Optimal entanglement-assisted one-shot classical communication

Brett Hemenway,1,* Carl A. Miller,2,† Yaoyun Shi,2,‡ and Mary Wootters1,§
1Mathematics Department, University of Michigan, Ann Arbor, Michigan 48109, USA
2Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109, USA

(Received 21 December 2011; revised manuscript received 5 February 2013; published 3 June 2013)

The one-shot success probability of a noisy classical channel for transmitting one classical bit is the optimal probability with which the bit can be successfully sent via a single use of the channel. Prevedel et al. [Phys. Rev. Lett. 106, 110505 (2011)] recently showed that for a specific channel, this quantity can be increased if the parties using the channel share an entangled quantum state. In this paper, we characterize the optimal entanglement-assisted protocols in terms of the radius of a set of operators associated with the channel. This characterization can be used to construct optimal entanglement-assisted protocols for a given classical channel and to prove the limits of such protocols. As an example, we show that the Prevedel et al. protocol is optimal for two-qubit entanglement. We also prove some tight upper bounds on the improvement that can be obtained from quantum and nonsignaling correlations.

DOI: 10.1103/PhysRevA.87.062301 PACS number(s): 03.67.Hk, 03.65.Ud

Suppose that two parties, Alice and Bob, communicate over a noisy classical channel. While there are many examples of how Alice and Bob may benefit when they upgrade to a quantum channel, examples in which shared entanglement improves communication over a classical channel have only recently been discovered [1–3]. That these examples exist at all is somewhat surprising, as neither shared entanglement [4] nor the assistance of nonsignaling correlations [3] can increase the classical capacity of the channel. So far, work in this direction has focused on the the (one-shot) zero error capacity, which measures the number of messages Alice can send to Bob perfectly [3,5–8], and the related notion of the one-shot success probability [2], which is the best probability with which Alice can successfully send a single bit to Bob. It is of interest to determine how shared entanglement affects these two quantities, as this will further our understanding of how resources from quantum mechanics can be used for communication.

Previous work on enhanced communication over a classical channel has focused on the assistance that can be provided by nonsignaling correlations. In this setting, both the zero error capacity and one-shot success probability can be written as the solution to linear programs [3,5]. Some upper bounds are known for the entanglement-assisted zero error capacity [6,8]; these bounds are often the best bounds available in the unassisted case, suggesting that there are strong limitations to the amount of assistance that entanglement can provide. Much less is known about the limits of quantum assistance for the one-shot success probability. In Ref. [2], Prevedel et al. give an example of a channel where the unassisted success probability Succ(N), the entanglement-assisted success probability SuccQ(N), and the nonsignaling-assisted success probability SuccNS(N) are all different. It is known that entanglement cannot be completely helpful: If Succ(N) is less than one, then so is SuccQ(N) [3]. However, the size of the gap between them has remained unquantified.

We use two distinct approaches to quantify the extent to which entanglement can help Alice and Bob. In our first approach, presented in Sec. II, we derive a simple formula for SuccQ(N) in terms of the dimension of the entanglement. This formula, which is given by maximizing a quantity over a family of positive semidefinite operators, is easy to work with. As an example of its applicability, we show in Sec. III that the protocol from Ref. [2] is in fact optimal for their channel and for two-dimensional entanglement assistance. Finally, in Sec. VI, we extend these results to the case where Alice’s message belongs to a set of arbitrary (finite) size, rather than being a single bit.

While our first approach is quite general, it does not give a closed form for the success probability. Our second approach obtains explicit closed-form upper bounds for the success probability. As a first step, we prove in Sec. IV the following general bound on nonsignaling assistance. Let $r$ be the number of elements in the input alphabet of $N$. Then,

$$\frac{\text{SuccNS}(N) - \frac{1}{2}}{\text{Succ}(N) - \frac{1}{2}} \leq 2 - \frac{2}{r}.$$ (1)

The quantity $\{\text{Succ}(N) - \frac{1}{2}\}$, called the bias (of the unassisted protocol), measures the advantage that Alice and Bob have over a random strategy; thus, relation (1) measures the additional advantage gained by nonsignaling correlations. Our proof of relation (1) uses the linear program characterization of SuccNS(N) from Ref. [5]. From this, in Sec. V we derive an upper bound on the amount of assistance from a binary quantum device; we use the fact that any quantum correlation can be decomposed into a probabilistic mixture of a local correlation and a nonsignaling correlation (the concept of local fraction). We show that both of these bounds are the best possible, in the sense that there are channels for which equality is achieved.

A common thread in both approaches above is the use of the radius of a subset of a normed vector space. The formulas we use for SuccQ(N) and SuccNS(N) both involve a concept of radius (under the operator norm and the $L^1$ norm, respectively).
In contrast, the unassisted quantity \( \text{Succ}(N) \) can be expressed in terms of the diameter of a particular set of vectors under the \( L^1 \) norm. Measuring nonsignaling assistance thus reduces to comparing diameter to radius. This is reminiscent of Jung’s theorem in Euclidean space [9].

Our results here can be seen as complementary to known results on using classical communication to simulate quantum and nonsignaling correlations [10,11].

I. NOTATION AND TERMINOLOGY

Throughout this paper, we assume that Alice is trying to transmit a single bit to Bob across a classical channel. Alice and Bob will have access to a two-part input-output device \( D \) (Fig. 1), which may be classical or quantum, or they may implement an arbitrary nonsignaling correlation.

Each two-part input-output device \( D \) gives rise to a correlation between Alice and Bob, given by

\[
\{D(pq|rs) \mid p \in P, q \in Q, r \in R, s \in S\}
\]

so that \( D(pq|rs) \) is the probability of outputs \( p \) and \( q \) given inputs \( r \) and \( s \). We abuse notation by identifying the device \( D \) with the correlation it induces.

We say a device is nonsignaling if the partial sums \( \sum_{p \in P} D(pq|rs) \) do not depend on \( r \), and the partial sums \( \sum_{q \in Q} D(pq|rs) \) do not depend on \( s \). We say a nonsignaling device \( D \) is quantum if there exist Hilbert spaces \( V_A \) and \( V_B \), families of positive operator-valued measures \( \{A^R_{pq}\}_{R\in I}, \{B^S_{pq}\}_{S\in I} \), and a density operator \( \Lambda \) on \( V_A \otimes V_B \) such that \( D(rs|pq) = \text{Tr}(A^S_{rs} \otimes B^R_{pq}) \Lambda \). The device is quantum with dimension \( n \) if both \( V_A \) and \( V_B \) are \( n \)-dimensional and binary if the input and output alphabets have size 2.

A classical channel \( N \) is given by a matrix of conditional probabilities \( \{N(y|x) \mid x \in X, y \in Y\} \), where \( N(y|x) \) is the probability of seeing an output \( y \in Y \) given the input \( x \in X \). For any channel \( N \), let Succ(\( N \)) denote the maximum probability with which a single bit can be sent across \( N \) (without assistance). Clearly, Succ(\( N \)) has the following simple formula.

**Proposition 1.** Let \( N \) be a classical channel, and for each \( x \in X \), let \( n_x = \{N(y|x) \mid y \in Y\} \in \mathbb{R}^Y \). Then,

\[
\text{Succ}(N) = \frac{1}{2} + \frac{1}{4} \text{Diam}_1(n_x \mid x \in X),
\]

where \( \text{Diam}_1(n_x \mid x \in X) := \max_{x,x' \in X} \|n_x - n_{x'}\|_1 \) denotes the diameter under the \( L^1 \)-norm.

Let Succ(\( N,D \)) denote the maximum probability for a single-bit transmission across \( N \) with the assistance of \( D \). If \( S \) is a set of two-part devices, write Succ(\( N \)) := sup_{\Delta \in S} Succ(\( N,D \)). We are concerned with three choices of \( S \). We consider the set NS of nonsignaling devices; the sets Q and Q_0 of quantum and n-dimensional quantum devices; and the set Q_0 of binary quantum devices.

II. GENERAL QUANTUM DEVICES

In this section, we derive a formula for Succ_{Q_0}(\( N \)) and give an example of how to use our formula. We use the radius of a finite set \( \{H_i\}_{i \in I} \) of Hermitian operators on a finite-dimensional Hilbert space \( V \), defined by \( \text{Rad} \{H_i\} := \min C \max \|H_i - C\| \), where the minimum is taken over all Hermitian operators \( C \) on \( V \). The following lemma gives an alternative expression for the radius through semidefinite programming duality.

**Lemma 2.** For any finite set \( \{H_i\}_{i \in I} \) of Hermitian operators on a finite-dimensional Hilbert space \( V \), the radius of \( \{H_i\} \) is equal to

\[
\max_{\lambda_i \geq 0, \lambda_i' \geq 0, \sum \lambda_i = \sum \lambda_i'} \left[ \sum_{i \in I} \text{Tr}(H_i - \lambda_i H_i^\dagger) \right],
\]

where the maximization is over all Hermitian operators \( \{\lambda_i\}_{i \in I} \) and \( \{\lambda_i'\}_{i \in I} \) on \( V \) satisfying the given constraints.

**Proof.** Any family of Hermitian operators \( \{H_i\} \) may be translated to a family \( \{H_i + W\} \) which contains the operator 0. This translation does not affect the radius or the expression from the statement of the lemma. Therefore, we may assume that \( \{H_i\} \) contains 0. By definition,

\[
\text{Rad} \{H_i\} = \min_{C,r} (r),
\]

where the maximization is over Hermitian operators \( C \) and real numbers \( r \). Since 0 \( \in \{H_i\} \), whenever the constraints in this maximization are satisfied we have in particular that \( C \geq -r \). Letting \( Z = C + r \), we obtain the following alternate expression:

\[
\text{Rad} \{H_i\} = \min_{Z,r} (r),
\]

By semidefinite programming duality, this expression is equal to

\[
\max_{\lambda_i \geq 0, \lambda_i' \geq 0, \sum \lambda_i = \sum \lambda_i'} \left[ \left( \sum_i \text{Tr}(\lambda_i H_i) - \sum_i \text{Tr}(\lambda_i' H_i) \right) \right],
\]

It is easy to see that this maximum is achieved by a pair of families \( \{\lambda_i\}, \{\lambda_i'\} \) satisfying \( \sum \lambda_i = \sum \lambda_i' \) and \( 2\text{Tr}(\sum \lambda_i') = 1 \).

Using the above lemma, we prove the following theorem, which characterizes Succ_{Q_0}(\( N \)).

**Theorem 3.** For any channel \( N \), and any integer \( n \geq 2 \),

\[
\text{Succ}_{Q_0}(N) = \frac{1}{2} + \max_{B_y} \left( \text{Rad} \left\{ \sum_{y \in Y} N(y|x)B_y \right\} \right),
\]

where the maximization is over all families \( \{B_y\}_{y \in Y} \) of Hermitian operators on \( \mathbb{C}^n \) satisfying \( 0 \preceq B_y \preceq I \).

**Proof.** Consider the following quantum-assisted protocol for transmitting a single bit across \( N \): Alice and Bob possess a bipartite quantum system represented by a density matrix \( \Lambda \) on
a Hilbert space $V_A \otimes V_b$. Alice wishes to transmit a message $a \in \{0,1\}$. Depending on the value of $a$, she applies one of two possible positive operator-valued measures (POVMs) $\{A_{i,a}^b\}_{i \in X}$ or $\{A_{i}^b\}_{i \in X}$ to $V_b$ and sends the result of the measurement to the channel $N$. Bob receives the output $y$ of the channel, and according to this output, applies one of a family of binary POVMs $\{B_{x,y}^b, B_{y}^0\}_{y \in Y}$ to $V_b$. The result of this output is Bob’s guess at Alice’s original message.

In order to compute the success probability for this protocol, it is not necessary to know the state $\Lambda$ or the operators $\{A_{x,a}^b\}_{x,a}$: It is only necessary to know the operators $\rho_0^a = \text{Tr}_X(A_{x,a}^b \otimes I)A_{x,a}^b$, which represent the state of Bob’s quantum system when the outcome of Alice’s measurement is $x$. These operators satisfy $\sum_x \rho_0^a = \sum_x \rho_1^a$ and $\text{Tr}(\sum_x \rho_0^a) = 1$, and, in fact, any family of operators satisfying those two constraints can be induced by an appropriately chosen state $\Lambda$ and appropriately chosen POVMs $\{A_{x,a}^b\}_{x,a} \in \mathcal{E}$. Thus, for our purposes, to specify an $(n,n)$-dimensional entanglement-assisted strategy for communicating a single bit across $N$, it suffices to specify a collection of binary POVMs $\{B_{x,y}^b, B_{y}^0\}_{y \in Y}$ on $C^n$ and a collection of positive semidefinite operators $\{\rho_n^a\}_{a \in \{0,1\}, x \in X}$ on $C^n$ satisfying

$$\sum_x \rho_0^a = \sum_x \rho_1^a \quad \text{and} \quad \text{Tr}(\sum_x \rho_0^a) = 1. \quad (5)$$

The success probability of the protocol is given by

$$\frac{1}{2} \left( \sum_{y \in Y} N(y|x) \left\{ \sum_{x \in X} \text{Tr}(\rho_0^a B_{y}^b) \right\} \right)^2 + \frac{1}{2} \sum_{y \in Y} N(y|x) \left\{ \text{Tr}(\rho_1^a B_{y}^b) \right\}^2.$$ 

Let $B_y^1 = I - B_y$. The expression above simplifies to

$$\frac{1}{2} + \frac{1}{2} \text{Tr} \left[ \sum_{x \in X} (\rho_0^a - \rho_1^a) \sum_{y \in Y} N(y|x) B_y^b \right]. \quad (6)$$

The quantity $\text{Succ}_{Q_0}(N)$ is the maximum of this expression over all $n \times n$ Hermitian operators $\{B_i\}_{i \in Y}$ satisfying $0 \leq B_i \leq I$ and all $n \times n$ positive semidefinite operators $\{\rho_n^a\}_{x \in X, a \in \{0,1\}}$ satisfying (5) above. Applying Lemma 2 yields the desired formula.

The following corollary is a stronger version of Theorem 3.

**Corollary 4.** The formula in Theorem 3 holds also when the maximum is taken only over families $\{B_i\}$ that consist of projections on $C^n$.

**Proof.** The radius function is convex in the following sense: For any families of operators $\{J_i\}_{i \in Y}$ and $\{K_i\}_{i \in Y}$, and real number $\alpha \in [0,1]$,

$$\text{Rad}(\alpha J_i + (1 - \alpha)K_i) \leq \alpha \text{Rad}(J_i) + (1 - \alpha)\text{Rad}(K_i).$$

(For, if we let $J'$ be such that the distance from $J'$ to $J_i$, is equal to $r = \text{Rad}(J_i)$, and we let $K'$ be such that the distance from $K'$ to $K_i$ is equal to $r' = \text{Rad}(K_i)$, then the distance from $\alpha J' + (1 - \alpha)K'$ to $\alpha J_i + (1 - \alpha)K_i$, is no more than $\alpha r + (1 - \alpha)r'$ by the triangle inequality.)

In particular, this convexity property implies that the radius of $\{\alpha J_i + (1 - \alpha)K_i\}$ is no more than the maximum of $\text{Rad}(J_i)$ and $\text{Rad}(K_i)$.

Since any Hermitian operator $B$ satisfying $0 \leq B \leq I$ is a convex combination of projection operators, the conclusion follows from Theorem 2.

As an example of the utility of Theorem 3, consider the channel $M$ in Fig. 2, which is defined in Ref. [2]. The input alphabet for $M$ is $\{1,2,3,4\}$, and the output alphabet is $\{1,2,3,4,5,6\}$. In Sec. III, we prove that for any $2 \times 2$ Hermitian operators $B_1, \ldots, B_6$ satisfying $0 \leq B_i \leq I$, the radius of the set $\{\sum_{y=1}^6 M(y|x)B_y | x = 1,2,3,4\}$ is no more than $\frac{1}{2} + \frac{1}{3\sqrt{2}}$. This maximum is achieved when $B_1 = 0$, $B_2 = I$, and $\{B_3, B_4\}$ and $\{B_5, B_6\}$ are two different Pauli measurements. Therefore,

$$\text{Succ}_{Q_2}(M) = \frac{2}{3} + \frac{1}{3\sqrt{2}},$$

and the protocol from Ref. [2] is optimal for two-dimensional entanglement assistance. (We note that this generalizes the paper [12], which showed the optimality of Ref. [2] within a more restricted class of protocols.)

We remark that the proof of Theorem 3 also gives a recipe for obtaining optimal protocols for a given channel $N$. Bob’s strategy is given by the optimizing operators $\{B_i\}_{i \in Y}$ and the state $\Lambda$ from the set $\{\lambda^i_{x} | x \in X\}$ and $\{\lambda^j_{y} | y \in Y\}$ that result from computing the radius of the set $\{\sum_{y \in Y} N(y|x)B_y | x \in X\}$ via the formulation in Lemma 2.

### III. Optimality for Two-Dimensional Entanglement Assistance

Let $M$ be the channel defined in Fig. 2. In this section we use Theorem 3 to calculate the quantity $\text{Succ}_{Q_2}(M)$.

First, we prove the following lemma which provides a simplified formula for $\text{Succ}_{Q_0}(N)$. For any projection operator $P$, let $P^\perp$ denote projection onto the orthogonal complement of $P$.

**Lemma 5.** For any $n \geq 1$, the quantity $\text{Succ}_{Q_0}(M)$ is equal to

$$\frac{1}{2} + \left(\frac{1}{2}\right) \max_{X,Y,Z} \text{Rad}(X + Y + Z, X + Y^\perp + Z^\perp, X^\perp + Y + Z^\perp, X^\perp + Y^\perp + Z),$$

where the maximum is taken over all projection operators $X,Y,Z$ on $C^n$.

**Proof.** For any sequence of Hermitian operators $B := (B_1, B_2, B_3, B_4, B_5, B_6)$ on $C^n$, define $F(B)$ as follows

$$F(B) := \text{Rad}(B_1 + B_3 + B_5, B_1 + B_4 + B_6, B_2 + B_3 + B_6, B_2 + B_4 + B_5).$$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1/3</td>
<td>1/3</td>
<td>0</td>
<td>0</td>
<td>1/3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
<td>0</td>
<td>1/3</td>
</tr>
</tbody>
</table>

**FIG. 2.** The channel $M$, from Ref. [2].
By the formula from Theorem 3,
\[ \text{Succ}_{Q_0}(M) = \frac{1}{2} + \left( \frac{1}{2} \right) \max_{0 \leq R_i \leq 1} F(B). \] (7)

Let
\[ m = \max_{0 \leq R_i \leq 1} F(B). \] (8)

It suffices to prove that this maximum is achieved by some six-tuple of the form \((X, X^+, Y, Y^+, Z, Z^+)\), where \(X, Y,\) and \(Z\) are projections.

As noted in the proof for Corollary 4, the radius function is convex in the sense that if \((H_1, H_2, H_3, H_4)\) and \((H'_1, H'_2, H'_3, H'_4)\) are Hermitian operators and \(\alpha \in [0, 1]\) is a real number,
\[ \text{Rad}[\alpha H_1 + (1 - \alpha)H'_1] \leq \alpha \text{Rad}[H_1] + (1 - \alpha) \text{Rad}[H'_1]. \]

It follows easily by linearity that a similar convexity property holds for \(F\): For any Hermitian operators \(B := (B_1, \ldots, B_6)\) and \(B' := (B'_1, \ldots, B'_6)\), and any \(\alpha \in [0, 1]\),
\[ F(\alpha B + (1 - \alpha)B') \leq \alpha F(B) + (1 - \alpha)F(B'). \] (9)

In particular,
\[ F(\alpha B + (1 - \alpha)B') \leq \max\{F(B), F(B')\}. \] (9)

Additionally, \(F\) is translation invariant in the following sense: For any Hermitian operators \(B_1, \ldots, B_6\), and any Hermitian operators \(K, L,\) and \(M\),
\[ F(B + M) = F(B). \] (10)

Let \(X_1, X_2, Y_1, Y_2, Z_1, Z_2\) be Hermitian operators satisfying \(0 \leq X_i, Y_i, Z_i \leq I\) such that \(F(X_1, X_2, Y_1, Y_2, Z_1, Z_2) = m\). Let \(X_+\) and \(X_-\) be a pair of positive semidefinite operators having mutual orthogonal supports which are such that
\[ X_1 - X_2 = X_+ - X_. \] (11)

Define \(Y_+, Y_-\) similarly. By property (10) above,
\[ F(X_+, X_-, Y_+, Y_-, Z_+ , Z_-) = F(X_1, X_2, Y_1, Y_2, Z_1, Z_2) = m. \] (12)

The pair \((X_+, X_-)\) can be expressed as a convex combination of pairs of projections \((P^{(i)}_+, P^{(i)}_-)\) where for each \(i\), the support of \(P^{(i)}_+\) is orthogonal to \(P^{(i)}_-\). A similar decomposition exists for \((Y_+, Y_-)\) and \((Z_+, Z_-)\). Therefore by property (9) above, there exist pairs of projections \((P_1, P_2)\), \((Q_1, Q_2)\), \((R_1, R_2)\), with each pair having mutually orthogonal supports, such that
\[ F(P_1, P_2, Q_1, Q_2, R_1, R_2) = m. \] (13)

Let \(P_3 = I - P_1 - P_2\), and define \(Q_3\) and \(R_3\) similarly. By Eq. (10),
\[ F\left( P_1 + \frac{P_3}{2} , P_2 + \frac{P_3}{2} , Q_1 + \frac{Q_3}{2} , Q_2 + \frac{Q_3}{2} , R_1 + \frac{R_3}{2} , R_2 + \frac{R_3}{2} \right) = m. \] (14)

The six-tuple on the left-hand side of the equation above is a convex combination of the six-tuples
\[ (P_1 + P_3, P_2, Q_1 + Q_3, R_1 + R_3) \quad \text{and} \quad (P_1 + P_3, Q_1, Q_2 + Q_3, R_1 + R_3). \]

By relation (9), at least one of these six-tuples must achieve the maximum \(m\). This completes the proof. \(\blacksquare\)

**Lemma 6.** For any projection operators \(X, Y, Z\) on the two-dimensional vector space \(\mathbb{C}^2\), the radius of the set
\[ \{X + Y + Z, X + Y^+, Z^+, X^+ + Y + Z^+, X^+ + Y^+ + Z\} \]
(15)
is less than or equal to \(\frac{1}{2} + \frac{1}{\sqrt{2}}\).

**Proof.** Case 1: The matrices \(X, Y, Z\) are all scalar matrices. In this case, each of \(X, Y,\) and \(Z\) is equal to either 0 or \(\mathbb{I}\). This case is trivial, since the radius of the set \([\mathbb{I}, \mathbb{I}]\) is 1, and the radius of the set \([2\mathbb{I}, 0]\) is 1.

Case 2: Two of the matrices \(X, Y, Z\) are scalar matrices and one is a nonscalar. We may assume without loss of generality that \(X\) is the nonscalar matrix. Then the set (15) is equal to either
\[ \{0, X + 2\mathbb{I}\} \]
(16)
or
\[ \{X, X + 2\mathbb{I}, \mathbb{I}\}. \] (17)

In the former case, the operator-norm distance from the operator \(\mathbb{I}\) to the set \([0, X + 2\mathbb{I}]\) is 1. In the latter case, the operator-norm distance from the operator \(X + \mathbb{I}\) to the set \([X, X + 2\mathbb{I}, \mathbb{I}]\) is 1. The desired result follows.

Case 3: Exactly one of the matrices \(X, Y, Z\) is a scalar matrix. We may assume that \(X\) and \(Y\) are nonscalar matrices and \(Z\) is scalar. Also, by replacing \((X, Y, Z)\) with \((X^+, Y, Z^+)\) if necessary, we may assume that \(Z = \mathbb{I}\).

Let \(X = |x\rangle\langle x|\) and \(Y = |y\rangle\langle y|\) where \(x, y \in \mathbb{C}^2\) are unit vectors, and let \(\theta = \arccos(|\langle x|y\rangle|)\). Both of the operators
\[ X + Y + \mathbb{I}, X^+ + Y^+ + \mathbb{I} \]
(18)
have eigenvalues \([2 + \cos \theta, 2 - \cos \theta]\), and both of the operators
\[ X + Y^+, X^+ + Y \]
(19)
have eigenvalues \([1 + \sin \theta, 1 - \sin \theta]\). If we let
\[ C = \frac{3}{2} + \frac{\cos \theta - \sin \theta}{2} \mathbb{I}, \] (20)
then the operator-norm distance from \(C\) to each of the elements of (15) is \(\frac{1}{2} + \frac{\cos \theta + \sin \theta}{2} \leq \frac{1}{2} + \frac{1}{\sqrt{2}}\).

Case 4: Each of \(X, Y, Z\) is a nonscalar matrix. As in case \(3\), let \(X = |x\rangle\langle x|\) and \(Y = |y\rangle\langle y|\) and let \(\theta = \arccos(|\langle x|y\rangle|)\). Let
\[ C = \mathbb{I} + \frac{1}{2} \cos \theta - \sin \theta \mathbb{I} Z + \left( \frac{1}{2} - \cos \theta - \sin \theta \right) Z^+. \]
(21)
Then, the operator norm of the difference
\begin{equation}
(X + Y + Z) - C = (X + Y) - \left( \frac{3}{2} + \frac{\cos \theta - \sin \theta}{2} \right) I
\end{equation}
(22)
is \frac{1}{2} + \frac{\cos \theta + \sin \theta}{2}, which is less than or equal to \( \frac{1}{2} + \frac{1}{\sqrt{2}} \). A similar calculation shows that the distance from \( C \) to each of the other three elements of set (15) is equal to \( \frac{1}{2} + \frac{\cos \theta + \sin \theta}{2} \).
This completes the proof.

For any angle \( \theta \in \mathbb{R} \), let \( P_{0} : \mathbb{C}^{2} \to \mathbb{C}^{2} \) denote projection onto the unit vector \( \cos(\theta)[0] + \sin(\theta)[1] \). Consider the set
\[ \{ P_{0} + P_{\pi/4} + I, P_{0} + P_{3\pi/4}, P_{\pi/2} + P_{\pi/4}, P_{\pi/2} + P_{3\pi/4} + I \}. \]
(23)
A direct calculation shows that the distance from the operator \( (\frac{3}{2} I) \) to set (23) is \( \frac{1}{2} + \frac{1}{\sqrt{2}} \). The next lemma asserts that this quantity is in fact the radius of set (23).

\textit{Lemma 7.} The radius of the set (23) is \( \frac{1}{2} + \frac{1}{\sqrt{2}} \).

\textit{Proof.} For any Hermitian operator \( H : \mathbb{C}^{2} \to \mathbb{C}^{2} \), let us write \( \overline{H} \) to denote the trace-zero operator \( H - (\text{Tr}(H))I/2 \). In the proof that follows, we make use of the following fact: For any two Hermitian operators \( Q, R : \mathbb{C}^{2} \to \mathbb{C}^{2} \),
\[ \| Q - R \| = | \text{Tr}(Q) - \text{Tr}(R) | + \| Q - R \|. \]
(24)
Suppose, for the sake of contradiction, that there exists a Hermitian operator \( Z \) whose distance from each of the elements of set (23) is strictly less than \( \frac{1}{2} + \frac{1}{\sqrt{2}} \). Then,
\begin{align*}
2 \left( \frac{1}{2} + \frac{1}{\sqrt{2}} \right) & > (\| P_{0} + P_{3\pi/4} - Z \| + \| P_{\pi/2} + P_{\pi/4} - Z \|) \\
& = (\| P_{0} + P_{3\pi/4} - I \| - Z\| + \| P_{\pi/2} + P_{\pi/4} - I \| - Z\|) \\
& + 2 | \text{Tr}(Z) | \\
& \geq (\| P_{0} + P_{3\pi/4} - (P_{\pi/2} + P_{\pi/4}) \| + 2 | \text{Tr}(Z) |) \\
& = \sqrt{2} + 2 | \text{Tr}(Z) |.
\end{align*}
Therefore, \( \text{Tr}(Z) < \frac{5}{2} \). Similarly,
\begin{align*}
2 \left( \frac{1}{2} + \frac{1}{\sqrt{2}} \right) & > (\| P_{0} + P_{\pi/4} + I - Z\| + \| P_{\pi/2} + P_{3\pi/4} - I \|) \\
& = (\| P_{0} + P_{\pi/4} + I \| - Z\| + \| P_{\pi/2} + P_{3\pi/4} - I \| - Z\|) \\
& + 2 | \text{Tr}(Z) | \\
& \geq (\| P_{0} + P_{\pi/4} - (P_{\pi/2} + P_{3\pi/4}) \| + 2 | \text{Tr}(Z) |) \\
& = \sqrt{2} + 2 | \text{Tr}(Z) |,
\end{align*}
which implies \( \text{Tr}(Z) > \frac{5}{2} \). This is a contradiction. •

By combining Lemmas 5–7, we have the following theorem.

\textit{Theorem 8.} The channel \( M \) satisfies
\[ \text{Succ}_{\text{OPT}}(M) = \frac{2}{3} + \frac{1}{3\sqrt{2}}. \]
(25)

\section{IV. Nonsignaling Devices}

Our next goal is to prove more explicit upper bounds on the limits of quantum assistance. We begin by considering assistance by nonsignaling correlations as an intermediate step. The next proposition asserts a formula (based on Ref. [5]) for the optimal nonsignaling-assisted success probability of a channel. For any finite set of vectors \( S \subseteq \mathbb{R}^{\mathcal{Y}} \), let \( \text{Rad}_{1}(S) \) denote the radius of \( S \) under the 1-norm.

\textit{Proposition 9.} Let \( N \) be a classical channel, and for each \( x \in \mathcal{X} \), let \( n_{x} = \{ N(y | x) \}_{y \in \mathcal{Y}} \in \mathcal{Y}^{\mathcal{Y}} \). Then,
\[ \text{Succ}_{\text{C}}(N) = \frac{1}{2} + \frac{1}{2} \text{Rad}_{1} \{ n_{x} : x \in \mathcal{X} \}. \]
(26)

Note that in the above formula, we take the radius of \( \{ n_{x} \} \) as a subset of \( \mathbb{R}^{\mathcal{Y}} \), not as a subset of the set of probability distributions on \( \mathcal{Y} \).

\textit{Proof of Proposition 9.} By Proposition 14 from Ref. [5],
\[ \text{Succ}_{\text{NS}}(N) = 1 - \max_{e \in \mathbb{R}^{\mathcal{Y}}} \min_{y \in \mathcal{Y}} (\text{Succ}(N(e | y)) - c_{y}/2). \]
Using the easily proved fact that \( \| u - v \|_{1} = \sum_{i} u_{i} + \sum_{i} v_{i} - 2 \sum_{i} \min(u_{i}, v_{i}) \), the above formula simplifies to
\[ \text{Succ}_{\text{NS}}(N) = \min_{e \in \mathbb{R}^{\mathcal{Y}}} \max_{y \in \mathcal{Y}} \left( \frac{1}{2} + \frac{1}{2} \| c - n_{x} \|_{1} \right), \]
which implies formula (26) above. •

Formula (26) allows us to relate the quantity \( \text{Succ}_{\text{NS}}(N) \) to the quantity \( \text{Succ}(N) \).

\textit{Theorem 10.} Let \( N \) be a classical channel, and let \( r = | \mathcal{X} | \) denote the size of the input alphabet of \( N \). Then,
\[ \text{Succ}_{\text{NS}}(N) - \frac{1}{2} \leq \left( \frac{2 - \frac{2}{r}}{r} \right) \left[ \text{Succ}(N) - \frac{1}{2} \right]. \]
(27)

\textit{Proof.} Let \( \{ n_{x} \} \) be the vectors defined in Proposition 9. By Proposition 1, the unassisted one-shot success probability can be expressed in terms of these vectors as in relation (2). A triangle-inequality argument shows that the distance from the mean vector \( \left( \sum_{x} n_{x} / r \right) \) to the set \( \{ n_{x} \} \) cannot exceed \( (1 - \frac{1}{r}) \text{Diam}_{1}(n_{x}) \). Therefore, \( \text{Rad}_{1}(n_{x}) \leq (1 - \frac{1}{r}) \text{Diam}_{1}(n_{x}) \), which implies the desired result. •

Theorem 10 is the best possible in the sense that there are channels where equality is achieved in relation (27).
Consider the following example, which is a generalization of the channel \( M \) from Fig. 2. Let \( s \) be a positive integer. For any \( i \in \{ 0, 1, 2, \ldots, 2^{s-1} \} \), let \( b_{i} \in \mathbb{F}_{2}^{r} \) denote the binary representation of \( i \). Define a channel \( T \) as follows. The input alphabet of \( T \) is \( \{ 0, 1, 2, \ldots, 2^{s-1} \} \), and the output alphabet is \( \{ 1, 2, \ldots, 2^{s-1} \} \times \{ 0, 1 \} \). On given input \( i \), the channel chooses an element \( j \in \{ 1, \ldots, 2^{s-1} \} \) uniformly at random and outputs the pair \( (j, b_{i} b_{j}) \) (where \( b_{i} b_{j} \) denotes the inner product of \( b_{i} \) and \( b_{j} \mod 2 \)).

For any \( i \in \{ 0, 1, 2, \ldots, 2^{s-1} \} \), let \( \ell_{i} \in \mathbb{R}^{2(2^{s}-1)} \) denote the probability vector which expresses the output of \( T \) on input \( i \). It is easy to see that the diameter of \( \{ \ell_{i} \} \) is \( 2^{s}/(2^{2^{s}}-1) \), and thus \( \text{Succ}(T) = \frac{1}{2} + 2^{-2^{s}}/(2^{2^{s}}-1) \). On the other hand, the radius of \( \{ \ell_{i} \} \) is 1, as can be seen from the following calculation. For
any $c \in \mathbb{R}^{2^{2^k-1}}$, 

$$\max_{0 \leq |c| \leq 2^k-1} \|\ell_i - c\|_1 \geq 2^{-k} \sum_{i=0}^{2^k-1} \|\ell_i - c\|_1$$

$$\geq 2^{-t} \sum_{1 \leq j \leq 2^t - 1} \sum_{t \in [0,1]} \{2^{t-1} |c_{jt} - (2^{t-1} - 1)| + 2^{t-1} |c_{jt} - 0|\}$$

Thus $\text{Succ}_{\mathbb{NS}}(T) = 1$. (Indeed, a perfect communication protocol for $T$ exists—see Appendix.) Therefore equality occurs in Theorem 10:

$$\text{Succ}_{\mathbb{NS}}(T) - \frac{1}{2} = \left(2 - \frac{2}{m}\right) \left[\text{Succ}(T) - \frac{1}{2}\right].$$

The following modified version of Theorem 10 is useful in our analysis of entanglement assistance.

Theorem 11. Let $N$ be a classical channel, and let $D$ be a nonsignaling correlation arising from a two-part device $(D_A, D_B)$. Let $m$ denote the size of the output alphabet of $D_A$. Then, 

$$\text{Succ}(N, D) - \frac{1}{2} \leq \left(2 - \frac{2}{m}\right) \left[\text{Succ}(N) - \frac{1}{2}\right]. \quad (28)$$

Proof. A protocol for communicating a single bit $a$ using $N$ and $D$ proceeds as follows. Alice uses $a$ to choose an input to $D_A$, and then uses $a$ and the output of $D_A$ to choose an input to $N$. Bob uses the output of $N$ to choose an input to $D_B$, and then uses the outputs of $N$ and $D_B$ together to guess the bit $a$.

The optimal success probability $\text{Succ}(N, D)$ can be achieved by a deterministic protocol (i.e., a protocol in which Alice and Bob make their choices according to deterministic functions). As there are only $2m$ possible inputs that Alice could make to $N$ in a deterministic protocol, the success probability of such a protocol is bounded by $[2 - 2/(2m)]\text{Succ}(N)$ by Theorem 10.

V. BINARY QUANTUM DEVICES

In this section, we use our bounds for nonsignaling devices from Sec. IV to obtain bounds for assistance by binary quantum devices.

A two-part device $D$ is local deterministic if the output of each part is a deterministic function of its input. A nonsignaling correlation is local if it is a convex combination of local-deterministic correlations. We define the local fraction of a nonsignaling correlation, a concept which is used in Refs. [13, 14].

Definition 12. Let $D$ be a nonsignaling correlation. The local fraction of $D$, denoted $\text{loc}(D)$, is the largest real number $\alpha \in [0,1]$ such that there exists a decomposition $D = \alpha L + (1 - \alpha)F$, where $L$ is a local correlation and $F$ is a nonsignaling correlation.

For any classical channel $N$, it is easy to see that when a decomposition $D = \alpha L + (1 - \alpha)F$ exists with $L$ local and $F$ nonsignaling,

$$\text{Succ}(N, D) \leq \alpha \text{Succ}(N, L) + (1 - \alpha)\text{Succ}(N, F) \leq \alpha \text{Succ}(N) + (1 - \alpha)\text{Succ}_{\mathbb{NS}}(N).$$

This implies the following stronger version of Theorem 11.

Theorem 13. Let $N$ be a channel, and let $D$ be a nonsignaling correlation arising from a two-part device $(D_A, D_B)$. Let $m$ denote the size of the output alphabet of $D_A$. Then

$$\frac{\text{Succ}(N, D) - \frac{1}{2}}{\text{Succ}(N) - \frac{1}{2}} \leq 1 + \left(1 - \frac{1}{m}\right) [1 - \text{loc}(D)].$$

Thus, to obtain improved upper bounds on $\text{Succ}(N, D)$ for quantum correlations $D$, it suffices to find lower bounds on the local fractions of quantum correlations. We use facts about the geometry of quantum and nonsignaling correlations [15] to prove the following bound for binary quantum correlations.

Proposition 14. Let $D$ be a binary quantum correlation. Then $\text{loc}(D) \geq 2 - \sqrt{2}$.

Proof. For any binary nonsignaling correlation $G$, let

$$f_i(G) = \sum_{a, x, y \in \{0,1\}} (-1)^{a \otimes b \otimes (y \wedge y)} G(xb|ay).$$

This is the function which defines the Clauser-Horne-Shimony Holt (CHSH) inequality [16]. Let $f_2, f_3,$ and $f_4$ be the functions defined by the same expression with $a \wedge y$ replaced by $\neg a \wedge y$, $a \wedge \neg y$, and $\neg a \wedge \neg y$, respectively.

We note the following facts. (See Ref. [15].)

1. A nonsignaling correlation $G$ is local if and only if $-2 \leq f_i(G) \leq 2$ for $i = 1, 2, 3, 4$.

2. If $G$ is a quantum correlation, then for $i = 1, 2, 3, 4$,

$$-2\sqrt{2} \leq f_i(G) \leq 2\sqrt{2}.$$  

3. There are eight nonsignaling correlations $\{P^+_i\}_{i=1}^4$ and $\{P^-_i\}_{i=1}^4$, satisfying

$$f_j(P^\pm_i) = \begin{cases} \pm 4 & \text{if } j = i \\ 0 & \text{otherwise.} \end{cases}$$

These are the Popescu-Rohrlich (PR) boxes.

4. Every nonsignaling correlation is a convex combination of local correlations and the eight PR boxes. Further, for any two distinct PR boxes $P$ and $P'$, the correlation $(P + P')/2$ is local.

From the second part of item 4, it follows that any convex combination of local boxes and PR boxes can be simplified into an expression of the form $\alpha L + (1 - \alpha)Q$, where $L$ is local, $Q$ is a PR box, and $\alpha \in [0,1]$. Any nonsignaling correlation can thus be expressed as a convex combination of a local correlation and a single PR box.

Let $D = \alpha L + (1 - \alpha)Q$, where $L$ is local and $Q$ is a PR box. First suppose that $Q = P^+_j$. Let $L_\beta = (\alpha L + (\beta - \alpha)P^+_j)/\beta$, for any $\beta \in [\alpha,1]$. Then $L_\beta$ is local whenever $f_j(L_\beta) \leq 2$. If $f_j(L_\beta) < 2$, then $L_1(=D)$ is local, and the proposition follows easily. Otherwise, there is a value $\beta \in [\alpha,1]$ such that $f_j(L_\beta) = 2$. We have $D = \beta L_\beta + (1 - \beta)P^+_j$. The quantity $\beta$ must be at least $2 - \sqrt{2}$, since otherwise relation (2) would be violated. Therefore $\text{loc}(D) \geq 2 - \sqrt{2}$.

A similar argument completes the proof in the case where $Q = P^-_j$. \qed
Theorem 13 and Proposition 14 yield the following corollary.

**Corollary 15.** For any classical channel $N$,

$$\frac{\text{Succ}_N^\alpha(N) - \frac{1}{2}}{\text{Succ}_N^\alpha(N) - \frac{1}{2}} \leq 1 + \frac{1}{\sqrt{2}}. \quad (29)$$

We note that equality is achieved in relation (29) by the protocol from Ref. [2].

**VI. NONBINARY QUANTUM DEVICES**

Finally, we extend our results from Sec. II to the case where Alice’s message is drawn from an alphabet of finite size. In this section, we prove an extension of Theorem 3 and provide an exact formula for the one-shot success probability in this new setting.

We direct the reader’s attention also to Ref. [17], which proves a related result: A bound, under somewhat different assumptions, for the one-shot communication capacity of an entanglement-assisted quantum channel.

If $Z = \{Z_j\}_{j \in J}$ and $Y = \{Y_j\}_{j \in J}$ are two families of Hermitian operators on the same Hilbert space $H$ with the same index set $J$, and if $c$ is a real number, let us write $cZ$ and $Y + Z$ for the sets $\{cZ_j\}$ and $\{Y_j + Z_j\}$, respectively. If we write $Z \geq Y$, we mean that $Z_j \geq Y_j$ for all $j \in J$.

We now introduce a quantity which measures the dissimilarity between a collection of POVMs.

**Definition 16.** Suppose that $M$, $\mathcal{X}$ are finite sets and that $H = \mathbb{C}^n$. Suppose that for each $x \in \mathcal{X}$, the set $C_x = \{C_x^m\}_{m \in M}$ is a POVM on $H$. Then, define $\Lambda(\{C_x^m\})$ by

$$\Lambda(\{C_x^m\}) = \min_{a, \mathcal{V} \succeq C_x} \max_{y} a^\mathcal{V} \sum_{x} \text{Tr}(C_x^m \mathcal{V}). \quad (30)$$

where the minimum is taken over all real numbers $a$ and POVMs $\mathcal{V} = \{V^m\}_m$ satisfying the constraints $a \mathcal{V} \succeq C_x$ for all $x \in \mathcal{X}$.

The reader can note that this definition bears some similarity to the definition of quantum conditional min-entropy. (See Sec. 3.1 of Ref. [18].) The following lemma (analog of Lemma 2) gives an alternate expression for $\Lambda$.

**Lemma 17.** Suppose that $M$, $\mathcal{X}$ are finite sets, and that for each $x \in \mathcal{X}$, the set $C_x = \{C_x^m\}_{m \in M}$ is a POVM indexed by $M$. Then,

$$\Lambda(\{C_x^m\}) = \max_{\mathcal{V}} \sum_{x} \text{Tr}(\rho^m \mathcal{V}), \quad (31)$$

where the maximum is taken over all density operators $\rho$ and decompositions $\rho = \sum_m \rho^m$, with $\rho^m \succeq 0$.

**Proof.** We can rewrite the definition of $\Lambda$ as a semidefinite program by setting $W = a \mathcal{V}$.

$$\Lambda(\{C_x^m\}) = \min_{a, \mathcal{V}} \max_{\mathcal{W} \succeq C_x} a, \quad (32)$$

$$\sum_{x} \text{Tr}(\rho^m \mathcal{V}), \quad (33)$$

By semidefinite programming duality,

$$\Lambda(\{C_x^m\}) = \max_{\mathcal{V}} \sum_{x} \text{Tr}(\rho^m \mathcal{V}), \quad (34)$$

which implies the desired formula.

Suppose that Alice and Bob share a classical channel $N$ which has input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Suppose that Alice possesses a message $m$ which is randomly chosen from a finite set $\mathcal{M}$ of size $q$ according to a uniform distribution. Suppose that Alice wishes to send this message to Bob. Let

$$\text{Succ}_N^q(N) \quad (35)$$

denote the maximum probability of success that can be achieved with a single use of the channel $N$ and with a shared bipartite quantum system of dimension $(n,n)$.

**Theorem 18.** Let $N$ be a classical channel with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. Let $\mathcal{M}$ be a finite set of size $q \geq 2$, and let $n$ be a positive integer. Then,

$$\text{Succ}_N^q(N) = \frac{\Lambda(\{C_x^m\})}{q}, \quad (36)$$

where the maximum is taken over all families of POVMs $\{C_x^m\}_{m \in \mathcal{M}}$ in $\mathbb{C}^n$ that can be obtained from a family of POVMs $\{B^m\}_{m \in \mathcal{M}}$ via the linear transformation $C_x^m = \sum_y N_y^m B_y^m$.

**Proof.** Suppose that Alice possesses a randomly chosen message $m$ from the set $\mathcal{M}$. An $(n,n)$-dimensional quantum-assisted protocol for communicating $m$ to Bob proceeds like so. Alice and Bob possess a bipartite $(n,n)$-dimensional quantum system $(Q_A,Q_B)$. Let $\rho$ denote the initial state of Bob’s subsystem $Q_B$. According to the value of message $m$, Alice performs a measurement on her subsystem and obtains an outcome $X$ which is a random variable over the input alphabet $\mathcal{X}$. The state of $(X,Q_B)$ at this point can be represented as

$$\sum_{x \in \mathcal{X}} |x\rangle \langle x| \otimes \rho^m_x, \quad (37)$$

where the operators $\rho^m_x$ satisfy $\rho^m_x \succeq 0$ and $\sum_x \rho^m_x = \rho$ for all $m \in \mathcal{M}$. Alice uses $X$ as input to the channel and Bob receives an output, $Y$. The joint state of $(X,Q_B)$ is now

$$\sum_{y \in \mathcal{Y}} |y\rangle \langle y| \otimes \sum_{x \in \mathcal{X}} N_y^m \rho^m_x, \quad (38)$$

Bob applies a joint measurement to $(Y,Q_B)$ in order to guess the value of $m$. Bob’s measurement can be expressed as a family of POVMs $\{B^m_y\}_{m \in \mathcal{M}}$. The probability that this protocol succeeds is

$$\frac{1}{q} \sum_m \text{Tr} \left[ \sum_{x} N_y^m B_y^m \left( \sum_{x} N_y^x \rho^x_m \right) \right]. \quad (39)$$

$$= \frac{1}{q} \sum_{m,x} \text{Tr} \left[ \rho^m_x \left( \sum_{y} N_y^m B_y^m \right) \right]. \quad (40)$$

We can obtain a formula for $\text{Succ}_N^q(N)$ by maximizing expression (40) over all density operators $\rho$ and decompositions $\rho = \sum_x \rho^m_x$, and over all families of POVMs $\{B^m_y\}_{y \in \mathcal{Y}}$. 


By Lemma 17, this yields
\[ \text{Succ}_{Q|a_0}^q(N) = \max_{B_y > 0, \sum B_y = 1} \frac{\Lambda\{\sum_y n_x B_y\}}{q}, \]
as desired.

**VII. CONCLUSION AND OPEN PROBLEMS**

We have given a formula for the \( n \)-dimensional entanglement-assisted one-shot success probability of a classical channel and have shown its utility by using it to show that the protocol in Ref. [2] is optimal for two-dimensional entanglement assistance. We derived a more explicit bound on the advantage gained by binary quantum correlations (which is an equality in the case of Ref. [2]). Along the way, we established a bound on the advantage from nonsignaling assistance and provided an example where equality is achieved.

Two major problems remain open. The first is to develop methods for determining \( \text{Succ}_{Q|a_0}^q(N) \) for explicit channels \( N \) without the constraint on the entanglement dimension. In particular, is the entanglement assistance provided by the protocol in Ref. [2] optimal for arbitrarily large dimensions? Our Theorem 3 and its application in Sec. III may provide a foundation for developing such methods.

The second is to determine the best possible quantum assistance among all channels. By Proposition 1 and Theorem 3, this is the same question as finding the infimum \( a_Q \) of constants \( c \) such that for all distributions \( n_x \), and all operators \( B_x, 0 \leq B_x \leq 1 \),
\[ \text{Rad}\{\sum_x n_x B_x\} \leq c. \]
Clearly \( a_Q \leq 2 \). We conjecture that \( a_Q < 2 \).

**Conjecture 19.** There exists a constant \( c < 2 \) such that for any channel \( N \),
\[ \frac{\text{Succ}(N) - \frac{1}{4}}{\text{Succ}_Q(N) - \frac{1}{2}} \leq c. \]

It is tempting to conjecture that the quantum assistance provided in Ref. [2] is the best possible for all channels, given that the protocol is based on the best quantum approximation of the extremal nonsignaling PR box. Some generalizations of Proposition 14 would thus be helpful for studying this problem.

Finally, an even more challenging direction is to ask similar questions for the larger alphabet size, in particular to explore the formula in Theorem 18 and to see what consequences it has for communication with multiple uses of a channel. While we know that entanglement does not increase the capacity of the classical channel, perhaps it would make the error tend to zero much faster (i.e., increasing the *error exponent*).

**ACKNOWLEDGMENTS**

The authors thank Vincent Russo for his help with the preparation and editing of this paper. We also thank Shmuel Friedland, Aubrey da Cunha, Xiaodi Wu, and Kim Winick for many useful discussions and the anonymous reviewers for helpful comments. This research was supported in part by the National Basic Research Program of China under Awards No. 2011CBA00300 and No. 2011CBA00301 and the NSF of the United States under Awards No. 1017335 and No. 1216729.

**APPENDIX: AN EXAMPLE OF OPTIMAL NONSIGNALING ASSISTANCE**

In this appendix we give more details about the example from Sec. IV. Let \( m \) be a positive integer. Let
\[ Z = F_2^m, \]
\[ \mathcal{W} = (F_2^m \setminus \{0\}) \times F_2. \]

Let \( K \) be a channel defined as follows:
1. The input alphabet of \( K \) is \( Z \), and the output alphabet of \( K \) is \( \mathcal{W} \).
2. For any given input \( v \in F_2^m \), the output of \( K \) is uniformly distributed over the set
\[ \{ (w, w \cdot v) \mid w \in F_2^m \setminus \{0\} \}. \]

Here, \( w \cdot v \in F_2 \) denotes the inner product of \( w \) and \( v \).

Let \( (E_1, E_2) \) be a two part input-output device defined as follows. (See Fig. 3.)

1. The input alphabet for \( E_1 \) is \( F_2 \), and the output alphabet for \( E_1 \) is \( Z \).
2. The input alphabet for \( E_2 \) is \( \mathcal{W} \), and the output alphabet for \( E_2 \) is \( F_2 \).
3. If the inputs to \( E_1 \) and \( E_2 \) are \( a \in \{0, 1\} \) and \((w, r) \in (F_2^m \setminus \{0\}) \times F_2 \), then the output of \( E_1 \) is uniformly distributed over all vectors \( a = (a_1, a_2, \ldots, a_m) \) that satisfy \( a_1 = a \), and the output of \( E_2 \) is \( a \oplus r \oplus (w \cdot a) \).

It can be checked that the correlation \( E \) arising from \((E_1, E_2)\) is nonsignaling. Additionally, one can see (by substitution) that using \( E \) to assist \( K \) yields a perfect transmission of a single bit. (See Fig. 4.)

Now, let us calculate the quantity \( \text{Succ}(K) \). For any two distinct vectors \( x_0, x_1 \in F_2^m \), the probability that a randomly chosen vector \( w \in F_2^m \setminus \{0\} \) will satisfy \( w \cdot x_0 \neq w \cdot x_1 \) is equal to \( 2^{m-1}/(2^m - 1) \). This fact has the following consequence: If Alice employs the deterministic encoding strategy \([0 \mapsto x_0, 1 \mapsto x_1]\) to send a single bit, then the optimal

![FIG. 4. A perfect communication protocol.](attachment:image.png)
probability with which Bob can decode is
\[
\left[ \frac{2^m - 1}{2^m - 1} \right] (1) + \left[ \frac{2^m - 1}{2^m - 1} \right] \left( \frac{1}{2} \right) \tag{A4}
\]
\[
= \frac{2^m + 2^{m-1} - 1}{2^{m+1} - 2}. \tag{A5}
\]
The unassisted success probability \( \text{Succ}(K) \) is equal to quantity (A5), while \( \text{Succ}_{\text{NS}}(K) \) is equal to 1. Therefore, the bound from Theorem 10 is an equality in this case:
\[
\text{Succ}_{\text{NS}}(K) \leq \frac{1}{2} + \left( 2 - \frac{2}{2^m} \right) \left[ \text{Succ}(K) - \frac{1}{2} \right]
\]
\[
= \frac{1}{2} + 2 \left( \frac{2^m - 1}{2^m} \right) \left( \frac{2^{m-1}}{2^{m+1} - 2} \right)
\]
\[
= 1.
\]