# Limits on the ability of quantum states to convey classical messages *

**Ashwin Nayak** † **Julia Salzman** ‡

January, 2006

**Abstract**

We revisit the problem of conveying classical messages by transmitting quantum states, and derive new, optimal bounds on the number of quantum bits required for this task. Much of the previous work on this problem, and on other communication tasks in the setting of bounded error entanglement-assisted communication, is based on sophisticated information theoretic arguments. Our results are derived from first principles, using a simple linear algebraic technique. A direct consequence is a tight lower bound for the Inner Product function that has found applications to privacy amplification in quantum key distribution protocols.

## 1 Introduction

The problem of characterising the "information capacity" of quantum states is of fundamental importance in the study of communication using quantum states. Since the classical description of a general superposition over $n$ quantum bits has size exponential in $n$, it is natural to expect that it be possible to encode a large amount of classical information into substantially fewer quantum bits. However, the information encoded in a quantum state can be accessed only *indirectly*, through the process of a measurement. Measurements typically disturb the state being observed, and it is seldom possible to infer the entire description of a quantum superposition (and thus recover an exponential amount of information) from measurements on the quantum system. A fundamental theorem due to Holevo [17] formalises this limitation, and quantifies the information content of a finite quantum system. Suppose $X$ is a random variable (say over $n$-bit messages), and let $Q$ be an encoding of $X$ into $m$ quantum bits. If $Y$ is any random variable obtained by performing a measurement on the encoding $Q$, then, by the Holevo theorem, the mutual information $I(X : Y)$ between $X$ and $Y$ is at most $m$.

*To appear in *Journal of the ACM*.  Preliminary versions of the results presented here were reported in Refs. [28] and [29].

†Institute for Quantum Computing, University of Waterloo, and Perimeter Institute for Theoretical Physics. Address: Department of Combinatorics and Optimization, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada. Email: `anayak@math.uwaterloo.ca` . Research supported in part by NSERC, CIAR, CFI, OIT (Canada). Much of this work was done while this author was at University of California, Berkeley, and later at California Institute of Technology.

‡Stanford University. Address: Department of Statistics, Sequoia Hall, 390 Serra Mall, Stanford University, Stanford, CA 94305-4065, USA. Email: `horence@stat.stanford.edu` . A part of this work was done while this author was in Mathematics Department, Princeton University, and was visiting California Institute of Technology on a Summer Undergraduate Research Fellowship.

The Holevo bound is based on sophisticated information theoretic properties of quantum states, in particular, on the *strong sub-additivity* of von Neumann entropy [30, Section 12.1.1]). Using further such properties, the bound was extended by Cleve, van Dam, Nielsen, and Tapp [13] to *interactive* communication protocols as well.

In this article, we revisit the problem of conveying classical bits over quantum channels. We explain the limitation of quantum communication highlighted by the Holevo bound in much simpler terms. In the process, we derive a new, optimal bound on the number of quantum bits required for this task.

**Theorem 1.1** *Suppose one party, Alice, wishes to convey $n$ bits to the other, Bob, by communicating over a quantum channel. In any protocol, possibly two-way, in which for any $x \in \{0,1\}^n$ the probability that Bob correctly recovers $x$ is at least $p \in (0,1]$, the total number of qubits $m$ exchanged by the two parties over all the rounds of communication is at least $n - \log \frac{1}{p}$.*

An application of the extended Holevo bound, along with Fano's inequality [14, Section 2.11], would result in the weaker bound $m \geq pn - H(p)$.

**Remark:** The error in the decoding of the classical message referred to above arises from the probabilistic nature of the measurement process, rather than any noise introduced into the quantum states during the communication. Indeed, in this paper, we restrict ourselves to the study of a noiseless quantum channel. ■

Central to the proof of Theorem 1.1 is the idea of bounding the probability of correct decoding, also known as the *fidelity*, when a random variable $X$ is transmitted over a quantum channel using $m$ quantum bits. We give a tight bound on this decoding probability by a direct argument which allows us to infer lower bounds for $m$ without appealing to Holevo's theorem.

An additional resource that may be available to parties communicating over a quantum channel is "shared entanglement": the two parties may be given some number of quantum bits jointly prepared in a fixed superposition, prior to communicating with each other. For example, they may jointly hold some number of EPR pairs, which consist of pairs of qubits prepared in the maximally entangled state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$. The quantum communication channel is then said to be "entanglement-assisted."

On the one hand, shared randomness does not help in the transmission of information from one party to another (cf. Appendix A), or significantly reduce the classical complexity of computing functions vis-a-vis "private-coin" protocols [24, Section 3.3]. On the other hand, prior entanglement leads to startling phenomena such as "quantum teleportation" [6] and "superdense coding" [8]. In particular, superdense coding allows us to transmit $n$ classical bits with perfect fidelity by sending only $n/2$ quantum bits. The problem of characterising the power of prior entanglement has baffled researchers (see, e.g., Refs. [11, 20]) especially in the setting of bounded-error protocols. It is open whether it leads to more than a factor of two savings (using superdense coding) or more than an additive $O(\log n)$ savings (when used to create shared randomness). Few lower bounds are known for communication problems in this setting [13, 27, 19, 22, 34], and most are derived using intricate information-theoretic techniques.

We again focus on the problem of transmitting classical bits from one party to another, perhaps the most basic problem in the setting of communication over an entanglement-assisted quantum channel. We derive an optimal bound on the number of quantum bits required for this task, for any given probability of error.

**Theorem 1.2** *Suppose Alice wishes to convey $n$ bits to Bob by communicating over an entanglement-assisted quantum channel. For any choice of the shared entangled state, and any protocol using this prior*

*entanglement, such that for any $x \in \{0,1\}^n$ the probability that Bob correctly recovers $x$ is at least $p \in (0,1]$, the total number of qubits $m_A$ sent by Alice to Bob, over all the rounds of communication, is at least $\frac{1}{2}(n - \log \frac{1}{p})$, independent of the number of qubits sent by Bob to Alice.*

This improves over the lower bound of $m_A \geq \frac{1}{2}(pn - H(p))$ implied by the extended Holevo bound of Cleve *et al.* [13, Theorem 2] mentioned above.

Lower bounds for bounded-error communication using prior entanglement prior to this work are based on complex information-theoretic arguments. (Razborov [34] has since devised a powerful proof technique that continues to work in the presence of arbitrary prior entanglement.) Theorem 1.2 is derived from first principles, using a simple linear algebraic technique that builds upon the proof of Theorem 1.1.

In order to prove Theorem 1.2, we give a new characterisation of the joint state at the end of a quantum protocol that complements the characterisation due to Yao [36]. It clarifies the role of shared entanglement in communication, and we expect that it will advance our understanding of quantum communication.

Theorems 1.1 and 1.2 are both implied by a general bound, which we prove in the following sections.

**Theorem 1.3** *Suppose Alice wishes to convey $n$ bits to Bob, over a quantum communication channel. Consider any protocol* without prior entanglement *such that the total number of qubits exchanged by the two parties is $m$, and the total number of qubits sent by Alice to Bob is $m_A$, over all the rounds of communication. Let $Y$ be the random variable denoting Bob's output (his guess for $X$), when Alice wishes to convey the random variable $X$. Let $P(X, \ell)$ denote the net probability of the $\ell$ most likely strings in the sample space of the random variable $X$. Then, irrespective of the number of rounds of communication, the probability that Bob correctly recovers $X$ is bounded as*

1. $\Pr[Y = X] \leq P(X, 2^m)$, *and also as*
2. $\Pr[Y = X] \leq P(X, 2^{2m_A})$, *independent of the number of qubits sent by Bob to Alice.*

For any given $X, m, m_A$, there is a natural communication protocol that saturates the stronger bound among the two, showing that the result is optimal. Note that we get Theorem 1.1 by considering the uniform distribution on the $n$-bit strings. Theorem 1.2 follows by considering the uniform distribution on the $n$-bit strings, and imagining that the shared entanglement was created through an initial message from Bob to Alice.

We also give an application of our results to communication complexity. Combining our results with a reduction due to Cleve, van Dam, Nielsen, and Tapp [13], we get a lower bound of $\frac{1}{2}n - \log \frac{1}{1-2\epsilon}$ for the $\epsilon$-error quantum communication complexity of the inner product function. The complexity of this function was shown to be $\Omega(n - \log \frac{1}{1-2\epsilon})$ by Kremer [23]. A more precise bound of $\frac{1}{2}(n - \log \frac{1}{1-4\epsilon})$ may be derived using a method due to Ambainis, Ta-Shma, Schulman, Vazirani, and Wigderson [2]. Our bound is also valid for bounded-error entanglement-assisted protocols and improves over the lower bound of $\frac{1}{2}((1-2\epsilon)^2 n - 1)$ in Ref. [13].

Van Dam and Hayden [35] independently proved a lower bound of $\frac{1}{2}n - \log \frac{1}{1-2\epsilon}$ for Inner Product under the assumption that the communication parties shared *EPR pairs*. However, their information-theoretic approach provably breaks down in the presence of arbitrary prior entanglement.

There is a classical $n - \log \frac{1}{1-2\epsilon} + 1$ bit one-way public-coin protocol for Inner Product, and hence a $\frac{1}{2}(n - \log \frac{1}{1-2\epsilon} + 1)$ qubit quantum protocol with shared EPR pairs. Our lower bound nearly matches this. Thus,

our results provide more examples where shared entanglement leads to at most a factor of two savings in communication.

The one-way communication complexity of Inner Product is of particular interest because of its use in proofs of security of privacy amplification techniques in quantum cryptography. The analysis due to Ben-Or [4] of the Bennett-Brassard protocol for quantum key distribution explicitly relies on the hardness of Inner Product. We derive a bound of $n - 2\log\frac{1}{1-2\epsilon}$ for the $\epsilon$-error one-way communication complexity of the inner product function. This gives a tight bound on the information an eavesdropper has about a secret key that is obtained by hashing a "raw key" with a random linear function.

Finally, we would like to point out that while the results presented here rule out obvious methods of compressing classical information into succinct quantum messages, certain other tasks can be accomplished at a significantly smaller communication cost, when compared to classical protocols [10, 2, 32, 3]. In a recent result, Raz [33] also demonstrates how to circumvent our results and those in Refs. [17, 1] by using Arthur-Merlin proofs. Evidently, the information theoretic aspects of quantum physics are very subtle, and call for deeper investigation.

### Organisation of the article

We refer the reader to texts such as [31, 30] for background in quantum information and computation. In Appendix A, we revise the role of shared randomness in transmitting classical information. The quantum communication model and some non-standard notation are described in Section 2. We begin by analysing quantum *encoding* of classical bits, first in the absence of prior entanglement (Section 3), then with entanglement (Section 4). In Section 4.1 we consider a very restricted kind of encoding, where the shared state consists of EPR pairs, and no ancillary qubits are used in the encoding. This contains the basic elements of the proof for encoding with general entanglement as well, which is the subject of Section 4.2. Building on the insight gained from the study of quantum encodings, we extend our results to the case of interactive communication in Section 5. We conclude with a discussion of the consequences of our results.

## 2 Preliminaries

### 2.1 Non-standard notation and a basic fact

We start by explaining the non-standard notation we use in this article. For a detailed treatment of the model of quantum computation and the associated mathematical formalism, we refer the reader to texts such as [31, 30].

For any random variable $X$ over a finite sample space, and integer $\ell \geq 1$, we use $P(X, \ell)$ to denote the net probability of the $\ell$ most likely outcomes in the sample space.

By a *mixed state* over a set of qubits, we mean a probability distribution $\{p_i, |\phi_i\rangle\}$ over superpositions (or *pure* states), where the state $|\phi_i\rangle$ occurs with probability $p_i$. We will sometimes use the notation $\{|\phi_i\rangle\}$ for a mixed state. Here the states $|\phi_i\rangle$ are in general unnormalised, and are such that $\sum_i \| \phi_i \|^2 = 1$.

We denote the identity operator on states over $k$ qubits by $I_k$. For any linear operator $U$ on these qubits, $U^\dagger$ denotes its adjoint and $U^\mathsf{T}$ denotes its transpose.

The following theorem gives a useful characterisation of bi-partite quantum states (see Ref. [30, Section 2.5]).

**Theorem 2.1 (Schmidt decomposition)** *Any unit vector $|\phi\rangle$ in a bi-partite Hilbert space $H \otimes K$ may be represented as*

$$|\phi\rangle \quad = \quad \sum_i \sqrt{\lambda_i}\, |e_i\rangle |f_i\rangle,$$

*where $\{|e_i\rangle\}$ and $\{|f_i\rangle\}$ are orthonormal sets of states in $H$ and $K$ respectively, and the $\lambda_i$ are non-negative reals such that $\sum_i \lambda_i = 1$.*

## 2.2 The quantum communication model

Formally, a two-party communication protocol for a function is a partition of a quantum circuit into two sets, where the input wires and the gates may be divided arbitrarily amongst the two, but all the output wires lie in one of the sets of the partition [36]. The complexity of the protocol is the number of wires crossing between the two parts of the circuit.

It is however simpler to describe a quantum protocol informally as follows. Two parties, Alice and Bob, hold an arbitrarily large supply of private qubits in some fixed basis state, say $|\bar{0}\rangle$. The initial joint state is thus $|\bar{0}\rangle_A \otimes |\bar{0}\rangle_B$, where a subscript indicates the player holding that set of qubits. When the game starts, they are given (classical) inputs $x \in X$ and $y \in Y$ respectively. The two parties then play in turns. Suppose it is Alice's turn to play. Alice can do an arbitrary unitary transformation (that depends only on her input $x$) on her qubits and then send one or more qubits to Bob. Sending qubits does not change the overall superposition, but rather changes the ownership of the qubits, allowing Bob to apply his next unitary transformation on the newly received qubits. The qubits sent in any round are independent of the inputs to the two players, and is pre-specified by the protocol. At the end of the protocol, one player measures one or more qubits in some basis, and declares the outcome as the result of the protocol. (In cases where a specific player is required to know the answer, that player makes the measurement.)

The complexity of a quantum protocol is the total number of qubits exchanged between the two players over all rounds. We say a protocol *computes* a function $f$ on $X \times Y$ with error $\epsilon \geq 0$ if, for any input $x \in X, y \in Y$, the probability that the result of the protocol is equal to $f(x, y)$ is at least $1 - \epsilon$. $Q_\epsilon(f)$ denotes the complexity of the best quantum protocol that computes $f$ with at most $\epsilon$ error.

Note that there is no loss of generality in not allowing the players to measure a subset of their quantum bits in the intermediate steps of a protocol. This is because all measurements may be postponed to the end by the *principle of safe storage (or delayed measurement)* [9]. While this may result in an increase in the (private) work qubits needed by the two parties, the over all communication between them remains unaffected.

In the quantum analogue of public-coin classical communication, Alice and Bob may initially share an arbitrary number of quantum bits which are in some pure state that is independent of the inputs. In this case, the initial joint state takes the form $|\bar{0}\rangle_A \sum_{i,j} \alpha_{ij} |i\rangle_A |j\rangle_B |\bar{0}\rangle_B$. This is known as *communication with prior entanglement* (see, e.g., Refs. [12, 13, 11]), or in information-theoretic terms, as *communication over an entanglement-assisted quantum channel* (see, e.g., Ref. [7]). We will denote the entanglement-assisted quantum communication complexity of a function $f$ as above by $Q^*_\epsilon(f)$.

On occasion, we will concentrate on communication in one round (i.e., with only one message), since this often sheds light on fundamental properties of protocols for certain problems. The message in a one-round protocol in which only one player gets an input is called an *encoding* of the input. The operations done by the other player, and her measurement are together referred to as the *decoding*.

## 3 Communication without prior entanglement

In this section, we prove part (1) of Theorem 1.3. The essence of the result is contained in the following theorem about quantum encoding of classical bits.

**Theorem 3.1** *Let $X$ be a random variable over bit strings in $\{0,1\}^n$ that are encoded as mixed states over $m$ qubits. If $Y$ is any random variable over $\{0,1\}^n$ obtained by making some measurement of the encoding of $X$, then*

$$\Pr[Y = X] \;\leq\; P(X, 2^m).$$

*In particular, if $X$ is uniformly distributed, the probability $\Pr[Y = X] \leq \frac{2^m}{2^n}$.*

**Proof:** Consider any encoding of strings $x \mapsto \{q_{x,i}, |\phi_{x,i}\rangle\}$ into mixed states over $m$ qubits. Assume that the random variable $Y$ is the outcome of a measurement given by orthogonal projections $\{P_x\}$ in the Hilbert space of the encoding, augmented with some ancillary qubits initialised to $|\bar{0}\rangle$. The probability $\Pr[Y = X]$ may then be bounded as

$$\begin{aligned}
\Pr[Y = X] &= \sum_x \Pr[Y = x | X = x] \cdot \Pr[X = x] \\
&= \sum_x p_x \sum_i q_{x,i} \parallel P_x |\phi_{x,i}\rangle |\bar{0}\rangle \parallel^2 \\
&\leq \sum_x p_x \parallel P_x |\phi_x\rangle \parallel^2 , \qquad (1)
\end{aligned}$$

where $p_x = \Pr[X = x]$, and $|\phi_x\rangle$ is the pure state $|\phi_{x,i}\rangle|\bar{0}\rangle$ that maximizes (over all $i$) the probability $\parallel P_x|\phi_{x,i}\rangle|\bar{0}\rangle \parallel^2$ of obtaining the correct outcome $x$. Thus, in an optimal encoding, we may assume that every string $x$ is encoded as a pure state.

We can now bound the decoding probability by using the following lemma.

**Lemma 3.2** $\sum_x \parallel P_x|\phi_x\rangle \parallel^2 \leq 2^m$.

**Proof:** Let $H$ be the subspace spanned by the codewords $|\phi_x\rangle$, and let $Q$ be the projection onto $H$. Since the code states $\{|\phi_{x,i}\rangle\}$ are over $m$ qubits, $H$ has dimension at most $2^m$.

Let $\{|e_{x,j}\rangle\}_j$ be an orthonormal basis for the range of $P_x$. Since the projection operators $\{P_x\}$ are orthogonal, the union of all these bases $\{|e_{x,j}\rangle\}_{x,j}$ is an orthonormal basis for the entire decoding Hilbert space. Now,

$$\begin{aligned}
\parallel P_x|\phi_x\rangle \parallel^2 &= \sum_j |\langle e_{x,j}|\phi_x\rangle|^2 \\
&\leq \sum_j \parallel Q|e_{x,j}\rangle \parallel^2 .
\end{aligned}$$

The last inequality follows because the length of the projection of the vector $|e_{x,j}\rangle$ onto the space $H$ is at least the length of its projection onto the one-dimensional subspace of $H$ that is spanned by $|\phi_x\rangle$.

Continuing the analysis, we have

$$\begin{aligned}
\sum_x \parallel P_x|\phi_x\rangle \parallel^2 &\leq \sum_{x,j} \langle e_{x,j}|Q|e_{x,j}\rangle \\
&= \mathrm{Tr}\; Q \;\leq\; 2^m, \qquad (2)
\end{aligned}$$

since $Q$ is an orthogonal projection onto the Hilbert space $H$ of dimension at most $2^m$. ■

By equation (1), the probability of correct decoding is at most $\sum_x p_x \parallel P_x|\phi_x\rangle \parallel^2$. From Lemma 3.2 above, this expression may be bounded by the maximum of $\sum_x p_x\lambda_x$, where $0 \leq \lambda_x \leq 1$ and $\sum_x \lambda_x \leq 2^m$. The theorem now follows from the next intuitive lemma.

**Lemma 3.3** *If $0 \leq \lambda_x \leq 1$, and $\sum_x \lambda_x \leq \ell$, then $\max_\lambda \sum_x p_x\lambda_x \leq P(X, \ell)$.*

Note that the upper bound in the lemma is achieved when $\lambda_x = 1$ for strings $x$ with the $\ell$ highest probability masses $p_x$, and zero for the rest.

We leave the proof of this lemma to the reader. ■

We now clarify how Theorem 3.1 may be applied in a communication complexity context (as opposed to a coding context) where more than one message may be exchanged by the communicating parties. The missing link is the following characterisation due to Yao [36], of the joint state of the communicating parties at the end of a quantum protocol.

**Lemma 3.4 (Yao [36])** *In any $m$-qubit quantum communication protocol between two parties Alice and Bob, with inputs $x$ and $y$ respectively, and initial joint state $|\bar{0}\rangle_A \otimes |\bar{0}\rangle_B$, the final state may be written as*

$$\sum_{c\in\{0,1\}^m} U_c^x|\bar{0}\rangle_A \otimes V_c^y|\bar{0}\rangle_B,$$

*where $U_c^x, V_c^y$ are linear operators that depend only on the inputs of Alice and Bob respectively, and the classical possibilities $c$ for the state of the qubits they send in the different rounds.*

We are interested in protocols in which only Alice receives an input. In such protocols, Bob's final state is always a combination of the states $V_c|\bar{0}\rangle$. This allows us to convert every such multi-round protocol into a protocol with one message from Alice to Bob. In other words, we can derive an equivalent *encoding* from it.

**Corollary 3.5** *Suppose $\mathcal{P}$ is any $m$-qubit, possibly multi-round, quantum communication protocol (without prior entanglement) in which only one party, say Alice, receives an input. Then, there is a protocol $\mathcal{P}'$ with only one message of length $m$ from Alice to Bob that is equivalent to $\mathcal{P}$.*

**Proof:** The idea is to recreate the joint state at the end of the protocol $\mathcal{P}$ by sending one length $m$ message. This is possible because Bob's share of the final state lies in a $d \leq 2^m$ dimensional subspace. The protocol $\mathcal{P}'$ compresses states in this $d$-dimensional space to at most $\log d$ qubits. The details follow.

The protocol $\mathcal{P}'$ is designed as follows. By Lemma 3.4, the final state at the end of $\mathcal{P}$ on input $x$ is

$$|\psi_x\rangle \quad = \quad \sum_{c\in\{0,1\}^m} U_c^x|\bar{0}\rangle_A \otimes V_c|\bar{0}\rangle_B \quad \in \quad H \otimes K,$$

where the qubits corresponding to $H$ are with Alice and those corresponding to $K$ with Bob. Alice constructs this state on her own; she has complete knowledge of the transformations used by Bob.

The states $\{V_c|\bar{0}\rangle\}_c$ all lie in a subspace $G$ of $K$ of dimension at most $2^m$. Thus, there is a unitary transformation $W$ on $K$ that maps $G$ to the subspace $|\bar{0}\rangle \otimes \mathbb{C}^m$. I.e., the $K$ part of the state may be compressed into $m$ qubits.

Alice applies the transformation $W$ to the $K$-part of the state $|\psi_x\rangle$, and sends the trailing $m$ qubits to Bob. Bob applies the inverse transformation $W^\dagger$ to the qubits he receives along with the requisite ancilla, and thus reconstructs the state $|\psi_x\rangle$. From here, both parties follow the original protocol. The protocol $\mathcal{P}'$ has all the features we desire. ■

Theorem 1.3(1) now follows by combining Corollary 3.5 with Theorem 3.1.

# 4 Bounds for encoding with prior entanglement

In this section we concentrate on one-way protocols, or *encodings*, in the presence of shared entanglement in which one party, Alice, wishes to send some number of classical bits to Bob. The main difficulty in extending Theorem 1.3(1) to this case is that because of the unlimited number of entangled qubits the two parties may share, the space in which the encodings reside has unbounded dimension. The bound on this dimension, in terms of the number of qubits sent, was crucial in deriving Theorem 1.3(1). We overcome this barrier by analysing the encodings in more detail. We demonstrate that the encodings all are of a special form that allows us to apply the technique used in the proof of Theorem 1.3(1).

## 4.1 Encoding over EPR pairs, without ancilla

We first prove our results in the case where Alice does not use any ancillary qubits in the encoding process, and Alice and Bob share some number of EPR pairs. This motivates the proof in the more general case, and illustrates its essential elements.

We start with a simple property of maximally entangled states, such as EPR pairs. This property has appeared in a different guise in earlier work such as that on the impossibility of quantum protocols for ideal bit commitment [26, 25]. The lemma allows us to analyse the encoding process easily.

**Lemma 4.1** *For any unitary transformation $U$ on $E$ qubits, and any orthonormal set $\left\{|\phi_a\rangle : a \in \{0,1\}^E\right\}$ over $E' \geq E$ qubits,*

$$\sum_{a\in\{0,1\}^E} U|a\rangle \otimes |\phi_a\rangle \quad = \quad \sum_{a\in\{0,1\}^E} |a\rangle \otimes \tilde{U}|\phi_a\rangle,$$

*where $\tilde{U}$ is any transformation on $E'$ qubits such that for all $a, a' \in \{0,1\}^E$, $\langle\phi_a|\tilde{U}|\phi_{a'}\rangle = \langle a'|U|a\rangle$.*
*In particular, if $E' = E$, and $|\phi_a\rangle = |a\rangle$ (i.e., we have an unnormalised EPR state) then*

$$\sum_{a\in\{0,1\}^E} U|a\rangle \otimes |a\rangle \quad = \quad \sum_{a\in\{0,1\}^E} |a\rangle \otimes U^{\mathsf{T}}|a\rangle.$$

**Proof:** Observe that for $b, c \in \{0,1\}^E$,

$$\begin{aligned}
\langle b|\langle\phi_c| \sum_{a\in\{0,1\}^E} U|a\rangle|\phi_a\rangle &= \langle b|U|c\rangle \\
&= \langle\phi_c|\tilde{U}|\phi_b\rangle \\
&= \langle b|\langle\phi_c| \sum_{a\in\{0,1\}^E} |a\rangle\tilde{U}|\phi_a\rangle.
\end{aligned}$$

The lemma follows. ■

We can now characterise the encoding process (without ancilla) as follows.

**Lemma 4.2** *Suppose that Alice performs a unitary transformation on her share of $E$ EPR pairs, and then sends $m$ of the $E$ qubits to Bob. Then, Bob has $E+m$ qubits in a mixed state that can be represented as $\{p_l, |\phi_l\rangle\}$ ($l \in \{0,1\}^{E-m}$) with $\{|\phi_l\rangle\}_l$ orthonormal, and $p_l = \frac{1}{2^{E-m}}$.*

**Proof:** Suppose that Alice applies a transformation $V$ to her part of the state. By Lemma 4.1, the resulting state is

$$\frac{1}{2^{E/2}} \sum_{a\in\{0,1\}^E} V|a\rangle_A |a\rangle_B \quad = \quad \frac{1}{2^{E/2}} \sum_{a\in\{0,1\}^E} |a\rangle_A V^{\mathsf{T}} |a\rangle_B.$$

After the communication, Alice and Bob's joint state may be written as (w.l.o.g., Alice sends the rightmost $m$ qubits to Bob):

$$\frac{1}{2^{(E-m)/2}} \sum_{l\in\{0,1\}^{E-m}} |l\rangle_A \ \ \frac{1}{2^{m/2}} \sum_{r\in\{0,1\}^m} |r\rangle_B V^{\mathsf{T}} |lr\rangle_B.$$

Consider the mixed state obtained on Bob's side if Alice measures her qubits in the standard basis. The probability $p_l$ of Alice observing any given $l$ is $\frac{1}{2^{E-m}}$. The state of Bob's $E+m$ qubits when Alice gets outcome $l$ is

$$|\phi_l\rangle \quad = \quad \frac{1}{2^{m/2}} \sum_{r\in\{0,1\}^m} |r\rangle V^{\mathsf{T}} |lr\rangle.$$

We may easily verify that these are orthonormal for different $l$:

$$\begin{aligned} \langle\phi_l|\phi_{l'}\rangle &= \frac{1}{2^m} \sum_r \langle lr| (V^{\mathsf{T}})^\dagger V^{\mathsf{T}} |l'r\rangle \\ &= \frac{1}{2^m} \sum_r \langle lr|l'r\rangle \\ &= \delta_{l,l'}. \end{aligned}$$

Note that the above measurement by Alice is one of several ways of arriving at a description of the state of Bob's qubits. It does not affect the decoding process; Bob's density matrix remains unchanged by it (see Ref. [30, Section 2.4], especially Section 2.4.3). Nonetheless, it allows us to express Bob's mixed state in a convenient form. ■

By a simple dimensional argument, we can now get an alternative proof of the fact that the superdense coding scheme in Ref. [8] is optimal (in the case of encoding without ancilla): Suppose we have an entanglement-assisted scheme for encoding $n$ bits into $m$ qubits that uses $E$ EPR pairs, does not use ancillary qubits in the encoding transformation, and makes no error in decoding. The lemma above says that Bob's state lies in a subspace of dimension $2^{E-m}$ for each $n$-bit string. Moreover, since the probability of correct decoding is 1, the subspaces corresponding to different strings are orthogonal. Thus, Bob's state space has dimension at least $2^n \cdot 2^{E-m}$. On the other hand, these states are $E+m$ qubits long, which implies that $m \geq n/2$.

If in a quantum (or a classical randomized) protocol a small amount of probabilistic error is allowed, the communication may be reduced drastically (see, for example, Refs. [10, 2, 24]). The following theorem places limits on the savings achieved when we wish to convey classical messages by quantum means.

**Theorem 4.3** *If Alice encodes messages $x \in \{0,1\}^n$ over EPR pairs without ancilla, and sends $m$ qubits to Bob, the probability of correct decoding of a message chosen uniformly at random is bounded as* $\Pr[\textit{correct decoding}] \leq \frac{2^{2m}}{2^n}$.

**Proof:** Suppose that the number of EPR pairs Alice and Bob share initially is $E$. Let $\{p_{x,l}, |\phi_{x,l}\rangle\}_l$ be Bob's mixed state when Alice has input $x \in \{0,1\}^n$, as given by Lemma 4.2.

Recall that the decoding procedure used by Bob consists of measuring the encoded state with some ancillary qubits (assumed to be initialised to state $|\bar{0}\rangle$) with some orthogonal projection operators $\{P_y\}$. Here, the outcome $y \in \{0,1\}^n$ corresponds to Bob's guess for the encoded message. We will omit the ancilla from the expressions below, for clarity of exposition.

Let $C$ be the event that Bob decodes a message correctly and $C_x$ ($C_{x,l}$) that he does so on receiving the encoding of $x$ ($|\phi_{x,l}\rangle$, respectively). Let $\mathbf{x}$ be the event that Alice encodes message $x$, and $\mathbf{x}_l$ that $|\phi_{x,l}\rangle$ is prepared given that $x$ is encoded. Then

$$\begin{aligned}
\Pr[C] &= \sum_x \Pr[C_x] \cdot \Pr[\mathbf{x}] \\
&= \sum_x \frac{\Pr[C_x]}{2^n} \\
&= \sum_{x,l} \frac{\Pr[C_{x,l}] \cdot \Pr[\mathbf{x}_l]}{2^n} \\
&= \sum_{x,l} \frac{\Pr[C_{x,l}]}{2^{E-m} 2^n}. \qquad (3)
\end{aligned}$$

It thus suffices to bound $\sum_{x,l} \Pr[C_{x,l}]$. Observe that

$$\Pr[C_{x,l}] = \| P_x |\phi_{x,l}\rangle \|^2 . \qquad (4)$$

We introduce some notation. For each $x$, let $H_x$ be the space spanned by $\{|\phi_{x,l}\rangle\}_l$. Note that $\{|\phi_{x,l}\rangle\}_l$ is an orthonormal basis for $H_x$. Let $R_x$ be the projection onto $H_x$. Since we allow a small error in the decoding process, the different spaces $H_x$ may not be orthogonal.

Let $H$ be the space spanned by all the vectors $\{|\phi_{x,l}\rangle\}_{x,l}$, and $Q$ the projection operator onto $H$. For each $x$, let the set $\{|e_{x,j}\rangle\}_j$ be an orthonormal basis for the range of $P_x$. Then $\{|e_{x,j}\rangle\}_{x,j}$ is an orthonormal basis for the entire decoding space.

Now, following the proof of Lemma 3.2,

$$\begin{aligned}
\sum_l \| P_x |\phi_{x,l}\rangle \|^2 &= \sum_{l,j} |\langle e_{x,j} | \phi_{x,l} \rangle|^2 \\
&= \sum_j \| R_x |e_{x,j}\rangle \|^2 \\
&\leq \sum_j \| Q |e_{x,j}\rangle \|^2 , \qquad (5)
\end{aligned}$$

since the length of the projection of $|e_{x,j}\rangle$ onto $H_x$ is at most the length of its projection on the space $H$ (of which $H_x$ is a subspace).

From equation (5),

$$\begin{aligned}\sum_{x,l} \| P_x|\phi_{x,l}\rangle \|^2 &\leq \sum_{x,j} \| Q|e_{x,j}\rangle \|^2 \\ &= \sum_{x,j} \langle e_{x,j}|Q|e_{x,j}\rangle \\ &= \mathrm{Tr}\, Q \quad = \quad \dim H \\ &\leq 2^{E+m}, \end{aligned} \tag{6}$$

since the space $H$ is generated by states over $E+m$ qubits.

Combining equations (3), (4), and (6), we get

$$\Pr[C] \leq \frac{2^{E+m}}{2^{E-m}2^n} = \frac{2^{2m}}{2^n},$$

as claimed. ■

Encoding with EPR pairs along with some ancilla leads to states very similar to those in Lemma 4.2, and Theorem 4.3 holds in that case as well. We will however skip ahead to encoding where Alice uses extra space, and an *arbitrary* entangled state.

### 4.2 Encoding with general prior entanglement

In general, in trying to transmit information, Alice and Bob may share an *arbitrary* entangled state (independent of their inputs) before they interact. In this section we show that the result in the previous section applies irrespective of which initial entangled state Alice and Bob share.

The main difficulty here is that the property of messages encoded over EPR pairs captured by Lemma 4.2 may fail to hold. However, we show a simple connection between encoding with EPR pairs and encoding with an arbitrary entangled state that allows us to conclude an identical result.

We start by observing that we need only consider protocols which make use of a special kind of shared state.

**Observation 4.4** *In any quantum communication protocol with prior entanglement, we may assume, without loss of generality, that the initial shared state is of the form*

$$\sum_{a\in\{0,1\}^E} \sqrt{\lambda_a}\, |a\rangle_A |a\rangle_B,$$

*where $\lambda_a$ are non-negative reals, and $\sum_a \lambda_a = 1$.*

**Proof:** This follows directly from the Schmidt decomposition theorem (Theorem 2.1). Consider a protocol $\mathcal{P}$ in which the quantum state shared by Alice and Bob has $E_A$ qubits on Alice's side and $E_B$ qubits on Bob's side. For concreteness, assume that $E_A \leq E_B$. By Theorem 2.1, the shared state may be expressed as

$$\sum_{b\in\{0,1\}^{E_A}} \sqrt{\mu_b}\, |\phi_b\rangle_A |\psi_b\rangle_B,$$

where the $\mu_b$ are non-negative reals summing up to 1, and the sets $\{|\phi_b\rangle\}$ and $\{|\psi_b\rangle\}$ are orthonormal. We may modify the protocol to a new protocol $\mathcal{P}'$, which has the same behaviour as $\mathcal{P}$ on each input, but where the shared state is of the form described in the observation above. Consider any unitary transformations $U, V$ on $E = E_B$ qubits such that for every $b \in \{0,1\}^{E_A}$,

$$\begin{aligned} U &: |\bar{0}, b\rangle \mapsto |\bar{0}\rangle|\phi_b\rangle \\ V &: |\bar{0}, b\rangle \mapsto |\psi_b\rangle. \end{aligned}$$

Let $\lambda_{\bar{0}b} = \mu_b$, for $b$ as above, and let the rest of the $\lambda_a$ be 0. The protocol $\mathcal{P}'$ begins with the shared state

$$\sum_{a \in \{0,1\}^E} \sqrt{\lambda_a}\, |a\rangle_A |a\rangle_B,$$

and then Alice and Bob apply $U$ and $V$ to their qubits respectively. Thereafter, the protocol proceeds exactly as in $\mathcal{P}$. By construction, the protocols behave the same way for each input. ■

We make another simplifying observation about the protocols that we need consider.

**Observation 4.5** *In any quantum communication protocol with prior entanglement, we may assume, without loss of generality, that neither Alice nor Bob uses any ancillary qubits in their local unitary operations or measurements.*

This is because all the ancillary qubits used may be considered as part of the initial shared state.

The above observations allow us to relate the encoding with a general entangled state to the encoding obtained when EPR pairs are used instead.

**Lemma 4.6** *Suppose that Alice performs a unitary transformation on her share of the joint state*

$$\sum_{a \in \{0,1\}^E} \sqrt{\lambda_a}\, |a\rangle_A |a\rangle_B,$$

*and then sends $m$ of the $E$ qubits to Bob. Then, Bob has $E + m$ qubits in a mixed state that can be represented as*

$$\left\{ 2^{m/2} (I_m \otimes \Lambda) |\phi_l\rangle \right\}_{l \in \{0,1\}^{E-m}}$$

*with $\{|\phi_l\rangle\}_l$ orthonormal, and $\Lambda = \sum_a \sqrt{\lambda_a}\, |a\rangle\langle a|$.*

**Proof:** The main technical insight here is that the entangled state $\sum_{a \in \{0,1\}^E} \sqrt{\lambda_a}\, |a\rangle_A |a\rangle_B$ may be viewed as a scaled version of some number of EPR pairs:

$$(I_E \otimes \Lambda) \sum_a |a\rangle|a\rangle,$$

where $\Lambda = \sum_a \sqrt{\lambda_a}\, |a\rangle\langle a|$ is a diagonal matrix with Frobenius norm $\mathrm{Tr}(\Lambda\Lambda^\dagger) = 1$. The matrix $\Lambda$ scales vectors along orthogonal dimensions according to the Schmidt coefficients $\{\lambda_a\}$.

Suppose Alice applies the transformation $V$ to her $E$ qubits. The resulting joint state is

$$\begin{aligned}
&(V \otimes I_E)(I_E \otimes \Lambda) \sum_a |a\rangle|a\rangle \\
&= (I_E \otimes \Lambda) \sum_a V|a\rangle|a\rangle \\
&= (I_E \otimes \Lambda) \sum_a |a\rangle V^{\mathsf{T}}|a\rangle \qquad (7) \\
&= \sum_{l \in \{0,1\}^{E-m}} |l\rangle \ (I_m \otimes \Lambda) \sum_{r \in \{0,1\}^m} |r\rangle V^{\mathsf{T}}|lr\rangle,
\end{aligned}$$

where equation (7) follows from Lemma 4.1. Let, as in Lemma 4.2,

$$|\phi_l\rangle = 2^{-m/2} \sum_{r \in \{0,1\}^m} |r\rangle V^{\mathsf{T}}|lr\rangle.$$

Suppose Alice sends $m$ of her qubits to Bob, and measures the remaining qubits in the standard basis. The residual, unnormalised, state with Bob is then $2^{m/2}(I_m \otimes \Lambda)|\phi_l\rangle$, when she observes $l \in \{0,1\}^{E-m}$. That the states $|\phi_l\rangle$ are orthonormal is shown in the proof of Lemma 4.2. ■

We can now prove the equivalent of Theorem 4.3 when Alice and Bob share an arbitrary entangled state.

**Theorem 4.7** *If Alice encodes $2^n$ messages over her part of an arbitrary shared entangled state (that is independent of her messages) and some ancillary qubits, and sends $m$ qubits to Bob, the probability of correct decoding of a message chosen uniformly at random is bounded as* $\Pr[\textit{correct decoding}] \leq \frac{2^{2m}}{2^n}$.

**Proof:** We use the same notation as in the proof of Theorem 4.3, adapted to the encoding we get here due to the more general entangled state.

By Observation 4.5, we may assume that Alice and Bob operate only on their shared entangled state. We may further assume that this state is of the special form described in Observation 4.4.

Let $\left\{2^{m/2}(I_m \otimes \Lambda)|\phi_{x,l}\rangle\right\}_l$ be Bob's mixed state when Alice encodes $x \in \{0,1\}^n$, as given by Lemma 4.6. Since no ancilla is used in the decoding procedure (i.e., in Bob's measurement to extract $x$, cf. Observation 4.5), the projection operators $P_y$ are over $E+m$ qubits. Now,

$$\Pr[C] = \sum_x \frac{\Pr[C_x]}{2^n}, \quad \text{and} \qquad (8)$$

$$\Pr[C_x] = 2^m \sum_l \| \, P_x(I_m \otimes \Lambda)|\phi_{x,l}\rangle \, \|^2 . \qquad (9)$$

Furthermore,

$$\begin{aligned}
\sum_{x,l} \| P_x(I_m \otimes \Lambda)|\phi_{x,l}\rangle \|^2 &= \sum_{x,l,j} |\langle e_{x,j}|(I_m \otimes \Lambda)|\phi_{x,l}\rangle|^2 \\
&= \sum_{x,j} \left\| R_x(I_m \otimes \Lambda^\dagger)|e_{x,j}\rangle \right\|^2 \\
&\leq \sum_{x,j} \| (I_m \otimes \Lambda)|e_{x,j}\rangle \|^2 \\
&= \sum_{x,j} \langle e_{x,j}|(I_m \otimes \Lambda^\dagger\Lambda)|e_{x,j}\rangle \\
&= \text{Tr } (I_m \otimes \Lambda^2) \\
&= 2^m \sum_a \lambda_a \;=\; 2^m. \qquad (10)
\end{aligned}$$

Combining equations (8), (9) and (10), we get $\Pr[C] \leq 2^{2m}/2^n$. ■

# 5 The extension to interactive communication

In this section, we analyse the most general quantum protocols for exchanging information. In these protocols, Alice and Bob share an arbitrary entangled state to begin with, and exchange messages both ways in order to communicate.

The essential idea behind the results below is contained in Lemma 4.6, and leads to a new characterisation of the joint state in quantum protocols (Lemma 5.1 below). In order to prove this lemma from first principles, we focus on protocols *in which there is no prior entanglement*. That it holds also for communication with prior entanglement may be inferred from the lemma itself by applying it to a modification of an entanglement-assisted protocol. In the modified protocol, the entangled state to be shared is generated by Bob, who sends the appropriate part of it to Alice. The parameter of interest in the characterisation in Lemma 5.1 is the Frobenius norm $\text{Tr}(\Lambda\Lambda^\dagger)$ of the matrix $\Lambda$ which generalises the scaling matrix of Lemma 4.6. This norm depends only on the number $m_A$ of qubits of communication from Alice to Bob, and hence remains unchanged despite the modification.

**Lemma 5.1** *Let $\mathcal{P}$ be any quantum communication protocol (without prior entanglement) in which the number of qubits sent by Alice to Bob (Bob to Alice) is $m_A$ (respectively, $m_B$), and the final number of qubits with Alice (Bob) is $q_A$ (respectively, $q_B$). Then, the joint state of Alice and Bob at the end of the protocol may be expressed as*

$$\sum_{a \in \{0,1\}^{q_A}} |a\rangle_A \; \Lambda|\phi_a\rangle_B,$$

*where*

1. *$\Lambda$ is a linear transformation that maps $q_B + 2m_B$ qubits to $q_B$ qubits, depends only on the unitary transformations of Bob, and satisfies $\text{Tr}(\Lambda\Lambda^\dagger) = 2^{2m_A}$, and*
2. *$\{|\phi_a\rangle\}$ is an orthonormal set of states over $q_B + 2m_B$ qubits, and depends only on the unitary transformations of Alice.*

**Proof:** The proof goes by induction on the number of rounds $t$.

In the beginning (for $t = 0$), the joint state (w.l.o.g.) is $|\bar{0}\rangle_A \otimes |\bar{0}\rangle_B$, which represents *all* the qubits the two players use during the protocol. This is of the form described in the lemma, with $\Lambda = \Lambda_0 = |\bar{0}\rangle\langle\bar{0}|$.

Let $q_{A,t}, q_{B,t}, m_{A,t}, m_{B,t}$ be the quantities corresponding to $q_A, q_B, m_A, m_B$ after $t \geq 0$ rounds of communication. Assume that at this stage, the joint state of Alice and Bob is

$$\sum_{a \in \{0,1\}^{q_{A,t}}} |a\rangle_A \; \Lambda_t |\phi_{a,t}\rangle_B,$$

where $\Lambda_t$ and $\{|\phi_{a,t}\rangle\}$ satisfy the conditions stated in the lemma (in terms of $q_{B,t}, m_{A,t}, m_{B,t}$). We look at two cases for the $(t+1)$'th round of communication.

Case (a). Alice applies a unitary transformation $U$ to her qubits and sends $p$ qubits to Bob.

The state after the unitary transformation is

$$\begin{aligned}
& (U \otimes I_{q_{B,t}})(I_{q_{A,t}} \otimes \Lambda_t) \sum_{a \in \{0,1\}^{q_{A,t}}} |a\rangle |\phi_{a,t}\rangle \\
&= (I \otimes \Lambda_t) \sum_{a \in \{0,1\}^{q_{A,t}}} U|a\rangle |\phi_{a,t}\rangle \\
&= (I \otimes \Lambda_t) \sum_{a \in \{0,1\}^{q_{A,t}}} |a\rangle \; \tilde{U} |\phi_{a,t}\rangle,
\end{aligned}$$

where $\tilde{U}$ is a unitary transformation on Bob's qubits as given by Lemma 4.1. Thus, after Alice sends $p$ of her qubits to Bob (w.l.o.g., these are the $p$ rightmost qubits), the joint state looks like

$$\sum_{l \in \{0,1\}^{q_{A,t+1}}} |l\rangle_A \; (I_p \otimes \Lambda_t) \sum_{r \in \{0,1\}^p} |r\rangle_B \; \tilde{U} |\phi_{lr,t}\rangle_B. \tag{11}$$

Here, $q_{A,t+1} = q_{A,t} - p$, $q_{B,t+1} = q_{B,t} + p$, $m_{A,t+1} = m_{A,t} + p$, and $m_{B,t+1} = m_{B,t}$.

Let

$$\begin{aligned}
\Lambda_{t+1} &= 2^{p/2} (I_p \otimes \Lambda_t), \quad \text{and} \\
|\phi_{l,t+1}\rangle &= 2^{-p/2} \sum_r |r\rangle \; \tilde{U} |\phi_{lr,t}\rangle.
\end{aligned}$$

Now,

$$\begin{aligned}
\text{Tr } \Lambda_{t+1} \Lambda_{t+1}^\dagger &= 2^p \text{ Tr } (I_p \otimes \Lambda_t \Lambda_t^\dagger) \\
&= 2^p \cdot 2^p \cdot 2^{2m_{A,t}} \\
&= 2^{2m_{A,t+1}}.
\end{aligned}$$

Moreover, for the same reasons as in the proof of Lemma 4.2, the set $\{|\phi_{l,t+1}\rangle\}$ is orthonormal. Thus, the state in equation (11) is of the form stated in the lemma.

Case (b). Bob applies a unitary transformation $V$ to his qubits and sends $p$ qubits to Alice. W.l.o.g., these are the $p$ leftmost qubits.

After the communication, the joint state looks like

$$\sum_{a\in\{0,1\}^{q_{A,t}}} \sum_{l\in\{0,1\}^p} |a\rangle_A|l\rangle_A \ \ (\langle l| \otimes I_{q_{B,t}-p}) V\Lambda_t|\phi_{a,t}\rangle_B,$$

which may be recast as

$$\sum_{a,l} |al\rangle \ \Lambda_{t+1}|\phi_{al,t+1}\rangle, \tag{12}$$

where

$$\begin{aligned} \Lambda_{t+1} &= \sum_{b\in\{0,1\}^p} (\langle b| \otimes I_{q_{B,t}-p}) \ V\Lambda_t \ (\langle b| \otimes I_{q_{B,t}}), \\ |\phi_{al,t+1}\rangle &= |l\rangle|\phi_{a,t}\rangle. \end{aligned}$$

Now $m_{A,t+1} = m_{A,t}$, $m_{B,t+1} = m_{B,t} + p$, $q_{A,t+1} = q_{A,t} + p$, and $q_{B,t+1} = q_{B,t} - p$.

The states $|\phi_{al,t+1}\rangle$ are orthonormal. Moreover,

$$\begin{aligned} &\mathrm{Tr}\ \Lambda_{t+1}\Lambda_{t+1}^\dagger \\ &= \mathrm{Tr} \sum_{b,b'} (\langle b| \otimes I)V\Lambda_t(\langle b| \otimes I)(|b'\rangle \otimes I)\Lambda_t^\dagger V^\dagger(|b'\rangle \otimes I) \\ &= \sum_b \mathrm{Tr}\left[(\langle b| \otimes I) \ V\Lambda_t\Lambda_t^\dagger V^\dagger \ (|b\rangle \otimes I)\right] \\ &= \sum_b \mathrm{Tr}\left[(|b\rangle\langle b| \otimes I) \ V\Lambda_t\Lambda_t^\dagger V^\dagger\right] \\ &= \mathrm{Tr}\left[(I \otimes I) \ V\Lambda_t\Lambda_t^\dagger V^\dagger\right] \\ &= \mathrm{Tr}\ \Lambda_t\Lambda_t^\dagger \ = \ 2^{2m_{A,t+1}}. \end{aligned}$$

Thus, the state in equation (12) is of the form described in the lemma.

This completes the induction step, and the proof. ■

We now sketch how our characterisation of quantum protocols enables us to prove Theorem 1.3(2).

**Proof of Theorem 1.3(2):** The proof is essentially the same as for Theorem 4.7, and we use the same notation here.

Lemma 5.1 shows that Bob's state remains of form similar to that in Lemma 4.6 as he interacts with Alice during the protocol. Let $\{\Lambda|\phi_{x,l}\rangle\}_l$ be Bob's mixed state at the end of the protocol when Alice has input $x \in \{0,1\}^n$, as given by Lemma 5.1. Note that $\Lambda$ is independent of $x$.

Since we may assume that all the ancillary qubits used by Bob are included in his state above (cf. the proof of Lemma 5.1), the projection operators $P_y$ are over $q_B$ qubits. Now, as before,

$$\Pr[C] \ = \ \sum_x p_x \ \Pr[C_x] \ = \ \sum_{x,l} p_x \ \| \ P_x\Lambda|\phi_{x,l}\rangle \ \|^2 ,$$

where $p_x = \Pr[X = x]$. Furthermore,

$$\begin{aligned}\sum_{x,l} \| P_x \Lambda |\phi_{x,l}\rangle \|^2 &= \sum_{x,l,j} |\langle e_{x,j}|\Lambda|\phi_{x,l}\rangle|^2 \\ &\leq \sum_{x,j} \left\| \Lambda^\dagger |e_{x,j}\rangle \right\|^2 \\ &= \mathrm{Tr}\, \Lambda\Lambda^\dagger = 2^{2m_A}.\end{aligned}$$

Combining these with Lemma 3.3, we get $\Pr[C] \leq P(X, 2^{2m_A})$. ■

# 6 Discussion

Since the two results involve different measures for the amount of classical information transmitted over a quantum channel, the strength of Theorem 1.3 is in general incomparable to that of the Holevo bound [17] (or its extensions). For example, the Holevo bound is more appropriate for deriving an optimal bound for *random access codes* [1]. This bound has implications in both classical and quantum complexity theory. For instance, Kerenidis and de Wolf use it to prove an exponential lower bound on the length of *locally decodable codes* [18], and Klauck uses it in deriving a bound for one-way communication and formula size in terms of VC-dimension [21]. On the other hand, Theorem 1.3 allows us to derive tight lower bounds for the communication complexity of Inner Product (as explained below). Recently, Ronald de Wolf applied our results to re-derive the best known lower bounds for *matrix rigidity* [15]. Thus, the two bounds—Holevo and ours—provide complementary views of the limitations of quantum communication that have proved to be of use in a variety of contexts.

In typical applications in complexity theory, the Holevo bound is used in a weaker form, which may be derived from our results. Here the messages are distributed uniformly, and the mutual information $I(X : Y)$ is bounded by $m$, the number of qubits used in the quantum encoding of $X$. Note that

**Lemma 6.1** *For any random variables $X, Y$ over $\{0,1\}^n$, there is a decoding procedure $\mathcal{D}$ such that*

$$\Pr[\mathcal{D}(Y) = X] \geq 2^{-H(X|Y)},$$

*where $H(X|Y)$ is the conditional Shannon entropy of $X$ with respect to $Y$.*

**Proof:** We describe a natural decoding procedure $\mathcal{D}$ and then show that it satisfies the requirement of the theorem. On input $y$, the decoding algorithm outputs $x$ such that $p_{x|y} = \max_{x'} p_{x'|y}$, where $p_{x|y} = \Pr[X = x|Y = y]$. Let $p_y^{\max}$ denote this probability and let $x_y$ denote the corresponding $x$.

We claim that the procedure $\mathcal{D}$ described above decodes correctly with probability at least $2^{-H(X|Y)}$. The probability of correct decoding is equal to

$$\begin{aligned}&\Pr[\mathcal{D}(Y) = X] \\ &= \sum_y \Pr[X = x_y|Y = y] \cdot \Pr[Y = y] \\ &= \mathrm{E}[p_Y^{\max}].\end{aligned}$$

Now, $H(X|Y=y) = -\sum_x p_{x|y} \log p_{x|y} \geq -\log p_y^{\max}$. So $p_y^{\max} \geq 2^{-H(X|Y=y)}$. Taking expectation over $Y$, and noting that $f(\alpha) = 2^{-\alpha}$ is a convex function, we have

$$\begin{aligned} \mathrm{E}[p_Y^{\max}] &\geq \mathrm{E}[2^{-H(X|Y=y)}] \\ &\geq 2^{-\mathrm{E}[H(X|Y=y)]} \\ &= 2^{-H(X|Y)}, \end{aligned}$$

which gives us the claimed lower bound on the decoding probability. ∎

Together with Theorem 1.1, Lemma 6.1 implies that when $X$ is distributed uniformly, and $Y$ is obtained by measuring an $m$-qubit encoding of it, the mutual information $I(X:Y)$ is at most $m$. Our bound thus obviates the need for a translation of in-probability statements into statements about mutual information in cases where the above weak form of the Holevo theorem is applied. In fact, our bound on decoding probability can give us sharper bounds on the number of qubits used in an encoding than an application of Holevo's theorem. We illustrate this with a toy example, and then with a more interesting example from quantum communication complexity.

Consider an encoding of $n$-bits into $n+1$ orthogonal states $|i\rangle, i = 0, 1, \ldots, n$. Half the strings are encoded as $|0\rangle$, a fourth as $|1\rangle$, an eighth as $|2\rangle$, and so on. A random codeword from this code can be decoded with probability exactly $(n+1)2^{-n}$. Theorem 1.1 tells us that $\log(n+1)$ qubits are necessary for any encoding achieving this probability of success. On the other hand, the mutual information between the original $n$ bits and its encoding is

$$\frac{1}{2} + \frac{2}{2^2} + \frac{3}{2^3} + \cdots + \frac{n}{2^n} + \frac{n}{2^n},$$

which sums up to $2 - 2^{-(n-1)}$. Thus, the mutual information with any *decoding* is also at most 2. (The mutual information inferred from the probability of correct decoding is even smaller.) This gives us a lower bound of at most 2 for the number of qubits in the encoding, when combined with Holevo's theorem.

Finally, we may apply Theorem 1.2 to obtain an improved lower bound for the entanglement-assisted quantum communication complexity of the inner product function $\mathrm{IP}_n$. (The inner product function $\mathrm{IP}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as $\mathrm{IP}_n(x,y) = \oplus_i (x_i \wedge y_i)$.) The connection between the two is provided by the following reduction due to Cleve *et al.*

**Theorem 6.2 (Cleve, van Dam, Nielsen, Tapp [13])** *If $Q_\epsilon^*(\mathrm{IP}_n) = m$, then there is an entanglement-assisted protocol for transmitting $n$ bits with probability of success at least $(1-2\epsilon)^2$, such that the total communication from* each *party to the other, over all the rounds of communication, is $m$ qubits.*

Theorem 1.2 now implies

**Corollary 6.3** $Q_\epsilon^*(\mathrm{IP}_n) \quad \geq \quad \frac{1}{2}n - \log \frac{1}{1-2\epsilon}$.

The previous best lower bound was $\frac{1}{2}((1-2\epsilon)^2 n - 1)$ due to Cleve *et al.* [13].

For any $\epsilon < 1/2$, there is a straightforward public-coin randomised protocol for $\mathrm{IP}_n$ with communication cost at most $\left\lceil n - \log \frac{1}{1-2\epsilon} \right\rceil + 1$. Along with the superdense coding scheme in Ref. [8], this means that

**Theorem 6.4** $Q_\epsilon^*(\mathrm{IP}_n) \quad \leq \quad \left\lceil \frac{1}{2}(n - \log \frac{1}{1-2\epsilon} + 1) \right\rceil$.

Thus, our lower bound for Inner Product is close to optimal, and for constant error, is within an additive $O(1)$ term of the upper bound. Since $Q_{1/3}(\mathrm{IP}_n) \leq n$, this provides more evidence that prior entanglement does not give us a saving of more than a factor 2 plus perhaps an additive term of $O(\log n)$ in communication cost. Proving this statement for an arbitrary boolean function however remains open.

Finally, observe that the reduction in Ref. [13] cited above can be adapted immediately so that a one-way protocol for Inner Product for $n$-bit inputs, with communication complexity $m$ is transformed into a *one-way* protocol for transmitting $n$ bits. As before, an error probability of $\epsilon$ in the original protocol results in an error of $(1-2\epsilon)^2$ in the new protocol. Moreover, the reduction extends, with the same features, if the original protocol is guaranteed to work with probability of error $\epsilon$ under the uniform distribution over inputs. Theorem 1.1 now gives us a factor of two improvement over the two-way communication complexity without prior entanglement.

**Corollary 6.5** *The $\epsilon$-error quantum one-way communication complexity of Inner Product is at least $n - 2\log\frac{1}{1-2\epsilon}$. The same lower bound continues to hold for $\epsilon$-error quantum one-way protocols under the uniform distribution over $n$-bit inputs.*

The one-way complexity of Inner Product is the basis of a proof of security due to Ben-Or [4] of the Bennett-Brassard quantum key distribution protocol [5]. It shows that "privacy amplification" via two-wise independent hashing is effective even in the presence of a quantum adversary. We briefly describe this application below to round up our discussion.

Two parties, Alice and Bob, share a common, uniformly random $n$-bit string $X$. Suppose that an adversary, Eve, has acquired some information about the random variable $X$ during the course of its transmission. This information is modeled as a quantum state $\rho_X$ over $m$ qubits that depends arbitrarily on the entire string $X$. In the Bennett-Brassard scheme, this is ensured by the testing/error-estimation stage of the protocol. The goal of privacy amplification is to extract a common random $k$-bit string $U$ from $X$ about which Eve has very little information (in a sense to be made precise shortly). Alice and Bob have, at their disposal, a *public* classical channel to assist in the extraction process. Thus, all their communication is visible to Eve.

Bennett and Brassard proposed a random hashing scheme for privacy amplification. In their scheme, the string $U$ is taken to be $A \cdot X$, where $A$ is a uniformly random $k \times n$ matrix over $\mathbb{Z}_2$ sent by Alice to Bob. This is equivalent to taking the Inner Product of $X$ with $k$ uniformly random $n$-bit strings. Eve, the adversary, may perform a measurement on her state $\rho_X$, possibly depending on the matrix $A$, to gain information about $U$. We would like that on average, the distribution of $U$ conditioned upon the random matrix $A$, and Eve's measurement outcome, be close to uniform on $k$ bits.

Using a lower bound for the one-way quantum communication complexity of Inner Product, Ben-Or proved this hashing scheme to be secure. Suppose that the conditional distribution of the extracted string $U$ is $\epsilon$ away from uniform in $\ell_1$ distance, on average over the measurement outcomes of Eve, and over the distribution of $A$. The Vazirani XOR lemma (see, e.g., Ref. [16]) implies that the Inner Product of $U$ with a random non-zero $k$-bit string is $2^{k/2}\epsilon$ away from uniform. Our lower bound for Inner Product now shows that that $\epsilon$ is at most $2^{(n-m-k)/2}$.

## Acknowledgements

We thank Michael Ben-Or, Leonard Schulman, and Umesh Vazirani for insightful discussions, and Michael Ben-Or also for asking about the one-way communication complexity of Inner Product.

## References


[1] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, 49(4):1–16, July 2002.

[2] A. Ambainis, L. J. Schulman, A. Ta-Shma, U. Vazirani, and A. Wigderson. Quantum communication complexity of sampling. *SIAM Journal on Computing*, 32:1570–1585, 2003.

[3] Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proceedings of the thirty-sixth Annual ACM Symposium on Theory of Computing*, pages 128–137, New York, NY, USA, 2004. ACM Press.

[4] M. Ben-Or. Simple security proof for quantum key distribution. 1999. Unpublished manuscript. See talk given during the MSRI special semester on quantum computation in 2002: http://www.msri.org/publications/ln/msri/2002/qip/ben-or/1/index.html.

[5] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, page 175. IEEE Press, 1984.

[6] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993.

[7] C. H. Bennett and P. W. Shor. Quantum information theory. *IEEE Transactions on Information Theory*, IT-44(6):2724–2742, 1998.

[8] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69:2881–2884, 1992.

[9] E. Bernstein and U. V. Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[10] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 63–68, New York, NY, USA, 1998. ACM Press.

[11] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity*, pages 120–130, Los Alamitos, CA, USA, 2001. IEEE Press.

[12] R. Cleve and H. Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997.

[13] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. In *Quantum Computing and Quantum Communications, Proceedings of the 1st NASA International Conference*, volume 1509 of *Lecture Notes in Computer Science*, pages 61–74, Heidelberg, Germany, 1998. Springer-Verlag.

[14] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA, 1991.

[15] R. de Wolf. Lower bounds on matrix rigidity via a quantum argument. Technical report, Arxiv.org Preprint Archive, 2005. Available at `http://www.arxiv.org/abs/quant-ph/0505188` .

[16] O. Goldreich. Three XOR-lemmas – an exposition. Technical Report TR95-056, Electronic Colloquium on Computational Complexity, 1995. Available at http://eccc.uni-trier.de/eccc/.

[17] A. Holevo. Some estimates of the information transmitted by quantum communication channels. *Problems of Information Transmission*, 9(3):177–183, 1973. Russian version in *Problemy Peredachi Informatsii 9* (1973), 3–11.

[18] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes. *Journal of Computer and System Sciences*, 69(3):395–420, 2004. Special issue for STOC 2003.

[19] H. Klauck. On quantum and probabilistic communication: Las Vegas and one-way protocols. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 644–651, New York, NY, USA, 2000. ACM Press.

[20] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 288–297, Los Alamitos, CA, USA, 2001. IEEE Press.

[21] H. Klauck. One-way communication complexity and the Neciporuk lower bound on formula size. Technical report, Arxiv.org Preprint Archive, 2001. Available at `http://www.arxiv.org/abs/cs.CC/0111062` .

[22] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of Set Disjointness. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 124–133, New York, NY, USA, 2001. ACM Press.

[23] I. Kremer. Quantum communication. Master's thesis, The Hebrew University of Jerusalem, Jerusalem, Israel, 1995.

[24] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, Cambridge, UK, 1997.

[25] H.-K. Lo and H. F. Chau. Is quantum bit commitment really possible? *Physical Review Letters*, 78:3410–3413, 1997.

[26] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78(17):3414–3417, 1997.

[27] A. Nayak. *Lower Bounds for Quantum Computation and Communication*. PhD thesis, University of California, Berkeley, 1999.

[28] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual IEEE Symposium on Foundations of Computer Science*, pages 369–376, Los Alamitos, CA, USA, 1999. IEEE Press.

[29] A. Nayak and J. Salzman. On communication over an entanglement-assisted quantum channel. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 698–704, New York, NY, USA, 2002. ACM Press. Appeared in the Joint Session with the *17th Annual IEEE Conference on Computational Complexity*, 2002.

[30] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, UK, 2000.

[31] J. Preskill. Quantum computation. Lecture Notes, available at `http://www.theory.caltech.edu/people/preskill/ph229/`, California Institute of Technology, Pasadena, CA, 1998.

[32] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 358–367, New York, NY, USA, 1999. ACM Press.

[33] R. Raz. Quantum information and the PCP theorem. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 459–468, Los Alamitos, CA, USA, 2005. IEEE Press.

[34] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003. Russian version in *Izvestiya Rossiiskoi Academii Nauk (seriya matematicheskaya) 67* (2003), 1, 159–176.

[35] W. van Dam and P. Hayden. Renyi-entropic bounds on quantum communication. Technical report, ArXiv.org Preprint Archive, 2002. Available at `http://www.arxiv.org/abs/quant-ph/0204093`.

[36] A. C.-C. Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, Los Alamitos, CA, USA, 1993. IEEE Press.

# A Shared randomness in classical communication

In this section, we briefly review the role of shared randomness in sending classical information from one party to another. Consider a scenario where two parties, Alice and Bob, are given a common random variable $R$. Alice wishes to relay a classical message, given by a random variable $X$ independent of $R$, to Bob. They engage in a classical communication protocol (see, e.g., Ref. [24] for a formal definition of a protocol) in which Alice sends $m_A$ bits to Bob, and Bob has a transcript $T$ at the end of the protocol. We make several observations about the protocol.

First, note that we may convert any multiple round communication protocol to an equivalent protocol in which there is only one message, that from Alice to Bob, of length $m_A$. This is because Bob's communication in any round is a *deterministic* function of the shared random variable $R$, and the history of messages sent till that point. Alice can compute all of Bob's communication bits herself and thus her portion $T_A$ (which is of length $m_A$) of the transcript $T$. In the one message protocol, she sends Bob $T_A$. Bob now reproduces $T$ from $T_A, R$.

Second, we may bound the mutual information between Bob's bits $R, T$ and Alice's input $X$ by $m_A$, the number of bits sent by Alice to Bob. Making essential use of the fact that the input $X$ is independent of

the previously shared random variable $R$, we have:[1]

$$\begin{aligned} I(X : RT) &\leq I(X : RT_A) \\ &= I(RX : T_A) + I(X : R) - I(R : T_A) \\ &= I(RX : T_A) - I(R : T_A) \\ &\leq I(RX : T_A) \\ &\leq H(T_A) \\ &\leq m_A. \end{aligned}$$

Thus, we see that at most $m_A$ "bits of information" may be transmitted using that number of classical bits. Twice as many bits of information may be transmitted with quantum communication using prior shared EPR pairs, via superdense coding [8].

Finally, we may also bound the probability that Bob is able to recover the input $X$ correctly from his view of the protocol. Suppose $Y$ is Bob's guess for $X$. Then,

$$\begin{aligned} \Pr[Y = X] &= \sum_r \Pr[Y = X | R = r] \cdot \Pr[R = r] \\ &\leq \max_r \Pr[Y = X | R = r] \\ &= \max_r \sum_x \Pr[X = x] \cdot \Pr[Y = x | X = x, R = r] \\ &= \max_r \sum_x \Pr[X = x] \cdot \mathbf{1}_{(Y=x|X=x,R=r)}. \end{aligned}$$

Since $Y$ is a deterministic function of $T_A$ and $R$, it assumes at most $2^{m_A}$ different values for a fixed $r$, i.e., $\sum_x \mathbf{1}_{(Y=x|X=x,R=r)} \leq 2^{m_A}$. So $\Pr[Y = X] \leq P(X, 2^{m_A})$, the sum of the $2^m$ largest probability masses in the distribution of $X$. In particular, if $X$ is uniformly distributed, we have $\Pr[Y = X] \leq 2^{m_A}/2^n$. This is exactly the probability of decoding we achieve if we transmit the first $m_A$ bits out of the $n$ in the input. (The recipient guesses the remaining bits uniformly at random.)

[1] We refer the reader to the text [14] for the definition and properties of mutual information used here.