Quantum and superquantum enhancements to two-sender, two-receiver channels

Yihui Quek and Peter W. Shor

Phys. Rev. A **95**, 052329 — Published 15 May 2017

DOI: 10.1103/PhysRevA.95.052329

Quantum and super-quantum enhancements to two-sender, two-receiver channels

Yihui Quek[1] and Peter W. Shor[1,*]

[1] *Massachusetts Institute of Technology, Departments of Physics and Mathematics*

(Dated: March 13, 2017)

We study the consequences of 'super-quantum non-local correlations' as represented by the PR-box model of Popescu and Röhrlich, and show PR-boxes can enhance the capacity of noisy interference channels between two senders and two receivers. PR-box correlations violate Bell/CHSH inequalities and are thus stronger – more non-local – than quantum mechanics; yet weak enough to respect special relativity in prohibiting faster-than-light communication. Understanding their power will yield insight into the non-locality of quantum mechanics. We exhibit two proof-of-concept channels: first, we show a channel between two sender-receiver pairs where the senders are not allowed to communicate, for which a shared super-quantum bit (a PR-box) allows perfect communication. This feat is not achievable with the best classical (senders share no resources) or quantum entanglement-assisted (senders share entanglement) strategies. Second, we demonstrate a class of channels for which a tunable parameter ϵ achieves a *double* separation of capacities; for some range of ϵ, the super-quantum assisted strategy does better than the entanglement-assisted strategy, which in turn does better than the classical one.

I. INTRODUCTION

Bell's influential paper in 1964 [1] brought to light the existence of correlations that can be obtained from bipartite measurements of a quantum state, that cannot be reproduced by a local theory. Quantum mechanics is a *non-local theory* because it is able to predict such correlations, whereas a local theory with spatially separated observers could never do so. Such a local theory would *prohibit* physical measurements (of, say, particle A's spin) in one place from affecting the measurement outcomes of another experimenter (who measures, say, particle B's spin) who is spacelike-separated from the first one, if there is no field between them. Whereas in a non-local theory, to borrow an analogy from Popescu [2], 'moving something here, something else instantaneously wiggles there'.

Research into this area (see [3] for a review) has been motivated by the desire to understand *how* the non-locality of quantum theory gives rise to the advantages of information processing with quantum resources. One of the main results in this research is the famous inequality of Clause, Horne, Shimony and Holt [4], which bounds the statistics of spatially-separated measurements by two experimenters on a physical state in local hidden-variable (LHV) models. They define a quantity

$$S := |\langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle| \quad (1)$$

where A_0 and A_1 are local measurement operators corresponding to spin up and spin down on experimenter A's spin-half particle, and B_0 and B_1 the analogous measurement operators for Bob, and $\langle . \rangle$ denotes expectation value, and show that for LHV models,

$$S_{\text{LHV}} \leq 2 \quad (2)$$

Since LHV theories must obey the inequality (2), while quantum theories, which are non-local, need not, the quantity in (2) has become a popular metric of the non-locality of a given theory and is sometimes referred to as the 'CHSH value'. *Quantum mechanics*(QM), as a non-local theory, is exempt from this bound. Measurements on an entangled state, such as the state $\frac{|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B}{\sqrt{2}}$, can satisfy

$$S_{\text{QM}} \leq 2\sqrt{2}. \quad (3)$$

Tsirelson[5] proved that with QM, $2\sqrt{2}$ is the maximal achievable violation of this inequality. But QM must also respect the causality/non-signaling property of special relativity, which prohibits information transfer at a speed faster than light. In fact, out of all our physical theories that are currently in use, QM is special in being non-local and *yet* satisfying the non-signaling constraint: two spacelike-separated observers may influence each other *(non-locality)*, and yet, cannot communicate with each other – the above-mentioned 'influence' must not allow for information transfer *(relativistic causality)*.

But note that even S_{QM} falls short of its algebraic maximum (see Equation (1)), which is 4. In 1994, Popescu and Rohrlich[6], asking 'Why isn't quantum theory *more* non-local?', proved that it is possible to construct causality-satisfying models that are more non-local than QM. To unify these theories, they proposed an abstraction to represent the probability distribution that they induce on measurement outcomes: a *non-local box*, visualized in figure 1. This is a bipartite correlated box with two ends, one of which is held by Alice and the other by Bob. Alice inputs x (respectively Bob inputs y) and

* shor@math.mit.edu

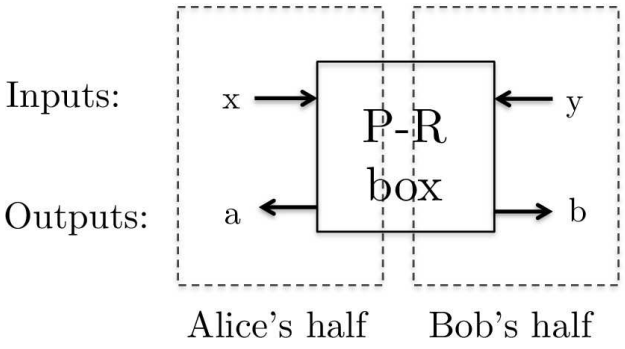

FIG. 1: A PR, or non-local box, whose inputs and outputs are governed by the distribution in Equation (4).

the box outputs a (respectively b) according to the probability distribution $P(a,b|x,y)$ (where $x,y,a,b \in \{0,1\}$):

$$P^{PR}(a,b|x,y) \begin{cases} 1/2 & \text{if } a \oplus b = xy \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

To calculate the CHSH value of the PR box, we now interpret A_0, A_1 (respectively B_0, B_1) from Equation (1) as the expected value of the box's output when Alice (respectively Bob) puts in 0, 1 into her end of the PR-box. This information-theoretic formulation of Alice and Bob's interaction with the theory is completely analogous with our previous language of measurements when construed within the measurement-operator formalism: in making measurements of a two-level system, Alice and Bob apply a set of measurement operators $\{\Pi_0, \Pi_1\}$ corresponding to the two possible outcomes, which correspond exactly to the set of inputs $\{0,1\}$ of both experimenters to the PR-box.

Thus, with such a PR box, we achieve the following super-quantum correlations:

$$S_{\text{QM}} = 4, \tag{5}$$

and we call a theory that predicts the correlations of PR-boxes a *super-quantum* theory. This is one that produces even stronger nonlocal correlations than quantum theory.

An important implication of this is that **if they share a PR-box, Alice and Bob could always win the CHSH game**. This is because the condition for outputs a, b to be produced by the PR box is exactly the winning condition of the CHSH game. On the other hand, if Alice and Bob share an entangled pair, they could win the CHSH game with a probability of at most $\cos^2(\frac{\pi}{8})$. This illustrates the super-quantum nature of the PR-box.

We summarize the theories under consideration in terms of their locality properties (as measured by their CHSH value) in Figure 2. We also refer the reader to [7] for a comprehensive review of PR-boxes and non-local correlations.

A major push of quantum information research has been to devise strategies that utilize quantum properties, such as entanglement, to aid communication tasks – quantum key distribution, quantum bit commitment and so on. This begs the question: could we use the maximally non-local correlations of *super-quantum* theories as a resource, and what tasks would they facilitate?

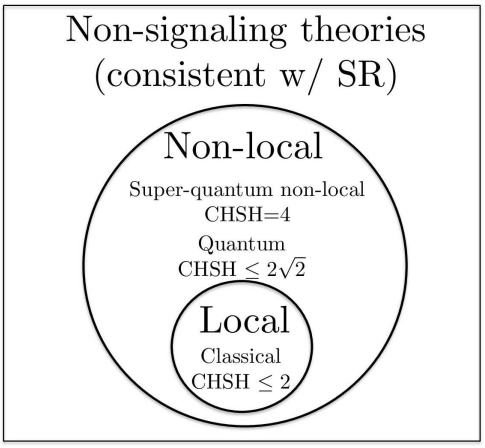

FIG. 2: Types of theories grouped by their locality properties (they must all not permit space-like separated observers to communicate and hence all fall under the banner of non-signaling)

Previously, PR-boxes had been shown to allow Alice and Bob to perform any two-party distributed computation by transmitting only a single bit of information [8], as well as the cryptographic primitives of unconditionally-secure bit-commitment and oblivious transfer [9]. This paper is the first survey of how super-quantum assistance could enhance communication over an interference channel.

It is organized as follows: In Section II, we introduce notation for the two-sender, two-receiver interference channel, as well as the information quantity we optimize. In Section III, we present our original result of a two-sender, two-receiver interference channel over which communication is more efficient with the aid of a PR-box, than with entanglement and/or a classical strategy. In Section IV, we present a variant of the above; a class of erasure channels characterized by a tunable parameter ϵ, whose capacities show a strict separation given these three classes of resources (classical, quantum-assisted and PR-box assisted). We finally conclude with a summary of results and suggestions for future research in Section V.

II. NOTATION

In the following sections, we will exhibit several two-sender, two-receiver channels that demonstrate capacity separations. We use the Shannon model of channel communication [10] to describe these channels, for which we follow the notation of [11] (in turn based on [12]). The basic model of a two sender-receiver pair channel is depicted in figure 3.

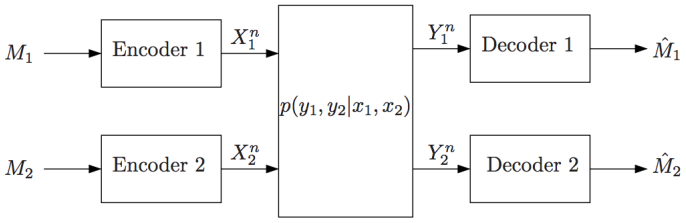

FIG. 3: General model of a two sender-receiver pair communication system. Figure taken from [11].

Such a channel is denoted $(\mathcal{X}_1 \times \mathcal{X}_2, p(y_1, y_2|x_1, x_2), \mathcal{Y}_1 \times \mathcal{Y}_2)$. A $(2^{nR_1}, 2^{nR_2}, n)$ code for this channel consists of:

- Two message sets $[1 : 2^{nR_1}]$ and $[1 : 2^{nR_2}]$
- Two encoders, where encoder 1 assigns a codeword $x_1^n(m_1)$ to each message $m_1 \in [1 : 2^{nR_1}]$ (respectively encoder 2 assigns $x_2^n(m_2)$ for $m_2 \in [1 : 2^{nR_2}]$).
- Two decoders, where decoder 1 uses a *decoding rule* to assign an estimate $\hat{m_1}$ or an error message e to each received sequence y_1^n, and decoder 2 does the same (ie. assigns $\hat{m_2}$ or e).

A rate pair (R_1, R_2) is said to be achievable for this channel if there exist a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ codes such that $\lim_{n\to\infty}[P_e^{(n)} \equiv P\{(\hat{M_1}, \hat{M_2}) \neq (M_1, M_2)\}] = 0$. For our channels, we will be concerned with their sum-capacity C_{sum} for classical information, which is the maximum, over all coding strategies, of the sum of the rates for each sender-receiver pair. That is,

$$C_{\text{sum}} = \max_{\text{coding strategy}} (R_1 + R_2) \tag{6}$$

Note that $C_{\text{sum}} \neq \max_{p(x_1)} R_1 + \max_{p(x_2)} R_2$ in general, because the rates must be attainable *simultaneously*. Whenever we speak of the 'capacity' of a channel, we shall refer to this information capacity.

III. CHANNEL I

In 2005, Cerf, Gisin, Massar and Popescu demonstrated a sense in which super-quantum non-locality encompasses quantum non-locality – they showed that a PR box could simulate the correlations obtained from any bipartite measurement of a maximally entangled pair of qubits without communication[13]. The reverse direction of simulation is impossible because PR-box correlations are more non-local than entanglement. Therefore, one expects that any communication task which is made more efficient with the aid of entanglement, could potentially benefit *even more* from PR-boxes. Table I represents a channel that demonstrates just such a non-locality separation.

In what follows, 'classical' strategies are those where senders are allowed to share no communication but may discuss a strategy before-hand, and 'entanglement-assisted' (alternatively, 'quantum-assisted) strategies imply strategies where the senders are allowed to share $2 \times n$ quantum entanglement – that is, a bipartite quantum state where each half is an n-level system represented as a n-dimensional Hilbert space. We make the usual assumption that the senders are spacelike-separated from each other and from any sources that they consult. This rules out any classical simulations of the quantum- and super-quantum strategies we mention.

$X_1 \backslash X_2$	**00**	**01**	**10**	**11**
00	00	11	01	10
01	11	00	10	01
10	10	01	00	11
11	01	10	11	00

TABLE I: Channel I: The senders each send two-bit codewords, X_1 and X_2 (codeword choices are in bold, on the axes), and the two-bit entries in the table (Y_1Y_2) correspond to the channel outputs; one bit goes to each receiver. Thus, if Sender 1 sends 01 and Sender 2 sends 10, Receiver 1 gets 1 and Receiver 2 gets 0.

The notation for this channel (which we shall call Channel I) is as follows: the senders and receivers shall be denoted by A_i and B_i; the bits they handle shall be denoted X_i (2 bit message that A_i inputs to the channel) and Y_i (1 bit message that B_i receives from the channel). To prevent confusion, we will try not to use A/B simultaneously with X/Y, unless it is necessary to make such a distinction.

On each use of Channel I, the senders send **two bits** out of the alphabet $\{00, 01, 10, 11\}$ and the channel outputs **one bit** to each receiver. Table I shows the output pairs that correspond to each input pair.

By definition, the maximum possible sum-capacity of Channel I (over all classes of resources) is 2: the two receivers each receive one bit. In fact, $C_{\text{sum}} = 2$ only if there exists a strategy where the receiver always decodes the sender's bit perfectly. In fact, it will turn out we fall far short of this maximum if the senders are restricted to using a purely classical probabilistic strategy; in that case the capacity is 1. We now show that this channel

demonstrates the capacity separations

$$\mathbf{C_{classical}},\ \mathbf{C_{quantum}} < \mathbf{C_{super\text{-}quantum}}.$$

A. Capacity of Channel I with a classical strategy

Let us build up our intuition about Channel I to understand why the classical strategy capacity should be so small. Channel I takes *two*-bit inputs but outputs only one bit to each receiver, so if the senders can ultimately communicate only one bit, the second bit seems redundant. Might the redundancy improve communication? We could note the following:

- Consider a strategy where each sender sends codewords according to a uniform probability distribution over the entire input alphabet, for both senders. Taking the marginal probability distribution for the first pair (over the second pair) results in the binary symmetric channel of Table II. It is the same for the other pair. This channel has a bit-flip probability $p = 0.5$. Since the capacity of the binary symmetric channel is $1 - H(p)$, the best possible joint rate with this strategy is 0.

$X_1 \backslash Y_1$	**0**	**1**
00	Pr = 0.5	Pr = 0.5
01	0.5	0.5
10	0.5	0.5
11	0.5	0.5

TABLE II: A uniform probability distribution results in a perfectly randomizing channel, evident from taking the marginal probability distributions for one sender-receiver pair (in this case, the first).

- The following coding strategy gives a joint rate of 1, and therefore 1 is an inner bound on the sum capacity: A_2 always sends 00 while A_1 encodes message bit 0 as 00 and message bit 1 as 01; then B_1 receives exactly the bit that A_1 intended to send. So the first sender-pair always communicates perfectly at the expense of the second pair.

The reader should persuade herself that other simple strategies such as reducing the size of either sender's alphabet will not achieve perfect coding either. In fact, as Lemma 1 shows, it is not even possible to do *better* than $R_1 + R_2 = 1$.

Lemma 1 (Classical capacity of Channel I)**.** *If the senders are limited to a classical (at most probabilistic) strategy with no aid from communication, entanglement or PR boxes, on the given channel the sum-capacity is strictly outer-bounded:*

$$R_1 + R_2 < 2. \tag{7}$$

In fact, we may show computationally that

$$C_{classical} = 1. \tag{8}$$

Proof. Here we sketch a proof for Equation (7). We show that $R_1 := I(X_1 : Y_1) = 1$ implies $R_2 := I(X_2 : Y_2) < 1$.

Suppose $I(X_1 : Y_1) = 1$. Using the chain rule for mutual information shows that $I(X_2 : Y_1|X_1) = 0$.

$$I(X_1 : Y_1) = 1 = \underbrace{I(X_1, X_2 : Y_1)}_{\text{takes on maximal value, 1}} - \underbrace{I(X_2 : Y_1|X_1)}_{=0} \tag{9}$$

Information-theoretically, the condition $I(X_2 : Y_1|X_1) = 0$ means that the first receiver's bit, Y_1, cannot possibly distinguish between the possibilities for the second sender's 2-bit message, X_2, for every choice of X_1 – and this is a restriction on what the the second sender's alphabet set could be. Consider the first row of Table I. The restriction says that **if Sender 1 sends** 00 **on a particular channel use, then there are only two possible non-trivial choices for Sender 2's alphabet**: a uniform probability distribution over $\{00, 10\}$ (both resulting in the output $Y_1 = 0$), OR a uniform probability distribution over $\{01, 11\}$ (both resulting in the output $Y_1 = 1$).

We similarly analyze the cases when Sender 1 sends 01, 10 or 11. The conclusion is that one of the following must hold (otherwise the restriction is never met):

1. Sender 1's alphabet is some subset of $\{00, 01\}$; Sender 2's alphabet is either $\{00, 10\}$ or $\{01, 11\}$.

2. Sender 1's alphabet is some subset of $\{10, 11\}$; Sender 2's alphabet is either $\{00, 11\}$ or $\{10, 01\}$.

Since the two senders are not allowed to communicate during the sending of the messages, they must choose an alphabet at the start and stick to it. Consequently, only one of these four cases can hold, and bearing in mind the other restriction that our coding strategy must fulfil the condition $I(X_1 : Y_1) = 1$, we may show that $R_2 < 1$ for all of them. For an example of this analysis, refer to Appendix A. □

But it is still possible that if one of the sender-receiver pairs is willing to accept a sub-optimal (less than 1) rate, the other pair could attain a high rate such that $R_1 + R_2 > 1$. To show that this never happens, we ran an algorithm based on modified gradient descent. This algorithm is given in pseudocode here (Algorithm 1). The inputs to the algorithm are two vectors $\vec{x_1} := (a_1, b_1, c_1, d_1), \vec{x_2} := (a_2, b_2, c_2, d_2)$, such that the square of the entries in the first vector $\{a_1^2, b_1^2, c_1^2, d_1^2\}$ represents the probabilities of Sender 1 sending $\{00, 01, 10, 11\}$ respectively, and correspondingly $\{a_2^2, b_2^2, c_2^2, d_2^2\}$ for Sender 2. The modification to the usual gradient descent algorithm was to respect the constraints

$$a_1^2 + b_1^2 + c_1^2 + d_1^2 = 1\ ;\ a_2^2 + b_2^2 + c_2^2 + d_2^2 = 1.$$

To do this, we treated the problem of simultaneous gradient descent where the component vectors had to lie on two 4-D unit spheres. After running gradient descent 10000 times with a *tol* set to $1e-6$ and never observing a value of the joint rate above 1, we concluded that the joint rate is, indeed, upper bounded by 1. Equation (8) follows.

Modified-Gradient-Descent($\vec{x}$)

ALGORITHM 1: Finds the maximum value of the function $I(X_1:Y_1)+I(X_2:Y_2)$ over all input distributions

$f(\vec{x_1},\vec{x_2}) := -I(X_1;Y_1) - I(X_2;Y_2)$ (Objective function)
$\vec{g_1} := \vec{\nabla}_{x_1} f$; $\vec{g_2} := \vec{\nabla}_{x_2} f$
Initialize $x_1, x_2, tol, maxiter$
while $iter < maxiter$ and $dx > tol$ **do**
  Evaluate $\vec{g_1}(\vec{x_1},\vec{x_2})$; $\vec{g_2}(\vec{x_1},\vec{x_2})$
  $\vec{h_1} \leftarrow \vec{g_1} - (\vec{g_1}\cdot\vec{x_1})\vec{x_1}$; $\vec{h_2} \leftarrow \vec{g_2} - (\vec{g_2}\cdot\vec{x_2})\vec{x_2}$
  $\alpha_1 \leftarrow \frac{h_1^2}{h_1^2+h_2^2}$; $\alpha_2 \leftarrow \frac{h_2^2}{h_1^2+h_2^2}$
  $\vec{n_1} \leftarrow \frac{\vec{h_1}}{|\vec{h_1}|}$; $\vec{n_2} \leftarrow \frac{\vec{h_2}}{|\vec{h_2}|}$
  $\phi' \leftarrow \arg\min_{\phi} f(\cos(\alpha_1\phi)\vec{x_1} + \sin(\alpha_1\phi)\vec{n_1}, \cos(\alpha_2\phi)\vec{x_2} + \sin(\alpha_2\phi)\vec{n_2})$
  $\vec{x_1'} \leftarrow \cos(\alpha_1\phi')\vec{x_1} + \sin(\alpha_1\phi')\vec{n_1}$; $\vec{x_2'} \leftarrow \cos(\alpha_2\phi')\vec{x_2} + \sin(\alpha_2\phi')\vec{n_2}$
  $dx \leftarrow \sqrt{x_1'^2 - x_1^2 + x_2'^2 - x_2^2}$
  $iter += 1$
end while

B. Capacity of Channel I with super-quantum assistance

We introduce the notion of super-quantum assisted capacity with a thought experiment: supposing that the two senders may coordinate their input alphabets in real-time, perhaps by using a non-classical resource. If we want both pairs to communicate perfectly, that is $I(X_1:Y_1)=1$ AND $I(X_2:Y_2)=1$, this imposes 4 conditions on the actual encodings that go into the channel:

1. If $X_1 \in \{00,01\}$, either $X_2 \in \{00,10\}$ or $X_2 \in \{01,11\}$.
2. If $X_1 \in \{10,11\}$, either $X_2 \in \{01,10\}$ or $X_2 \in \{00,11\}$.
3. If $X_2 \in \{00,01\}$, either $X_1 \in \{00,10\}$ or $X_1 \in \{01,11\}$.
4. If $X_2 \in \{10,11\}$, either $X_1 \in \{01,10\}$ or $X_1 \in \{00,11\}$.

That, is, only the shaded outputs in either the left or the right subtables of Table III could be produced. Obviously, this is not a set that can be produced with only classical resources. Lemma 2 states that it is possible with a PR-box.

$X_1 \backslash X_2$	**00**	**01**	**10**	**11**
00	00	11	01	10
01	11	00	10	01
10	10	01	00	11
11	01	10	11	00

$X_1 \backslash X_2$	**00**	**01**	**10**	**11**
01	00	11	01	10
00	11	00	10	01
11	10	01	00	11
10	01	10	11	00

TABLE III: Hypothetically, the demand that perfect coding happen requires that only the shaded outputs be produced by the channel. Only these two coding strategies will allow both $I(X_1:Y_1)=1$ and $I(X_2:Y_2)=1$. Returning back to the classical realm, since the senders cannot communicate with each other, they cannot coordinate their inputs so as to only produce the shaded outputs, so perfect coding is not possible classically. But if they share a PR-box, they can. Our super-quantum strategy achieves exactly the left-hand-side set of outputs.

Lemma 2 (Capacity of Channel I with super-quantum resources)**.** *If the senders are allowed to share a PR-box, the capacity of the given channel is exactly 2. This is the algebraically maximal sum-capacity of the channel.*

We have all but spelled out our super-quantum strategy. In this strategy, the senders can communicate only one bit m. They encode this bit into a two-bit codeword by concatenating it with the single-bit output of the PR box which results from feeding m into their respective sides of the box. That is, Sender 1 sends $X_1 =$ "$m_1\, a$" and Sender 2 sends $X_2 =$ "$m_2\, b$" where a, b are the outputs of the PR box. This strategy guarantees $a \oplus b = m_1 m_2$. The possible sets of encoded channel inputs produced by this strategy are listed in Table IV. Comparing that to Table III reveals that the resulting encoded message pairs are special for our channel: *they are exactly the combinations whereby Receiver 1 and Receiver 2 respectively receive the original 1-bit messages that Sender 1 and Sender 2 intended to send.* Hence, this super-quantum strategy enables perfect message transmission.

Corollary 1 (Capacity of Channel I with 1 bit of communication between senders)**.** *If senders are allowed to share one bit of communication, they will achieve a joint rate of 2.*

Proof. This strategy follows straightforwardly from the technique described in the proof of the previous lemma. This time, Sender 1 encodes her message bit by duplicating it. She then uses her one bit of communication by sending this message bit to Sender 2. Sender 2 uses this knowledge to replicate the action of the PR-box to pad his own one-bit message, and Table IV shows that this is always possible. Since this strategy achieves (deterministically) exactly the same input sets as the PR-box assisted strategy described above, it too achieves a joint rate of 2. □

(m_1, m_2) (PR-box input)	(a, b) (PR-box output)	Encoding (Sender 1, Sender 2)
(0,0)	(1,1) or (0,0)	(01,01) or (00,00)
(0,1)	(1,1) or (0,0)	(01,11) or (00,10)
(1,0)	(1,1) or (0,0)	(11,01) or (10,00)
(1,1)	(0,1) or (1,0)	(10,11) or (11,10)

TABLE IV: The rightmost column shows all possible combinations of the two senders' inputs to the channel using the encoding strategy described above: each sender's PR-box output (either a or b) is concatenated with her input m_i.

C. Capacity of Channel I with quantum assistance

We saw that Alice and Bob can coordinate their inputs using a non-classical resource to achieve perfect coding. Does a quantum resource suffice, or only a super-quantum one?

Lemma 3 (Capacity of Channel I if senders share entanglement)**.** *If the senders are allowed to share an entangled quantum state $|\Phi\rangle$ of dimension $2 \times n$,*

$$C_{sum} < 2.$$

Proof. For this proof, we borrow notation from [14]. Let $\mathcal{P}$ denote the set of all POVMs acting on a single qubit, and $O_{\mathcal{P}}$ denote the set of all outcomes for the POVM $\mathcal{P}$. Let m_i, X_i, Y_i denote the message bits, encoded message bits (input to channel) and channel output bits respectively, where the subscript i denotes the respective sender-receiver pair.The two senders share an entangled state $|\Phi\rangle$.

$$\begin{gathered} \mathcal{C}_1 : m_1 \to \mathcal{P}_1 \quad \mathcal{C}_2 : m_2 \to \mathcal{P}_2 \\ \mathcal{E}_1 : m_1 \times O_{\mathcal{P}_1|\Phi\rangle} \to X_1 \quad \mathcal{E}_2 : m_2 \times O_{\mathcal{P}_2|\Phi\rangle} \to X_2 \\ \text{Channel} : (X_1, X_2) \to (Y_1, Y_2) \\ \mathcal{D} : (Y_1, Y_2) \to (\hat{m_1}, \hat{m_2}) \end{gathered} \tag{10}$$

Any quantum strategy for communication can be mathematically represented as four consecutive mappings $(\mathcal{C}_i, \mathcal{E}_i, \text{Channel}, \mathcal{D})$. The senders independently choose a POVM $(\mathcal{C}_i)$ depending on their message bit m_i, apply that POVM to their share of the entangled state, and apply an encoding function $(\mathcal{E}_i)$ that maps the measurement outcome of the POVM to a 2-bit input to the channel. These bits go through the channel and the output of the channel is decoded $(\mathcal{D})$ by the two receivers. This process is illustrated in Figure 4.

This is indeed the most general form of a quantum communications strategy; Naimark's theorem guarantees that a POVM is mathematically equivalent to a general measurement, and the most general decoder looks at both the channel outputs (including as a special case a restricted decoding strategy where receivers do not communicate). This model (and the proof it inspires) is in very much the same spirit as the model in [14], which was used to prove a similar result for two-player pseudo-telepathy games.

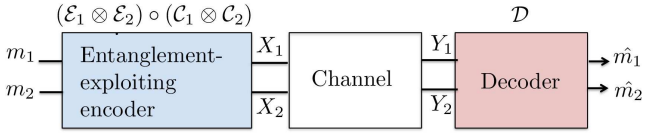

FIG. 4: Model of a quantum communication system over this channel.

Our goal is to show that that if there exists a quantum strategy that achieves rate 2 (ie. perfect coding), there is a classical strategy that achieves the same rate. But, since there is *not* a classical strategy that achieves perfect coding, there cannot be a quantum one.

We assume that any decoding strategy depends deterministically, and solely, on the bits that the receivers receive. That is, every time a particular (Y_1, Y_2) is received, the decoding step infers a fixed, corresponding, $\hat{m_1}$ and $\hat{m_2}$. Demanding a rate of 2 rules out any probabilistic decoding strategy, so the inferred $\hat{m_1}$ and $\hat{m_2}$ have to be the right ones. The question is now whether there exist functions $(\mathcal{E}_1 \otimes \mathcal{E}_2) \circ (\mathcal{C}_1 \otimes \mathcal{C}_2)$ such that the overall map from $m_1 \times m_2$ to $(X_1 \times X_2)/(Y_1 \times Y_2)$ is injective. This means we can group the 16 options for (X_1, X_2) based on the resulting (Y_1, Y_2) and stipulate that the entanglement-assisted $(\mathcal{E}_1 \otimes \mathcal{E}_2) \circ (\mathcal{C}_1 \otimes \mathcal{C}_2)$ must achieve the following map:

$(\mathbf{m_1}, \mathbf{m_2})$	$(\mathbf{X_1}, \mathbf{X_2})$
$(m_1, m_2)_a$	$(00, 00), (01, 01), (10, 10), (11, 11)$
$(m_1, m_2)_b$	$(00, 01), (01, 00), (10, 11), (11, 10)$
$(m_1, m_2)_c$	$(00, 10), (01, 11), (10, 01), (11, 00)$
$(m_1, m_2)_d$	$(00, 11), (01, 10), (10, 00), (11, 01)$

TABLE V: Entanglement-assisted map between message bits and their encoding. $(m_1, m_2)_a, (m_1, m_2)_b, (m_1, m_2)_c, (m_1, m_2)_d$ must correspond to some permutation of the message set $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$.

Therefore, all we are asking of our copycat classical strategy is that, for any combination of message bits, it should encode them as some subset of the allowed encodings in the right column of the corresponding row – since this suffices for perfect decoding. Note that the assumption that our quantum strategy is *perfect* is key – our classical strategy only needs to never produce an illegal output, even though some legal outputs may never occur.

It turns out that it is entirely possible to devise a classical strategy that never produces an output that is illicit from a POVM, and this is proved in Appendix B. □

IV. CHANNEL II

In this section, we present a class of related channels to Channel I that displays yet stronger capacity separations:

$$\mathbf{C_{classical} < C_{quantum} < C_{super\text{-}quantum}}.$$

To get Channel II, we modify Channel I by allowing now two types of outputs. Consider the cells in Table VI. The cells with *ee* are outputs that always get erased. All other cells are outputs that are *erased* with probability $1-\epsilon$, but with probability ϵ output the two bits stated. We will see later that the parameter ϵ can be tuned to change the magnitude of the capacity separations. We prove the desired inequalities for this channel when ϵ is taken to be small.

$X_1 \backslash X_2$	**00**	**01**	**10**	**11**
00	00/ee	ee	01/ee	ee
01	ee	00/ee	ee	01/ee
10	10/ee	ee	ee	11/ee
11	ee	10/ee	11/ee	ee

TABLE VI: Channel II: a variation on Channel I in which the channel outputs not corresponding to the PR-box-encoded joint inputs are erased with probability 1, and the channel outputs corresponding to the PR-box-encoded joint inputs are erased with probability $1-\epsilon$. Erased bits are denoted by 'e'.

A. Super-quantum and entanglement-assisted capacities of Channel II

Lemma 4 (Capacity of Channel II with super-quantum resources)**.** *There exists a super-quantum-assisted strategy on Channel II that achieves* $R_1 + R_2 = 2\epsilon$.

Proof. Encoding proceeds exactly as in in the previous section. Why this works is best visualized by comparing our channel in Table VI to the set of encoded messages produced by the PR-box strategy from the previous section, summarized in the left half of Table III – *the encoding only produces the channel inputs whose outputs are erased with probability* $1-\epsilon$ *by Channel II.* Since preserved outputs contain exactly the first bits of each sender's message, they are perfectly decoded by each receiver. This therefore amounts to a binary erasure channel for each sender-receiver pair with erasure parameter $1-\epsilon$. This gives a joint rate of 2ϵ. □

This intuition is this: any classical choice of input alphabets for the two senders results in at least one combination of inputs that is always erased by the channel. Using a PR-box helps us avoid these 'bad' input combinations, and using entanglement helps us avoid them with probability $\cos^2(\frac{\pi}{8}) \approx 0.854$, as we will see next.

Lemma 5 (Achievable rate with senders sharing entanglement)**.** *There exists an entanglement-assisted strategy on Channel II that achieves* $R_1 + R_2 = \left[2\cos^2(\frac{\pi}{8})\right]\epsilon$.

Proof. We will describe such a strategy. The encoding step is a simple extension of the previous one: in place of the PR-box, let the two senders share the CHSH entangled pair, $|\Psi\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$. This is the same state that they can use to win the CHSH game with higher-than-classical probability. The essence of the strategy is that they play a CHSH game to communicate. Recall that the winning condition of the CHSH game is that

$$a \oplus b = r \wedge s \tag{11}$$

where $r :=$ player 1's question, $s :=$ player 2's question, $a :=$ player 1's response, $b :=$ player 2's response. With a shared Bell state, the two players can perform measurements on their state in such a way that their question and response bits fulfill Equation (11) with probability $\cos^2(\frac{\pi}{8})$. But observe that this is exactly the equation that *always* holds true for all licit input-output pairs (inputs: r, s, outputs: a, b) from a PR-box. Therefore, instead of concatenating the PR-box output with their message bit, the senders now concatenate a or b with their message bit, where a and b are obtained by measurements on their shared entangled state. That is, a and b are their 'response' bit in the CHSH game if their desired message had been their 'question' bit from the referee.

This encoding strategy allows for pretty-good communication. We may observe that we obtain a 'good' encoding (one that lands on the double-valued cells in Table VI) with probability $\cos^2\left(\frac{\pi}{8}\right) \approx 0.854$; we obtain a 'bad' encoding (one that lands on the single-valued cells, thus always gets erased) with probability $\sin^2\left(\frac{\pi}{8}\right) \approx 0.147$. Hence, each sender gets his input bit erased with probability $\alpha = \sin^2(\frac{\pi}{8}) + (1-\epsilon)\cos^2(\frac{\pi}{8})$, and transmitted perfectly with probability $\epsilon\cos^2(\frac{\pi}{8})$. This amounts to each sender-receiver pair experiencing a binary erasure channel with erasure probability α. Since the capacity of a binary erasure channel is $1-\alpha$, the joint rate achieved by such a strategy is $2(1-\alpha) \approx 1.707\epsilon$. □

B. Capacity separations for Channel II

Finally, we reach the capstone lemmas of this section:

Lemma 6 (Classical vs. quantum capacities of Channel II)**.** *For sufficiently small* ϵ, $\boldsymbol{C^{(\epsilon)}}_{\boldsymbol{classical}} < \boldsymbol{C^{(\epsilon)}}_{\boldsymbol{quantum}}$.

Proof. In Appendix C we prove a lemma that upper-bounds $\mathbf{C}^{(p)}{}_{\mathbf{classical}}$ by $1.255\epsilon + O(\epsilon^2)$. This proof rests on the following Lemma.

Lemma 7. *Suppose the inputs to a channel are five symbols, 1, 2, 3, 4, and ?. The first four symbols are replaced by ? with probability $1-\epsilon$ and transmitted intact with probability ϵ, and the last symbol is always sent as ?. Furthermore, suppose that these symbols must be sent with probabilities p_1, p_2, p_3, p_4, and $p_?$, with these probabilities adding up to 1. Then the capacity of this channel is*

$$-\epsilon\left(p_1\log p_1 + p_2\log p_2 + p_3\log p_3 + p_4\log p_4\right) + O(\epsilon^2) \tag{12}$$

(This lemma is perfectly mappable to our problem; we need only consider the p_is to be the probabilities of an xx/ee state being sent, where xx is one of $\{00, 01, 10, 11\}$. That is, eventually we wish to make the replacement:

$$\begin{aligned} p_1 &\leftarrow pq\alpha \\ p_2 &\leftarrow p(1-q)\beta \\ p_3 &\leftarrow (1-p)q\gamma \\ p_4 &\leftarrow (1-p)(1-q)\delta) \end{aligned} \tag{13}$$

Proof: The probability of the output symbol ? is

$$p_?+(1-\epsilon)(p_1+p_2+p_3+p_4) = p_?+(1-\epsilon)(1-p_?) = 1-\epsilon+\epsilon p_?.$$

We simply plug into Shannon's formula (X = channel input, Y = channel output):

$$I(X;Y) = H(Y) - H(Y|X),$$

where

$$\begin{aligned} H(Y|X) &= \Sigma_{i=1}^4 p_i H(\epsilon) \\ H(Y) &= -\Sigma_{i=1}^4 \epsilon p_i \log(\epsilon p_i) \\ &\quad - (\epsilon p_? + 1 - \epsilon)\log(\epsilon p_? + 1 - \epsilon). \end{aligned}$$

The second term of these goes to zero as $\epsilon \to 0$, so we ignore it henceforth.

$$\begin{aligned} &H(Y) - H(Y|X) \\ &= \Sigma_{i=1}^4 - p_i[\epsilon\log(\epsilon p_i) + H(\epsilon)] \\ &= \Sigma_{i=1}^4 - p_i[\epsilon\log(\epsilon p_i) - \epsilon\log(\epsilon) - (1-\epsilon)\log(1-\epsilon)] \\ &= \Sigma_{i=1}^4 - \epsilon p_i\log(p_i) + (1-\epsilon)p_i\log(1-\epsilon) \end{aligned}$$

Taking the limit as $\epsilon \to 0$, the last term of the above disappears and we get the desired expression. □
Please refer to Appendix C for the rest of the proof that $\mathbf{C}^{(p)}{}_{\mathbf{classical}} < 1.255\epsilon + O(\epsilon^2)$. We have also seen an entanglement-assisted strategy that achieves a joint rate of 1.707ϵ, which must therefore be a *lower*-bound for the entanglement-assisted capacity. Therefore, if ϵ is chosen small enough such that the second-order terms can be ignored, we may achieve $\mathbf{C}^{(p)}{}_{\mathbf{classical}} < \mathbf{C}^{(p)}{}_{\mathbf{quantum}}$. This is suffices to prove the desired capacity separation. □

The following is a corollary of Lemma 7:

Lemma 8 (Quantum vs. Super-quantum capacities of Channel II)**.** *For sufficiently small ϵ, $\boldsymbol{C}^{(\epsilon)}{}_{\boldsymbol{quantum}} < \boldsymbol{C}^{(\epsilon)}{}_{\boldsymbol{super\text{-}quantum}}$*

Proof. This statement follows straightforwardly from Equation (12), which gives us an expression (up to first order in ϵ) for the capacity of the channel in terms of the probabilities of the xx/ee and ee states being sent. This expression is valid for any coding strategy, no matter what types of resources are used.

We also know that even with entanglement, the maximum percentage of time that an xx/ee state is sent is 0.8536, and this follows from CHSH (refer to the proof of Lemma 5 for why). Using our convention for defining the p_is, this translates to the constraint that

$$\Sigma_i p_i = 0.8536. \tag{14}$$

Therefore, the entanglement-assisted capacity is upper-bounded by the maximum value of the LHS of Equation (12),

$$-\epsilon\left(p_1\log p_1 + p_2\log p_2 + p_3\log p_3 + p_4\log p_4\right) + O(\epsilon^2).$$

Under the constraint of Equation (14), this is strictly less than 2ϵ (the super-quantum capacity). Furthermore, the bound holds even if the senders are allowed to share other types of entanglement than just $2\times n$ entanglement, since that does not affect the maximum success probability of the CHSH game (from which we derived the constraint (14)). □

V. DISCUSSION

This work continues in the vein of many studies (including but not limited to: [8, 9, 15]) exploring the implications of super-quantum theories (as represented by non-local boxes) for communication. In both types of channels that we have proposed, all the PR-boxes are assumed to be perfect. We would like to see a rigorous proof that these separations can be maintained even if the senders are provided a noisy PR-box and allowed multiple uses of it for non-locality distillation.

We have also only considered interference channels operating on discrete-variable bits because this is a proof-of-concept. In real life, many communication scenarios where multiple uncoordinated links share a common communication medium can be represented as interference channels (albeit ones where transmitted messages take on continuous values in $\mathbb{C}$ subject to Gaussian noise). Therefore, some work is needed to replicate the above separations on an general interference channel, or at the very least, characterize channels and coding strategies in a way that optimizes them for each of the three classes of resources.

Our choice to limit our channel to handling only classical information (as opposed to density matrices representing quantum information) proved fruitful, as it paved

the way for proofs that rely on classical information theory, as well as some results from pseudo-telepathy games where the referee, too, accepts only a discrete (albeit distributed) set of outcomes. In hindsight, this connection seems natural; pseudo-telepathy games exhibit the twin boons of being *known* to demonstrate super-quantum-to-quantum separations, and having had winning strategies (in a few cases) characterized and generalized to an arbitrarily large number of parties[16, 17]! For this reason, the literature on pseudo-telepathy and XOR games is therefore a natural starting point for the task of mapping the capacity separations described in this paper to the multi-sender ($n \geq 3$) case.

VI. CONCLUSION

We have exhibited two types of interference channels that show the following new separations in classical capacity on the given classes of resources:

- Channel I: $\mathbf{C_{classical}}, \mathbf{C_{quantum}} < \mathbf{C_{super\text{-}quantum}}$
- Channel II: $\mathbf{C_{classical}} < \mathbf{C_{quantum}} < \mathbf{C_{super\text{-}quantum}}$.

The takeaway point from this research is that PR-boxes shared between a set of transmitters can be used for better channel communication – a task for which they have never been considered.

VII. ACKNOWLEDGMENTS

The authors thank Sandu Popescu and Isaac Chuang for discussions and suggestions to improve this paper. Y.Q. was supported by a DSTA Undergraduate Scholarship (Overseas) from the Singapore government. P.W.S. was supported in part by NSF grant CCF-121-8176, and by the NSF through the Science and Technology Center for Science of Information under Grant CCF0-939370.

Appendix A: Remainder of proof of Lemma 1 for Channel I

Here we show that if Sender 1 uses the alphabet $\{00, 01\}$ and Sender 2 uses $\{00, 10\}$ (Table VII depicts this schematically), then the second sender-receiver pair cannot communicate perfectly, and therefore $R_2 < 1$ as we asserted. The same turns out to be true for the other 3 cases.

To get $I(X_1 : Y_1) = H(X_1) - H(X_1|Y_1) = 1$ when there are only two options for X_1, the first term must take its maximal value of 1, which can only happen if X_1 is uniformly distributed over $\{00, 01\}$. Let X_2 send 00 with probability c and 01 with probability $1 - c$. This is shown on the left in Table VII. Since we will be interested in calculating $I(X_2 : Y_2)$, we also calculate the input-output probability distribution experienced by sender-receiver pair 2, shown on the right in Table VII.

$X_1 \backslash X_2$	**00**	**10**
00	00	01
01	11	10

$X_2 \backslash Y_2$	**0**	**1**
00	$\frac{c}{2}$	$\frac{c}{2}$
10	$\frac{1-c}{2}$	$\frac{1-c}{2}$

TABLE VII: Left: reduced alphabets of senders and resulting output to the receivers (in the format Y_1Y_2). Right: Joint probability distribution experienced by the second sender-receiver pair on this coding scheme.

Referring to the right side of Table VII, we obtain

$$\begin{aligned} I(X_2 : Y_2) &= H(X_2) + H(Y_2) - H(X_2, Y_2) \\ &= [-c \log c - (1-c)\log(1-c)] + 1 \\ &\quad - \left[2\left(-\frac{c}{2}\log\frac{c}{2}\right) + 2\left(-\frac{1-c}{2}\log\frac{1-c}{2}\right)\right] \\ &= 0 \end{aligned} \tag{A1}$$

We have therefore shown that $I(X_1 : Y_1) = 1$ implies that $I(X_2 : Y_2) = 0$, so that $I(X_1 : Y_1) + I(X_2 : Y_2) = 2$ will never be achieved. **Put another way, perfect coding between one pair implies that the other pair can do no better than random guessing.**

Appendix B: A classical strategy that performs as well as a hypothetical perfect entanglement-assisted strategy on Channel I

The strategy will follow after the subsequent lemmas:

Lemma 9. *For any two-sender-receiver pair communication strategy that relies on the senders sharing some state $|\Phi\rangle$ of dimension 2×2, there exists a communication strategy that achieves the same rate where the senders are restricted to sharing a state of the form $|\Psi\rangle = \alpha|00\rangle + \beta|11\rangle$, where α and β are well-chosen positive real numbers.*

Proof. The key idea is to re-write $|\Phi\rangle$ in terms of its Schmidt decomposition, and then apply a unitary transformation to get $|\Psi\rangle$. Then, the senders may apply the quantum strategy whose existence we have assumed. More precisely, there exist orthogonal bases $\{|A_0\rangle, |A_1\rangle\}$ for Sender 1 and $\{|B_0\rangle, |B_1\rangle\}$ for Sender 2 such that $|\Phi\rangle$ can be rewritten as

$$|\Phi\rangle = \alpha |A_0\rangle |B_0\rangle + \beta |A_1\rangle |B_1\rangle .$$

From there it is easy to see that Sender 1 may apply the unitary transformation $|A_0\rangle\langle 0| + |A_1\rangle\langle 1|$, and Sender 2 may apply the unitary transformation $|B_0\rangle\langle 0| + |B_1\rangle\langle 1|$, to their qubits, to transform $|\Psi\rangle$ into $|\Phi\rangle$. Any such unitary U is completely accounted for in our model of communication in 10 by applying it to the POVMs

M_i that the senders choose for their states (which preserves its POVM properties), that is, using the property $U|\Phi\rangle = |\Psi\rangle \rightarrow \langle\Phi| M_i |\Phi\rangle = \langle\Phi| U M_i U^\dagger |\Phi\rangle$. □

Since the following two lemmas are almost identical to the ones in [14], we merely cite them and leave the reader to refer to [14] for their proofs.

Lemma 10. *For any two-party quantum communication protocol that uses an entangled state of dimension $d_A \times d_B$, there exists a two-party quantum communication protocol that uses a state of dimension $d \times d$ where $d := \min(d_A, d_B)$.*

This justifies the audaciously general claim made in Lemma 3 that *no* quantum state of dimension $2 \times d$ could possibly enable a perfect joint rate for communication. The proof is similar to the proof of Lemma 9 and relies on the following fact from the Schmidt decomposition: if H_1 and H_2 are Hilbert spaces of dimensions n, m respectively, and we assume without loss of generality that $n \geq m$, for any vector $w \in H_1 \otimes H_2$, there exist orthonormal bases $\{u_i, 1 \leq i \leq n\}$ for H_1 and $\{v_j, 1 \leq j \leq m\}$ for H_2 respectively such that

$$w = \Sigma_{i=1}^{m} \alpha_i u_i \otimes v_i. \tag{B1}$$

Lemma 11. *Any POVM can be written in a way such that all its elements are proportional to one-dimensional projectors. Each such projector can be re-written in the form*

$$P = \begin{pmatrix} \cos^2(\theta) & e^{-i\phi}\sin(\theta)\cos(\theta) \\ e^{i\phi}\sin(\theta)\cos(\theta) & \sin^2(\theta) \end{pmatrix} \tag{B2}$$

for appropriate angles $0 \leq \theta \leq \frac{\pi}{2}$ and $0 \leq \phi \leq 2\pi$. Since this representation is unique, we may associate each such projector with a three-dimensional unit vector $\vec{v} = (\sin(2\theta)\cos(\phi), \sin(2\theta)\sin(\phi), \cos(2\theta))$.

Finally, the classical strategy promised three lemmas ago is described. Thanks to Lemma 9, we may assume that the two senders are using an entangled state of the form $|\Psi\rangle = \alpha|00\rangle + \beta|11\rangle$, where α and β are strictly positive real numbers.

Suppose a quantum strategy exists and the POVMs applied by the two senders, $M^x := \mathcal{X}(x) = \{\gamma_i^x P_i^x\}$ and $N^y := \mathcal{Y}(y) = \{\gamma_j^y Q_j^y\}$ have been fixed beforehand for each $x, y \in \{0, 1\}$. We will show that any measurement outcome (i, j) on $|\Psi\rangle$ as described in the first row of Equations 10 can be replicated perfectly classically. The probability of getting the tuple (i, j) is:

$$\begin{aligned} \Pr[i,j] &= \langle\Psi| (\gamma_i^x P_i^x) \otimes (\gamma_j^y Q_j^y) |\Psi\rangle \\ &= \gamma_i^x \gamma_j^x \left[\alpha^2 \cos^2(\theta_i^x)\cos^2(\theta_j^y)\right. \\ &\quad +2\alpha\beta\left[\cos(\phi_i^x + \phi_j^y)\sin\theta_i^x \cos\theta_i^x \sin\theta_j^y \cos\theta_j^y\right] \\ &\quad \left.+\beta^2 \sin^2(\theta_i^x)\sin^2(\theta_j^y)\right] \\ &= \gamma_i^x \gamma_j^x (a^2 + b^2 + 2abc) \end{aligned} \tag{B3}$$

where $a := \alpha\cos(\theta_i^x)\cos(\theta_j^y)$, $b := \beta\sin(\theta_i^x)\sin(\theta_j^y)$ and $c := \cos(\phi_i^x + \phi_j^y)$. Using the AM-GM inequality and the fact that $|c| \leq 1$ we may show that $\Pr[i, j]$ can only vanish if one of the following two things are true of the POVMs used by the two senders ($\{\gamma_i^x P_i^x\}, \{\gamma_j^y Q_j^y\}$):

- $a = b = 0$
  Attained if $\theta_i^x = 0, \theta_j^y = \pi/2$ or vice versa – that is, either P_i^x or Q_j^y belongs to neither hemisphere.
- $a = b$ and $c = -1$.
  Attained if $\phi_i^x + \phi_j^y = \pi$(both projectors in eastern hemisphere) or $\phi_i^x + \phi_j^y = 3\pi$ (both projectors in western hemisphere).

But all our classical strategy needs to do is to choose a classical tuple, (i, j), such that the corresponding quantum POVM elements, P_i^x and Q_j^y, would not fulfill either of these conditions. To do this, it suffices for Sender 1, knowing $M^x := \{\gamma_i^x P_i^x\}$, to choose an i such that P_i^x belongs to the eastern hemisphere and for Sender 2, knowing $N^y := \{\gamma_j^y Q_j^y\}$, to choose a j such that Q_j^y belongs to the western hemisphere (without actually measuring anything). This is always possible since POVM elements have to sum to the identity. They may then carry out the (classical) mappings $\mathcal{A}$ and $\mathcal{B}$ on their message bits and POVM 'outcomes' as per normal.

Appendix C: A proof of an upper bound on the classical capacity of Channel II

Channel II has been replicated in Table VIII for your convenience. We would like to prove that for some values of the parameter ϵ, the entanglement-assisted capacity beats the classical capacity.

$X_1 \backslash X_2$	**00**	**01**	**10**	**11**
00	00/ee	ee	01/ee	ee
01	ee	00/ee	ee	01/ee
10	10/ee	ee	ee	11/ee
11	ee	10/ee	11/ee	ee

TABLE VIII: Channel II, reproduced here. In bold are the senders' inputs, and the table shows the resulting channel outputs. Erased bits are denoted by e, and the ee has probability $1 - \epsilon$ in the squares it shares with numerical values.

Let the probability of X_1 sending 00 or 01 be p, and the probability of X_2 sending 00 or 01 be q.

The proof proceeds in three steps. We aim to show that the parameter ϵ governing the rate for the best classical strategy can be tuned small enough that that quantum-assisted capacity is larger than the classical capacity. Therefore, we first establish a relation that constrains the probabilities of the various possible output symbols for any classical strategy. Next, we find an expression

for the classical capacity of the channel up to first order in ϵ. Using this relation, we find an upper bound on the classical capacity in terms of ϵ. This completes the proof.

The first thing to prove is that

Lemma 12. *There are numbers α, β, γ and δ with*

$$\alpha+\beta+\gamma+\delta \leq 3 \qquad \textit{and} \qquad \alpha,\beta,\gamma,\delta < 1$$

such that if we look at the output,

$$\begin{aligned}
\Pr(00) &= \epsilon pq\alpha, \\
\Pr(01) &= \epsilon p(1-q)\beta, \\
\Pr(10) &= \epsilon (1-p)q\gamma, \\
\Pr(11) &= \epsilon (1-p)(1-q)\delta.
\end{aligned} \tag{C1}$$

Proof: First, let's assume that Alice and Bob input a product distribution. The most general thing they can do is input a convex combination of product distributions, and the result for convex combinations follows straightforwardly from the result for product distributions.

Now, let Alice's input be expressed as a vector.

$$(p_{00}, p_{01}, p_{10}, p_{11})$$

meaning that with probability p_{ij} Alice inputs bit string ij. Note that $p_{00}+p_{01}=p$ and $p_{10}+p_{11}=1-p$. We can decompose this vector into a sum of 'basis vectors' $\{u_i\}$, each with two non-zero entries a_i and b_i. We may stipulate $a_i/b_i = p/(1-p)$:

$$\begin{aligned}
&(p_{00}, p_{01}, p_{10}, p_{11}) = \Sigma_{i=1}^{4} u_i \\
&= (a_1, 0, b_1, 0) + (a_2, 0, 0, b_2) + (0, a_3, b_3, 0) \\
&+ (0, a_4, 0, b_4).
\end{aligned}$$

Similarly, we decompose Bob's input distribution into $\{v_i\}$ such that $c_i/d_i = q/(1-q)$.

$$\begin{aligned}
&(q_{00}, q_{01}, q_{10}, q_{11}) = \Sigma_{i=1}^{4} v_i \\
&= (c_1, 0, d_1, 0) + (c_2, 0, 0, d_2) + (0, c_3, d_3, 0) \\
&+ (0, c_4, 0, d_4).
\end{aligned}$$

This notation permits the senders' joint inputs to be written as a linear combination of 16 terms, $\Sigma_{i,j} u_i v_j$. Each such term induces a probability distribution over channel outputs, which we shall express using the same naming convention for the proportionality factors as in Equation (C1), but with an additional subscript i,j. We claim that for each basis vector of the joint input distribution (indexed by i,j), $\alpha_{ij}+\beta_{ij}+\gamma_{ij}+\delta_{ij} \leq 3$.

For instance, if we take the vectors $u_2 = (a_2, 0, 0, b_2)$ and $v_1 = (c_1, 0, d_1, 0)$, we know that $\Pr(u_2) = a_2 + b_2$ and $a_2 = p\Pr(u_2)$, $b_2 = (1-p)\Pr(u_2)$ by our convention for choosing the entries of the basis vectors. Similarly, $c_1 = q\Pr(u_2, v_1)$ and $d_1 = (1-q)\Pr(u_2, v_1)$. Then we have

$$\begin{aligned}
\Pr_{21}(00) &= \epsilon a_2 c_1 = \epsilon pq\alpha_{21} \Pr(u_2, v_1) \\
\Pr_{21}(01) &= \epsilon a_2 d_1 = \epsilon p(1-q)\beta_{21} \Pr(u_2, v_1) \\
\Pr_{21}(10) &= 0 \quad = \epsilon (1-p)q\gamma_{21} \Pr(u_2, v_1) \\
\Pr_{21}(11) &= \epsilon b_2 d_1 = \epsilon (1-p)(1-q)\delta_{21} \Pr(u_2, v_1)
\end{aligned}$$

where each $\alpha_{ij}, \beta_{ij}, \gamma_{ij}, \delta_{ij}$ is 1 if the corresponding probability is non-zero and 0 otherwise. In this instance, $\alpha_{21} = \beta_{21} = \delta_{21} = 1$, $\gamma_{21} = 0$. In particular, $\alpha_{21}+\beta_{21}+\gamma_{21}+\delta_{21} \leq 3$ and we can easily check that this is true for all choices of i,j. It is straightforward to extend this property to $\alpha+\beta+\gamma+\delta$. We have

$$\alpha = \Sigma_{ij} \Pr(u_i, v_j)\alpha_{ij}$$

and so on for the other greek letters. So

$$\begin{aligned}
&\alpha+\beta+\gamma+\delta \\
&= \Sigma_{ij} \Pr(u_i, v_j)(\alpha_{ij}+\beta_{ij}+\gamma_{ij}+\delta_{ij}) \\
&\leq 3\left(\Sigma_{ij} \Pr(u_i, v_j)\right) = 3
\end{aligned}$$

□

The next step is to find an upper bound to the classical capacity of this channel up to first order in ϵ. It follows from Lemma 7 a channel with these probabilities cannot send much more than

$$\begin{aligned}
&-\epsilon\,[pq\alpha \log(pq\alpha) + p(1-q)\beta \log(p(1-q)\beta) \\
&\quad +(1-p)q\gamma \log((1-p)q\gamma) \\
&\quad +(1-p)(1-q)\delta \log((1-p)(1-q)\delta)]
\end{aligned} \tag{C2}$$

information. Now, we relax the problem. We no longer require that we have a product distribution. Choose p_i as described in Equations 13 and choose α, β, γ, δ with $\alpha+\beta+\gamma+\delta \leq 3$ as in Equation C1. The capacity of our channel, by the above lemma, is at most

$$\begin{aligned}
&-\epsilon(\alpha k_{00} \log \alpha k_{00} + \beta k_{01} \log \beta k_{01} \\
&+ \gamma k_{10} \log \gamma k_{10} + \delta k_{11} \log \delta k_{11}) + O(\epsilon^2).
\end{aligned} \tag{C3}$$

where we have further defined $k_{00} = pq, k_{01} = p(1-q), k_{10} = (1-p)q, k_{11} = (1-p)(1-q)$ such that $\Sigma_{i,j} k_{ij} = 1$.

Our aim now is to find values of $(k_{00}, k_{01}, k_{10}, k_{11})$ and $(\alpha, \beta, \gamma, \delta)$ which maximize this expression, which would give us an upper bound on the classical capacity.

We can do this in several steps, which we outline below.

First, we observe that one of αk_{00}, βk_{01}, γk_{10}, δk_{11}, is at most $3/16$, and recall that $\alpha, \beta, \gamma, \delta < 1$. But $f(x) = x \log x$ is maximized when $x = 1/e$. These two facts let us assume that

$$\alpha+\beta+\gamma+\delta = 3$$

at the point where Equation (C3) is maximized because of the following: suppose $\alpha k_{00} < \frac{3}{16}$ (and therefore $<$

$\frac{1}{e}$). Then if $\alpha+\beta+\gamma+\delta<3$, we could increase the capacity, Equation (C3), by increasing α, and therefore our original choice could not have maximized the capacity.

Next, we show that at the maximum $\alpha=3k_{00}$, $\beta=3k_{01}$, $\gamma=3k_{10}$, $\delta=3k_{11}$.

Proof: This eventually falls out from formulating the problem with Lagrange multipliers with the constraints $\alpha+\beta+\gamma+\delta=3$ and $k_{00}+k_{01}+k_{10}+k_{11}=1$. But we take a quicker tack: we show that we can increase the rate if this is not the case.

Suppose $\alpha-3k_{00}=\delta_1$ and $\beta-3k_{01}=-\delta_2$. Let ϵ be $\frac{1}{2}\min(\delta_1,\delta_2)$. We can increase αk_{00} and βk_{01} by replacing

$$\begin{aligned}\alpha' &= \alpha-\epsilon\\ \beta' &= \beta+\epsilon\\ k'_{00} &= k_{00}+\epsilon/3\\ k'_{01} &= k_{00}-\epsilon/3\end{aligned}$$

which respects the constraints while leaving the other four variables unchanged. Since the probability of getting a faithfully-transmitted output increases with both αk_{00} and βk_{01} (recalling how k_{ij} was defined), so should the rate increase. □

Finally, we need to show that the capacity is maximized when $k_{00}=k_{01}=k_{10}=k_{11}=1/4$ and $\alpha=\beta=\gamma=\delta=3/4$.

Proof: Let

$$f=-3x^2\log 3x^2.$$

We need to find

$$\max_{k_{ij}} f(k_{00})+f(k_{01})+f(k_{10})+f(k_{11})$$
$$s.t.\ \Sigma_{i,j}\,k_{ij}=1$$

Define

$$\begin{aligned}\mathcal{L}(k_{ij},\lambda) = f(k_{00})+f(k_{01})+f(k_{10})+f(k_{11})\\ +\lambda(k_{00}+k_{01}+k_{10}+k_{11}-1)\end{aligned}$$

and we would like to get $\nabla_{ij,\lambda}\mathcal{L}=0$, so we need $f'(k_{00})=f'(k_{01})=f'(k_{10})=f'(k_{11})$.

Looking at the graph of f' shows that for any t, there are at most two points x_1 and x_2 with $0<x_1\le x_2<1$ where $f'(x_1)=f'(x_2)=t$. This shows that in the maximum of $f(k_{00})+f(k_{01})+f(k_{10})+f(k_{11})$, there are at most two different values of k_{ij}. However, the asymmetry of the graph for the portions where $0<x<1$ and $f'(x)>0$ makes it clear that one cannot assign more than one value to k_{ij}in such a way that $\Sigma k_{ij}=1$. We may thus conclude that there is only one value of k_{ij} that maximizes this expression, and that must be $k_{ij}=1/4$.

Putting the numbers into Equation (C3), we may conclude that the upper bound on the classical capacity is $1.255\epsilon+O(\epsilon^2)$.

[1] J. Bell, Physics, 1 , 195 (1964).
[2] S. Popescu, Nature **10**, 264 (2014).
[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86** (2014).
[4] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[5] B. S. Cirelson, Letters in Mathematical Physics **4**, 93 (March 1980).
[6] D. Rohrlich and S. Popescu, Foundations of Physics **24** (1994).
[7] P. Lamontagne, "Non-local boxes," (2014).
[8] W. van Dam, Journal of Natural Computing **12**, 9 (2013).
[9] H. Buhrman, M. Christandl, F. Unger, S. Wehner, and A. Winter, Proc. Royal Soc. A. (2006).
[10] C. E. Shannon, *A Mathematical Theory of Communication*, Tech. Rep. (ATT Bell Laboratories, 1948).
[11] A. El Gamal and Y.-H. Kim, *Network information theory* (Cambridge University Press, 2012).
[12] H. Sato, IEEE Transactions on Information Theory **23**, 295 (1977).
[13] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu, Phys. Rev. Lett. **94**, 220403 (2005).
[14] G. Brassard, A. A. Methot, and A. Tapp, Quantum Information & Computation **5**, 275 (2005).
[15] S. Wolf and J. Wullschleger, in *Information Theory Workshop, 2006. ITW '06 Punta del Este. IEEE* (2006).
[16] A. Broadbent and A. A. Méthot, Journal of Theoretical Computer Science **358** (2006).
[17] A. Arkhipov, *Extending and Characterizing Quantum Magic Games*, Master's thesis, MIT (2012).