski, and M. Bourennane, Dimensional discontinuity in quantum communication complexity at dimension seven, Phys. Rev. A 95, 020302(R) (2017).

[21] A. Tavakoli, B. Marques, M. Pawłowski, and M. Bourennane, Spatial versus sequential correlations for random access coding, Phys. Rev. A **93**, 032336 (2016).

[22] A. Hameedi, D. Saha, P. Mironowicz, M. Pawłowski, and M. Bourennane, Complementarity between entanglement-assisted and quantum distributed random access code, Phys. Rev. A **95**, 052345 (2017).

[23] M. Pawłowski, and M Żukowski, Entanglement-assisted random access codes, Phys. Rev. A **81**, 042326 (2010).

[24] A. Tavakoli, and M. Zukowski, Higher-dimensional communication complexity problems: Classical protocols versus quantum ones based on Bell's theorem or prepare-transmit-measure schemes, Phys. Rev. A **95**, 042305 (2017).

[25] N. Tishby, F. C. Pereira and W. Bialek, The information bottleneck method, Proc. of the 37th Annual Allerton Conference on Communication, Control and Computing, pages 368-377, (1999)

[26] N. Datta, C. Hirche and A. Winter, Convexity and Operational Interpretation of the Quantum Information Bottleneck Function, Proc. ISIT 2019, 7-12 July 2019, Paris, pp. 1157-1161.

[27] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A **71**, 052108 (2005).

[28] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Prepara-

tion Contextuality Powers Parity-Oblivious Multiplexing, Phys. Rev. Lett. **102**, 010401 (2009).

[29] A. Hameedi, A. Tavakoli, B. Marques and M. Bourennane, Communication Games Reveal Preparation Contextuality, Phys. Rev. Lett. **119**, 220402 (2017).

[30] T. V. Himbeeck, E. Woodhead, N. J. Cerf, R. Garcia-Patron, and S. Pironio, Semi-device-independent framework based on natural physical assumptions, Quantum **1**, 33 (2017).

[31] J. B. Brask, A. Martin, W. Esposito, R. Houlmann, J. Bowles, H. Zbinden, and N. Brunner, Megahertz-Rate Semi-Device-Independent Quantum Random Number Generators Based on Unambiguous State Discrimination, Phys. Rev. Applied **7**, 054018 (2017).

[32] Y. Wang, I. W. Primaatmaja, E. Lavie, A. Varvitsiotis, C. C. W. Lim, Characterising the correlations of prepare-and-measure quantum networks, npj Quantum Information **5**, 17 (2019).

[33] R. Chaves, J. B. Brask, and N. Brunner, Device-Independent Tests of Entropy, Phys. Rev. Lett. **115**, 110501 (2015).

[34] M. Hayashi1, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita, (4,1)-Quantum random access coding does not exist - one qubit is not enough to recover one of four bits, New J. Phys. **8** 129 (2006).

[35] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, Optimal bounds for parity-oblivious random access codes, New J. Phys. **18** 045003 (2016).

A Characterisation of classical correlations

We describe a classical scheme, starting with deterministic strategies. Alice uses an encoding function $E : [n] \to [d]$ to associate her input to a d-valued message $m = E(x)$ and sends it to Bob. No limitation on d is assumed. Bob uses a decoding function $D : [d] \times [l] \to [k]$ to map the pair (m, y) into an k-valued output $b = D(m, y)$. Since there are $Z_{\mathrm{A}} = d^n$ ($Z_{\mathrm{B}} = k^{dl}$) possible encoding (decoding) functions, the number of deterministic strategies is $Z = Z_{\mathrm{A}} Z_{\mathrm{B}}$. We index them by $(E_{\lambda_{\mathrm{A}}}, D_{\lambda_{\mathrm{B}}})$ for $\lambda_{\mathrm{A}} \in [Z_{\mathrm{A}}]$ and $\lambda_{\mathrm{B}} \in [Z_{\mathrm{B}}]$ respectively. Via the shared randomness $\lambda = (\lambda_{\mathrm{A}}, \lambda_{\mathrm{B}})$, classical correlations are written

$$p^{\mathrm{C}}(b|x,y) = \sum_{\lambda} p(\lambda) \sum_{m=1}^{d} \delta_{m,E_{\lambda_{\mathrm{A}}}(x)} \delta_{b,D_{\lambda_{\mathrm{B}}}(m,y)}. \tag{20}$$

We now characterise $p^{\mathrm{C}}(b|x,y)$ when $\mathcal{I}_X \leq \alpha$ for some real $\alpha \geq 0$. To this end, we need to eliminate the dimension d. Below, in section A.3 we show that without loss of generality one can choose $d = n$ (i.e. the dimension equal to the number of inputs for Alice). We will use this fact to characterise the polytope of classical correlations and leave the proof for the end of this section.

A.1 The classical polytope

We use that classical messages of dimension $d = n$ are sufficient. Therefore, we can denote all encoding functions and decoding functions $(E_{\lambda_{\mathrm{A}}}, D_{\lambda_{\mathrm{B}}})$ where the index $\lambda = (\lambda_{\mathrm{A}}, \lambda_{\mathrm{B}})$ acts as a shared random variable (whose cardinality is now finite) allowing the coordination of deterministic encoding and decoding strategies. For a fixed deterministic strategy, we obtain a distribution $p'_\lambda(b|x,y)$. This distribution is a vertex of the polytope $\mathbb{P}$ which is the space of all probabilities $p(b|x,y)$. However, many deterministic strategies give rise to the same vertex in the probability space. Therefore, we write $\{p_\gamma(b|x,y)\}_\gamma$ for the unique elements in the set $\{p'_\lambda(b|x,y)\}_\lambda$. We define

$$E_\gamma = \{\lambda = (\lambda_{\mathrm{A}}, \lambda_{\mathrm{B}}) | p_\gamma(b|x,y) = p'_\lambda(b|x,y)\}, \tag{21}$$

where $\{p_\gamma(b|x,y)\}$ is the list of vertices of $\mathbb{P}$ (without duplicates). In other words, E_γ is the set of all deterministic strategies that generate the vertex $p_\gamma(b|x,y)$.

To each vertex of $\mathbb{P}$ we associate the smallest amount of information needed to generate it (for simplicity, we work with the guessing probability). That is,

$$P_g^{(\gamma)} = \min_{\lambda \in E_\gamma} P_g^{(\lambda_{\mathrm{A}})} \tag{22}$$

where the guessing probability of the deterministic strategy is given by

$$P_g^{(\lambda_A)} = \max_{\mu} \sum_x p_X(x) \sum_{m=1}^{d} \delta_{m,E_{\lambda_A}(x)} \delta_{x,\tilde{D}_\mu(m)}, \tag{23}$$

where the maximisation is over all the deterministic decoding strategies $\tilde{D} : [d] \to [n]$ (of which there are n^d).

We now impose the information restriction, $I_X \leq \alpha$. This can be formulated as a linear constraint in the shared randomness. The characterisation of the set of information restricted classical correlations reads

$$p(b|x,y) = \sum_\gamma p(\gamma) p_\gamma(b|x,y) \tag{24}$$

$$\sum_\gamma p(\gamma) P_g^{(\gamma)} \leq 2^{\alpha - H_{\min}(X)} \tag{25}$$

$$\sum_\lambda p(\gamma) = 1 \tag{26}$$

$$p(\gamma) \geq 0. \tag{27}$$

This defines a convex polytope. Its facets can be obtained using standard polytope software. We label this polytope $\mathbb{P}_\alpha$ and note that it is contained inside $\mathbb{P}$.

As an illustration of how the polytope $\mathbb{P}_\alpha$ may look, we have displayed in Fig. 3 a schematic of the polytope in the simplest case of Alice having two inputs and Bob performing a single binary outcome measurement ($n = k = 2$, $l = 1$), for which the polytopes $\mathbb{P}$ and $\mathbb{P}_\alpha$ are polygons.

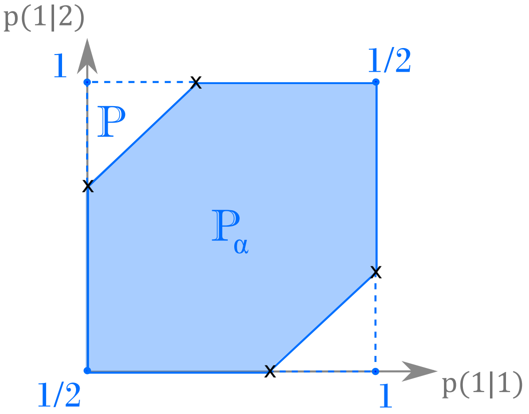

Figure 3: The classical set of correlations for a scenario with two preparations and one binary outcome measurement $(n, l, k) = (2, 1, 2)$. The polytope $\mathbb{P}$ has four vertices, each corresponding to a guessing probability of either one or one half (written in blue). The facets are lines. Therefore there is only one pair of vertices per facet, for each of which we inscribe a new vertex (represented by a tick) as imposed by limiting the guessing probability. Thus, the blue region is the polytope $\mathbb{P}_\alpha$.

A.2 Optimal classical correlations via linear programming

Since the set of classical correlations forms a convex polytope for $\mathcal{I}_X \leq \alpha$, one can determine whether a given $p(b|x,y)$ belongs to said polytope via a linear program. This allows one to determine whether $p(b|x,y)$ is classically realisable with information no more than α.

Moreover, given any linear functional of probabilities,

$$F = \sum_{x,y,b} r_{xyb}\, p(b|x,y), \tag{28}$$

one can determine the exact classical bound through the evaluation of the linear program

$$\begin{aligned} F^{\mathrm{C}} &= \max_{p(\lambda)} F[p(b|x,y)] \\ &\text{such that } \sum_\lambda p(\lambda) P_g^{(\lambda_A)} \leq 2^{\alpha - H_{\min}(X)}, \\ &\sum_\lambda p(\lambda) = 1, \quad \text{and} \quad p(\lambda) \geq 0. \end{aligned} \tag{29}$$

This allows to obtain witnesses for classical correlations.

A.3 Dimension n is sufficient for classical messages

Here, we show that the optimal classical correlations, for any correlation witness constrained by bounded guessing probability (or equivalently, bounded information) with shared randomness, is obtained with a message dimension not larger than the cardinality of the input of Alice, i.e. $d = n$.

Any classical strategy can be decomposed as a mixture of deterministic strategies, as given by Eq. (20). For a fixed value of the shared variable λ, the encoding strategy E_{λ_A} is fixed. Since x can take at most n different values, there is then at most n different values of $E_{\lambda_A}(x)$. Thus, for a fixed λ, at most n message symbols are used. Whether there is any advantage in using message dimensions $d > n$ thus becomes a question of whether there is any advantage in using different sets of message symbols for different λ.

We first show that any value of the maximum in Eq. (29) obtained with different sets of message symbols for different λ can also be achieved using the same set of n symbols for all λ. This can be seen from Eq. (20). For each value of λ, the factor $\delta_{m,E_{\lambda_A}(x)}$ is nonzero for at most n different values of m. The decoding function $D_{\lambda_{\mathrm{B}}}(m)$ hence needs to be defined only on these values. If any of these values lie outside $[n] = \{1, \dots, n\}$ then there must be corresponding values in $[n]$ which are not used. We can then redefine $E_{\lambda_{\mathrm{A}}}$ and $D_{\lambda_{\mathrm{B}}}$ to use these values instead.

Specifically, for some fixed λ, say that $E_{\lambda_{\mathrm{A}}}(x_0) = \nu \notin [n]$ for some x_0. Then there exists $\nu' \in [n]$ such that $E_{\lambda_{\mathrm{A}}}(x) \neq \nu'$ for all x. We then define

$$E'_{\lambda_{\mathrm{A}}}(x) = \begin{cases} \nu' & \text{if } x = x_0, \\ E_{\lambda_{\mathrm{A}}}(x) & \text{otherwise,} \end{cases} \tag{30}$$

$$D'_{\lambda_{\mathrm{B}}}(m) = \begin{cases} D_{\lambda_{\mathrm{B}}}(\nu) & \text{if } m = \nu', \\ D_{\lambda_{\mathrm{B}}}(m) & \text{otherwise.} \end{cases} \tag{31}$$

Substituting $E_{\lambda_{\mathrm{A}}} \to E'_{\lambda_{\mathrm{A}}}$ and $D_{\lambda_{\mathrm{B}}} \to D'_{\lambda_{\mathrm{B}}}$ in (20) leaves the probabilities $p(b|x,y)$ unchanged. Repeating this process, the message symbols can be chosen in $[n]$ for every λ, without changing the probabilities and hence a distribution achieving the optimum in Eq. (29) remains optimal.

The only remaining question is now, whether this remapping to a strategy using the same n symbols for all λ can lead to violation of the information constraint. From (23), we can see that this is not the case. Let $\tilde{D}_{\mu^*}$ be the optimal decoding function which achieves the maximum on the right-hand side of (23), for some fixed λ. When $E_{\lambda_{\mathrm{A}}}$ is replaced by $E'_{\lambda_{\mathrm{A}}}$ as above, the maximum remains unchanged and is achieved by

$$\tilde{D}'_{\mu^*}(m) = \begin{cases} \tilde{D}_{\mu^*}(\nu) & \text{if } m = \nu', \\ \tilde{D}_{\mu^*}(m) & \text{otherwise.} \end{cases} \tag{32}$$

Thus, following the recipe above, we can replace all the encoding and decoding functions $E_{\lambda_{\mathrm{A}}} : [n] \to [d]$, $D_{\lambda_{\mathrm{B}}} : [d] \to [n]$, and $\tilde{D}_{\mu} : [d] \to [n]$ by other functions $E_{\lambda_{\mathrm{A}}} : [n] \to [n]$, $D_{\lambda_{\mathrm{B}}} : [n] \to [n]$, and $\tilde{D}_{\mu} : [n] \to [n]$ without changing the probabilities $p(b|x,y)$ or the guessing probabilities $P_g^{(\lambda_{\mathrm{A}})}$. It follows that the optimum of Eq. (29) can always be attained using a message dimension of at most n.

B Case study for $(n, l, k) = (3, 2, 2)$

We have obtained the facets of the polytope for several simple scenarios. The simplest scenario in which we have found non-trivial facets is $(n, l, k) = (3, 2, 2)$. One can consider different values for the information bound $\mathcal{I}_X \leq \alpha$. We have considered different values of α for each of which we have found two non-trivial inequalities (i.e. they are not positivity nor the information restriction). More precisely, we considered eleven evenly spaced values of the guessing probability in the range $(1/3, 1)$. The facets are

$$F_1 = \sum_{x,y} t^1_{x,y} E(x,y) \leq 6P_g - 1 \tag{33}$$

$$F_2 = \sum_{x,y} t^2_{x,y} E(x,y) \leq 12P_g - 4. \tag{34}$$

where $t^1_{x,y} = \{[-1,-1],[-1,1],[1,0]\}$ and $t^2_{x,y} = \{[-1,-1],[-1,1],[2,0]\}$. Note that for convenience, we have expressed the upper bounds in terms of the guessing probability instead of the information. Both inequalities can be violated in quantum theory. For the first inequality, a violation valid for any non-trivial information was presented in the main text using a quantum strategy with one bit of shared randomness. Notably, said strategy also violates the second inequality but not in the entire range $\mathcal{I}_X \in (0, \log 3)$.

Moreover, we have numerically explored whether larger violations of the first inequality are possible. We considered the case in which Alice prepares general qutrit states and found it to be advantageous. We have employed a brute-force numerical search using the function “fmincon“ in MATLAB. We employ an effective Lagrange multiplier λ and seek to maximise the function

$$\tilde{F}_1 = F_1 - \lambda|\mathcal{I}_X - \alpha|, \tag{35}$$

for a given information bound α. We have chosen $\lambda = 100$. In every step, we evaluate the information $\mathcal{I}_X$ in the three preparations via a semidefinite program. Then, we evaluate the largest possible value of F_1 for the given preparations, which thanks to the binary outcomes can be cast as an eigenvalue problem. We then ask MATLAB to maximise $\tilde{F}_1$. In Fig 4 the results are compared to those of the strategy in the main text. In the range $\mathcal{I}_X \in (0, 1)$ we find an improvement, but not in the range $\mathcal{I}_X \in [1, \log 3]$. However, we have not found a simple parameterisation of these quantum strategies. Also, it could be possible that even better results can be obtained with higher-dimensional preparations.

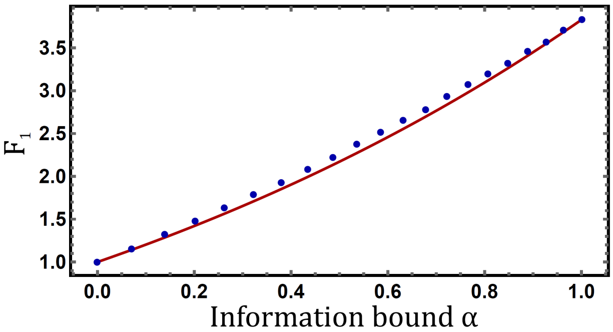

Figure 4: Witness value F_1 as a function of the information $\mathcal{I}_X \leq \alpha$. The quantum strategy from the main text is displayed (red curve) and the numerically obtained quantum violations based on qutrits are displayed in blue. In the range $\alpha \in (0,1)$ these improve on the first quantum strategy. Notably, numerics showed that an improvement on the red curve is possible already with qubit preparations.

C Arbitrary large advantage over dimension-bounded quantum ensembles without shared randomness

In the main text, we showed that one bit of communication is not always optimally encoded in a qubit ensemble but sometimes in an ensemble of higher-dimensional quantum systems. Here, we show that such advantages over dimension-bounded systems can become more significant in scenarios without shared randomness.

Consider the following variant of a quantum Random Access Code (without shared randomness). Alice has a uniformly random variable $X \in [2^n]$ with values $x = x_1 \dots x_n \in [2]^n$. She sends m bits of information to Bob, who has a random variable $Y \in [n]$ with values y from which he produces an outcome $b \in [2]$. The aim is to maximise the *worst-case* success probability of finding $b = x_y$, i.e.,

$$\mathcal{A}_n^m = \min_{x,y} p(b = x_y|x,y). \tag{36}$$

Let us first choose $m = 1$. It is known that with two-valued classical messages or with two-dimensional quantum systems, it is impossible to achieve a better result than that obtained with random guessing, i.e. $\mathcal{A}_n^1 = 1/2$, when $n > 3$ [34]. In contrast, for $n = 2$ and $n = 3$, qubits hold an advantage over classical two-valued messages. The reason is that for $n = 4$ (and analogously for $n > 4$) it is impossible to cut the Bloch sphere into $2^4 = 16$ symmetric parts with four planes passing through the origin. By a similar argument using the generalised higher-dimensional Bloch sphere, it has been shown [34] that for general integers $m \geq 1$, sending m classical two-valued messages or sending m qubits (2^m-dimensional quantum systems) cannot achieve a better result than $\mathcal{A} = 1/2$ when n is choosen as at least 2^{2m}.

We compare this with sending a general quantum ensemble of limited information. Again, we first choose $m = 1$ and $n = 4$. Using the ensemble and measurements specified in the main text for four-bit Random Access Code (average success probability variant), one immediately finds that $\forall x, y : p(b = x_y|x,y) = 3/4$, and therefore that $\mathcal{A}_4^1 = 3/4$. Thus, the ensemble of mixed four-dimensional systems provides an advantage over two-valued classical messages when qubit ensembles fail to provide any better-than-classical result.

Refs. [21, 35] derived Bell inequalities for Random Access Codes. Using the results of Ref. [35], Alice and Bob can share an entangled state of local dimension $D = 2^{\lfloor \frac{n}{2} \rfloor}$ and use their inputs as settings for testing the Bell inequalities of [21, 35]. Then, if Alice communicates her binary outcome to Bob, he can satisfy the relation $b = x_y$ with probability

$$\forall x, y : \qquad p(b = x_y|x,y) = \frac{1}{2} + \frac{1}{2\sqrt{n}}. \tag{37}$$

In the main text we showed that any correlations achievable by means of entanglement-assisted classical communication also is achievable by means of quantum communication without sending more information (and without the need of share randomness). Therefore, we can obtain the correlations (37) using the quantum communication model discussed in the main text. Consequently, using only a single bit of quantum information (encoded in a general ensemble), we can achieve

$$\mathcal{A}_n^1 = \frac{1}{2} + \frac{1}{2\sqrt{n}}. \tag{38}$$

Note that this is strictly greater than 1/2 for all $n \geq 2$. Therefore, if we choose $n \geq 2^{2m}$ but use only a single bit of information, we outperform the best possible quantum protocols in which the allowed m bits are encoded in 2^m-dimensional quantum systems. Thus, the advantage is unbounded in the sense that a fixed amount (one bit) of general quantum information holds an advantage over the m bits carried by m qubits, for any (potentially) arbitrarily large choice of m.