Exponential Separation between Quantum Communication and Logarithm of Approximate Rank

Makrand Sinha†
CWI
The Netherlands
makrand.sinha@cwi.nl

Ronald de Wolf*
QuSoft, CWI and University of Amsterdam
The Netherlands
rdewolf@cwi.nl

***Abstract*—Chattopadhyay, Mande and Sherif [CMS19] recently exhibited a total Boolean function, the sink function, that has polynomial approximate rank and polynomial randomized communication complexity. This gives an exponential separation between randomized communication complexity and logarithm of the approximate rank, refuting the log-approximate-rank conjecture. We show that even the *quantum* communication complexity of the sink function is polynomial, thus also refuting the quantum log-approximate-rank conjecture.**

Our lower bound is based on the fooling distribution method introduced by Rao and Sinha [RS15] for the classical case and extended by Anshu, Touchette, Yao and Yu [ATYY17] for the quantum case. We also give a new proof of the classical lower bound using the fooling distribution method.

***Keywords*-Quantum Communication, Log-rank conjecture, Approximate rank**

I. Introduction

Communication complexity [KN97], [RY18] is a basic model of distributed computing where one only cares about the resource of *communication* between the various distributed parties doing the computation. This is a beautiful and fundamental computational model in its own right, and has many applications to other areas, in particular for lower bounds. For concreteness consider the two-player communication complexity of some Boolean function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Here Alice receives input $x \in \{0,1\}^n$ and Bob receives input $y \in \{0,1\}^n$, and they want to compute $f(x,y)$ with minimal communication between them.

Much research has gone into relating the (deterministic, randomized, nondeterministic, quantum, ...) communication complexity of f to its combinatorial or algebraic properties. In particular, we may consider the relation between the communication complexity and the *rank* (over the reals) of the $2^n \times 2^n$ Boolean matrix M_f whose entries are the values $f(x,y)$. Mehlhorn and Schmidt [MS82] showed that the log of this rank lower bounds the deterministic communication complexity of f, and Lovász and Saks [LS93] conjectured that this lower bound is polynomially tight; in other words, that deterministic communication complexity is upper bounded by a polynomial in the logarithm of the rank of M_f. This *log-rank conjecture* is one of the main open problems in communication complexity and remains wide open. On the one hand, the best upper bound on deterministic communication complexity in terms of rank is roughly the *square root* of the rank [Lov16], [Lov14] (see also [Rot14]). On the other hand, the biggest known gap between deterministic communication complexity and log-rank is only quadratic [GPW15].

One may similarly consider the relation between *randomized* communication complexity (say with private coin flips, and error probability $\leq 1/3$ on every input x, y) and log of the *approximate* rank, which is the minimal rank among all matrices that approximate M_f entrywise up to $1/3$. The log of the approximate rank lower bounds randomized communication complexity (even *quantum* communication complexity with unlimited prior entanglement [BW01]), and Lee and Shraibman [LS09, Conjecture 42] conjectured that this lower bound is polynomially tight. This is known as the *log-approximate-rank* conjecture. Until very recently, the biggest separation known between randomized communication complexity and the log of approximate rank was a fourth power [GJPW17]. But then, in an important breakthrough, Chattopadhyay, Mande and Sherif [CMS19] devised a function, called the *sink* function[1], that refutes this conjecture.[2]

Their function is as follows. Let $n = \binom{t}{2}$. The function

†Supported by the Netherlands Organization for Scientific Research, Grant Number 617.001.351.

*Partially supported by ERC Consolidator Grant 615307-QPROGRESS and by QuantERA project QuantAlgo 680-91-034.

[1]It turns out that the sink function was already discovered in another context in the 1970s: Aanderaa introduced it as a counterexample to a conjecture of Rosenberg (c.f. [BEL74]). The unpublished report [BEL74] also introduced several other functions that may equally well be used for separating randomized communication and approximate rank following the ideas introduced by Chattopadhyay *et al.* [CMS19].

[2]Their function is a so-called *XOR function*, of the form $f(x,y) = g(x \oplus y)$ for some n-bit Boolean function g, and thus even refutes the special case of the log-approximate-rank conjecture restricted to XOR functions. This special case has received much attention recently [TWXZ13], [Zha14], [HHL16] (in part thanks to the fact that the rank of M_f equals the *Fourier sparsity* of g), and remains open.

$\mathsf{sink} : \{0,1\}^n \to \{0,1\}$ is defined on the edges of the complete graph on the vertex set $[t]$.[3] For each edge $e \in \binom{[t]}{2}$, the corresponding input bit z_e assigns an orientation to the edge e (such an oriented complete graph is called a *tournament*). The function $\mathsf{sink}(z) = 1$ iff there is a vertex that is a *sink* (i.e., that has no outgoing edges). Note that a tournament can have at most one sink, since the orientation of the edge between vertices v and w eliminates one of them as a possible sink. The communication problem is defined by Alice and Bob receiving the inputs $x, y \in \{0,1\}^n$ and wanting to compute the function $\mathsf{sink}(x \oplus y)$ where $x \oplus y$ is the bitwise parity. In other words, together they compute the sink function after putting the label $x_e \oplus y_e$ on the edge e. With slight abuse of notation, we denote the $2n$-bit communication function by sink as well.

The approximate rank of the $2^n \times 2^n$ Boolean matrix M_{sink} associated to the sink problem is only polynomial in n, which can be seen as follows. Consider vertex $v \in [t]$, let $N(v)$ denote the set of edges incident on v and let $x_{N(v)}$ (and $y_{N(v)}$) denote the projection of the input x (and y) to the edges in $N(v)$. Let $z_{N(v)} \in \{0,1\}^{t-1}$ be the unique string of orientations that makes v the sink of the graph. Note that v is a sink in the tournament $x \oplus y$ iff $x_{N(v)} = y_{N(v)} \oplus z_{N(v)}$. The latter problem corresponds to a (shifted) equality problem on strings of $t-1$ bits, and it is well known that this problem has a cheap randomized private-coin protocol that uses $O(\log t) = O(\log n)$ bits of communication, that outputs 1 with probability 1 if v is the sink in tournament $x \oplus y$, and outputs 1 with probability $\in [0, 1/(3t)]$ if v is not a sink. This in turns implies the existence of a $2^n \times 2^n$ matrix M_v of rank polynomial in n, whose (x,y)-entry is 1 if v is the sink in $x \oplus y$, and whose (x,y)-entry is $\in [0, 1/(3t)]$ if v is not a sink. Thanks to the fact that at most one of the t vertices is a sink, we can now get a good entry-wise approximation of M_{sink} by just adding up all the M_v-matrices over all $v \in [t]$: the resulting matrix $\widetilde{M} = \sum_{v=1}^{t} M_v$ will have (x,y)-entry $\in [0, 1/3]$ whenever $x \oplus y$ has no sink, and will have (x,y)-entry in $[1, 4/3]$ whenever $x \oplus y$ has a sink (if v is the sink in $x \oplus y$, then M_v contributes 1 to the entry $\widetilde{M}_{xy}$, and the other M_w's together contribute at most $1/3$). By subadditivity of rank, the rank of $\widetilde{M}$ is at most the sum of the ranks of the M_v's, which is polynomial in n. Hence the log of the approximate rank of M_{sink} is $O(\log n)$. In contrast, Chattopadhyay *et al.* show that the randomized communication complexity of the sink function is exponentially bigger:

Theorem I.1 ([CMS19])**.** *The $1/3$-error randomized communication complexity of the function* sink *on $n = \binom{t}{2}$ bits is $\Omega(t) = \Omega(\sqrt{n})$.*

This lower bound is optimal even for *deterministic* protocols: by looking at one edge, Alice and Bob can rule out one vertex from being a sink. Proceeding this way, they read $t-1$ edges until they have eliminated all but one vertex v from being a sink. At this point, they look at the $t-1$ edges incident to v, and find out if v is a sink or not. This gives an $O(t)$-bit deterministic communication protocol, since the parties exchange two bits per edge.

This separation refutes the log-approximate-rank conjecture, showing that randomized communication complexity is not always upper bounded by polylog of the approximate rank. However, *quantum* communication complexity can be much smaller than randomized communication complexity: polynomial gaps are known for some total functions [BCW98], [AA05], [ABBD$^+$16] and exponential gaps are known for some partial functions [Raz99], [KR11]. Thus one might still entertain the weaker conjecture that quantum communication complexity is upper bounded by polylog of the approximate rank, and indeed Lee and Shraibman [LS09, Conjecture 57] made this conjecture explicitly. Prior to this work, the biggest separation known between quantum communication complexity and log of the approximate rank, was only quadratic [ABDG$^+$17]. Indeed, one of the main problems left open by Chattopadhyay *et al.* asks about the quantum communication complexity of the sink function. If this is large then it would refute the quantum log-approximate-rank conjecture, but if it is small then it would provide the first superpolynomial separation between quantum and classical communication complexity for a total function. We answer their open question by proving a polynomial lower bound on the quantum communication complexity of the sink function, thus refuting the quantum log-approximate-rank conjecture:

Theorem I.2. *The $1/3$-error quantum communication complexity of the function* sink *on $n = \binom{t}{2}$ bits is $\Omega(t^{1/3}) = \Omega(n^{1/6})$.*

As Chattopadhyay *et al.* noted, the quantum communication complexity of the sink function is polynomially smaller than the randomized complexity: using Grover's algorithm [Gro96] to search for a sink, combined with an efficient low-error equality protocol to test whether a specific vertex is a sink, one gets an $\widetilde{O}(\sqrt{t})$-qubit protocol. We suspect that this upper bound is tight up to the log-factor, and that our quantum lower bound should be improvable.

Independent Work: In independent and simultaneous work, Anshu, Boddu and Touchette [ABT19] obtained the same $\Omega(t^{1/3})$ lower bound using a reduction to quantum information complexity of the equality function, but our techniques to prove Theorem I.2 are different, as we describe below.

[3] We use t for the number of vertices in the graph instead of m as used in [CMS19].

Proof Outline

Our approach to proving Theorem I.2 is to first give an alternate and arguably simpler proof of Theorem I.1 using the fooling distribution method (and other tools) introduced by Rao and the first author in [RS15], and then we show that the same approach can be used to give a (weaker) quantum lower bound using tools from a paper by Anshu, Touchette, Yao and Yu [ATYY17], which generalized some of the techniques used in [RS15] to the quantum setting. Our proofs are relatively straightforward and short given the tools in these papers. Below we give a high-level outline.

Let us look at the classical case first. To prove a lower bound on the randomized communication complexity, it suffices to give a distribution on the inputs that is hard for deterministic protocols. Let $p_0(X,Y)$ denote the uniform distribution on 0-inputs to sink and $p_1(X,Y)$ denote the uniform distribution on 1-inputs to sink. Our hard distribution for deterministic protocols will be the distribution which samples from $p_0(X,Y)$ with probability $\frac{1}{2}$ and from $p_1(X,Y)$ with probability $\frac{1}{2}$. Note that the messages of any low-error protocol look very different under these two distributions: $p_0(M)$ and $p_1(M)$ have statistical distance close to 1, where $p_b(M)$ denotes the distribution induced on the messages under $p_b(X,Y)$ for $b \in \{0,1\}$.

To show that this is a hard distribution for deterministic protocols, we show that there is another distribution $u(X,Y)$ such that for any protocol with communication at most ϵt, the induced distribution $u(M)$ on the messages satisfies $u(M) \approx p_0(M)$ as well as $u(M) \approx p_1(M)$, where $\approx$ denotes closeness in statistical distance. This in turn implies that $p_0(M) \approx p_1(M)$ for small-communication protocols, giving us a lower bound on communication. Such a distribution $u(X,Y)$ is called a *fooling* distribution.

The fooling distribution $u(X,Y)$ for sink will just be the uniform distribution on $\{0,1\}^{n+n}$. Note that under the uniform distribution $u(X,Y)$, the function sink takes value 0 with probability $1-2^{-\Omega(t)}$, and since $p_0(X,Y) = u(X,Y|\mathsf{sink} = 0)$, the input distributions $p_0(X,Y)$ and $u(X,Y)$ are already very close in statistical distance, and so are the corresponding distributions on the messages. The interesting part is to argue that the message distribution $p_1(M) \approx u(M)$ even though the respective input distributions $p_1(X,Y)$ and $u(X,Y)$ are actually very far apart. For this purpose, let us first note that the distribution $p_1(X,Y)$ can be generated from $u(X,Y)$ by first picking a uniformly random vertex v as the sink and conditioning on the event that $X_{N(v)} = Y_{N(v)} \oplus z_{N(v)}$ (recall that $N(v)$ is the set of edges incident on v, $X_{N(v)}$ and $Y_{N(v)}$ are projections of X and Y to the edges in $N(v)$, and $z_{N(v)}$ is the unique string that encodes the orientation of the edges for which vertex v is the sink).

To argue that $p_1(M) \approx u(M)$, first one can use Shearer's inequality (see Lemma II.7) to conclude that under the distribution $u(X,Y)$, the messages M reveal only a small amount of information about $X_{N(v)}$ and $Y_{N(v)}$ for a random vertex v. In particular, since an edge appears in $N(v)$ with probability $2/t$ for a random v, one would expect M to reveal at most $(2/t) \cdot |M| \leq \epsilon$ bits of information about $X_{N(v)}$ and $Y_{N(v)}$ each (this is also the reason for working with the fooling distribution: since all the inputs are independent of each other, one may use Shearer's inequality). Now to relate the fooling distribution $u(X,Y)$ to the input distribution $p_1(X,Y)$ we need to condition on the event $X_{N(v)} = Y_{N(v)} \oplus z_{N(v)}$. A lemma from [RS15] (see Lemma III.3 in Section III) exactly captures this situation and says that conditioning on such a collision event, when the messages reveal little information about the colliding variables, does not change the distribution of the messages too much, so we can conclude that $p_1(M) \approx u(M)$.

The proof for the quantum case proceeds more or less analogously. It is still true that the output of a low-error quantum protocol must look very different under distributions supported only on 0-inputs and 1-inputs respectively. We show that $u(X,Y)$ is still a fooling distribution for small-communication quantum protocols. As in the classical case, it is easy to argue using a quantum version of Shearer's inequality (see Lemma II.24) that small-communication quantum protocols do not reveal too much information about $X_{N(v)}$ and $Y_{N(v)}$ for a random vertex v under the fooling distribution $u(X,Y)$. To condition on the collision event $X_{N(v)} = Y_{N(v)} \oplus z_{N(v)}$, we use a lemma from [ATYY17] (see Lemma IV.2 in Section IV) which allows us to argue that for a typical vertex v, conditioning on the collision event does not change the output too much. So, it must be the case that for a small-communication quantum protocol, the output on an input distribution where v is the sink (for a typical v) must be close to the output when the input distribution is $p_0(X,Y)$. This implies that small-communication quantum protocols for the sink function must have large error.

Organization: We introduce preliminaries on information theory, quantum information theory and communication complexity in the next section (Section II). Section III contains the proof described above for the classical case. The quantum lower bound is given in Section IV.

II. Preliminaries

A. Classical Probability Theory

Probability Spaces and Variables: Throughout this paper, $\log$ denotes the logarithm taken in base two. We use $[k]$ to denote the set $\{1,2,\ldots,k\}$ and $[k]^{<n}$ to denote the set of all strings of length less than n over the alphabet $[k]$, including the empty string. The notation $|z|$ denotes the length of the string z.

Random variables are denoted by capital letters (e.g. A) and values they attain are denoted by lower-case letters

(e.g. a). Events in a probability space will be denoted by calligraphic letters (e.g. $\mathcal{E}$). Given $a = (a_1, a_2, \ldots, a_n)$, we write $a_{\leq i}$ to denote $a_1, \ldots, a_i$. We define $a_{<i}$ similarly. We write a_S to denote the projection of a to the coordinates specified in the set $S \subseteq [n]$.

Given a probability space p and a random variable A in the underlying sample space, we use the notation $p(A)$ to denote the probability distribution of the variable A in the probability space p. We will often consider multiple probability spaces with the same underlying sample space, so for example $p(A)$ and $q(A)$ will denote the distribution of the random variable A under the probability spaces p and q, respectively, with the underlying sample space of p and q being the same. We write $p(A|b)$ to denote the distribution of A conditioned on the event $B = b$. We write $p(a)$ to denote the number $\mathbb{P}_p[A = a]$ and $p(a|b)$ to denote the number $\mathbb{P}_p[A = a|B = b]$. Given a distribution $p(A, B, C, D)$, we write $p(A, B, C)$ to denote the marginal distribution on the variables A, B, C. We often write $p(AB)$ instead of $p(A, B)$ for conciseness of notation. Similarly, $p(a, b, c)$ will denote the probability according to the marginal distribution $p(A, B, C)$ and we will often write it as $p(abc)$ for conciseness.

If $\mathcal{W}$ is an event, we write $p(\mathcal{W})$ to denote its probability according to p. For two events $\mathcal{W}$ and $\mathcal{W}'$, the probability of their intersection $\mathcal{W} \cap \mathcal{W}'$ is denoted by $p(\mathcal{W}, \mathcal{W}')$. Given a probability space p and a random variable A, when we write $A \in \mathcal{W}$ for an event $\mathcal{W}$ we only consider events in the space of values taken by the variable A.

Given a fixed value c, we denote by $\mathbb{E}_{p(b|c)}[g(a, b, c)] := \sum_b p(b|c) \cdot g(a, b, c)$, the expected value of the function $g(a, b, c)$ under the distribution $p(B|c)$. If the probability space p is clear from the context, then we will just write $\mathbb{E}_{b|c}[g(a, b, c)]$ to denote the expectation. For a Boolean function $h(a, b)$ and a probability distribution $p(A, B)$, we use $\mathbf{1}[h(a, b) = 0]$ to denote the indicator function for the event $h(a, b) = 0$, and we write $p(h = 0) := \mathbb{E}_{p(ab)}[\mathbf{1}[h(a, b) = 0]]$ as the probability that h is 0 under inputs drawn from p.

We write $A - M - B$ as a shorthand to say that the random variables A, M and B form a *Markov chain*, or in other words, that A and B are independent given M: $p(amb) = p(m) \cdot p(a|m) \cdot p(b|m)$ for every a, b, m.

To illustrate the notation, consider the following example. Let $A \in \{0, 1\}^2$ be a uniformly distributed random variable in a probability space p. Then, $p(A)$ is the uniform distribution on $\{0, 1\}^2$, and if $a = (0, 0)$ then $p(a) = 1/4$. Let A_1 and A_2 denote the first and second bits of A, then if $B = A_1 + A_2 \bmod 2$, then when $b = 1$, $p(A|b)$ is the uniform distribution on $\{(0, 1), (1, 0)\}$. If $a = (1, 0)$ and $b = 1$, then $p(a|b) = 1/2$ and $p(a, b) = 1/4$. If $\mathcal{E}$ is the event that $A_1 = B$, then $p(\mathcal{E}) = 1/2$. Let $q(A) = p(A|\mathcal{E})$, then $q(A)$ is the uniform distribution on $\{(0, 0), (1, 0)\}$ and $q(A_2)$ is the distribution over the sample space $\{0, 1\}$ which takes the value 0 with probability 1.

Statistical Distance: For two distributions $p(A), q(A)$, the *statistical* (or *total variation*) *distance* $\|p(A) - q(A)\|_{\mathsf{tv}}$ between them is defined to be $\|p(A) - q(A)\|_{\mathsf{tv}} = \max_{\mathcal{Q}} (p(A \in \mathcal{Q}) - q(A \in \mathcal{Q}))$ where $\mathcal{Q}$ ranges over all events. The following propositions are easy to prove.

Proposition II.1. $\|p(A) - q(A)\|_{\mathsf{tv}} = \frac{1}{2}\sum_a |p(a) - q(a)| = \sum_{a:p(a)>q(a)}(p(a) - q(a))$.

We say $p(A)$ and $q(A)$ are ϵ-*close* if $\|p(A) - q(A)\|_{\mathsf{tv}} \leq \epsilon$ and we write it as $p(A) \overset{\epsilon}{\approx} q(A)$.

Proposition II.2. *If $p(AB), q(AB)$ are such that $p(A) = q(A)$, then*

$$\|p(B) - q(B)\|_{\mathsf{tv}} = \mathop{\mathbb{E}}_{p(a)} \left[\|p(B|a) - q(B|a)\|_{\mathsf{tv}}\right].$$

Lemma II.3. *If $\mathcal{E}$ is an event such that $p(\mathcal{E}) = 1 - \delta$, then $\|p(A|\mathcal{E}) - p(A)\|_{\mathsf{tv}} = \delta$.*

Proof: Note that for any $a \notin \mathcal{E}$, $p(a|\mathcal{E}) = 0$ and for $a \in \mathcal{E}$, using Bayes' rule, we get that

$$p(a|\mathcal{E}) = \frac{p(a, \mathcal{E})}{p(\mathcal{E})} = \frac{p(a)}{p(\mathcal{E})} = \frac{p(a)}{1 - \delta}. \tag{1}$$

By Proposition II.1, we have that

$$\begin{aligned}
\|p(A|\mathcal{E}) - p(A)\|_{\mathsf{tv}} &= \frac{1}{2}\sum_{a\in\mathcal{E}} |p(a|\mathcal{E}) - p(a)| \\
&\quad + \frac{1}{2}\sum_{a\notin\mathcal{E}} |p(a|\mathcal{E}) - p(a)| \\
&= \frac{1}{2}\sum_{a\in\mathcal{E}} |p(a|\mathcal{E}) - p(a)| + \frac{\delta}{2} \\
&\overset{(1)}{=} \frac{1}{2}\sum_{a\in\mathcal{E}} \left(\frac{p(a)}{1-\delta} - p(a)\right) + \frac{\delta}{2} \\
&= \frac{1}{2} \cdot \frac{\delta}{1-\delta} \cdot p(\mathcal{E}) + \frac{\delta}{2} = \delta,
\end{aligned}$$

where the second inequality follows from (1). ∎

Divergence and Mutual Information: The *divergence* between distributions $p(A)$ and $q(A)$ is defined to be

$$\mathbf{D}\left(p(A) \,||\, q(A)\right) = \sum_a p(a) \log \frac{p(a)}{q(a)}.$$

In a probability space p, the *mutual information* between A, B conditioned on C is defined as

$$\begin{aligned}
\mathbf{I}_p\left(A : B|C\right) &= \mathop{\mathbb{E}}_{p(bc)} \left[\mathbf{D}\left(p(A|bc) \,||\, p(A|c)\right)\right] \\
&= \mathop{\mathbb{E}}_{p(ac)} \left[\mathbf{D}\left(p(B|ac) \,||\, p(B|c)\right)\right] \\
&= \sum_{a,b,c} p(abc) \log \frac{p(a|bc)}{p(a|c)}.
\end{aligned}$$

Basic Information Theory Facts: The proofs of the following basic facts can be found in the book by Cover and Thomas [CT06]. In the following, p and q are probability spaces (over the same sample space), and A and B are random variables on the underlying sample space.

Proposition II.4. $\mathbf{D}(p(A) \,||\, q(A)) \geq 0$.

Proposition II.5. *If* $A \in \{0,1\}^\ell$, *then* $\mathbf{I}_p(A:B) \leq \ell$.

Proposition II.6 (Pinsker's Inequality)**.**

$$\begin{aligned}&\|p(A) - q(A)\|_{\mathsf{tv}}^2 \\ &\leq \frac{\ln 2}{2} \cdot \mathbf{D}(p(A) \,||\, q(A)) \leq \mathbf{D}(p(A) \,||\, q(A)).\end{aligned}$$

Lemma II.7 (Shearer's Inequality [GKR16])**.** *Let* $A = (A_1, \ldots, A_n)$ *where the* A_i*'s are mutually independent. Let* M *be another random variable and* $S \subseteq [n]$ *be a random set independent of* A *and* M*, such that* $p(i \in S) \leq \mu$ *for every* $i \in [n]$*. Then, we have*

$$\mathbf{I}_p(A_S : M|S) \leq \mu \cdot \mathbf{I}_p(A:M).$$

B. Classical Communication Complexity

The *communication complexity* of a protocol is the maximum number of bits that may be exchanged by the protocol. Communication protocols may use *shared randomness* and henceforth we will refer to such protocols as randomized protocols. We say a randomized protocol computing a Boolean function $f(x,y)$ has error δ, if for every input, the protocol outputs the correct answer with probability at least $1-\delta$, where the probability is over the shared randomness.

We briefly describe some basic properties of communication protocols that we need. For more details see the textbooks [KN97] or [RY18]. For a deterministic protocol π, let $\pi(x,y)$ denote the sequence of messages (i.e., the transcript) of the protocol on inputs x, y. For any transcript m of the protocol, define the events:

$$\begin{aligned}\mathcal{S}_m &= \{x \mid \exists y \text{ such that } \pi(x,y) = m\}, \\ \mathcal{T}_m &= \{y \mid \exists x \text{ such that } \pi(x,y) = m\}.\end{aligned}$$

We then have:

Proposition II.8 (Messages Correspond to Rectangles)**.** *If* m *is a transcript and* x, y *are inputs to a deterministic protocol* π*, then,* $\pi(x,y) = m \iff x \in \mathcal{S}_m \wedge y \in \mathcal{T}_m$.

Proposition II.8 implies:

Proposition II.9 (Markov Property of Protocols)**.** *Let* X *and* Y *be random inputs to a deterministic protocol and let* M *denote the transcript of this protocol. If* X *and* Y *are independent, then* $X - M - Y$.

Lemma II.10 (Errors and Statistical Distance)**.** *Let* $h(x,y)$ *be a boolean function and* $p(X,Y)$ *be a distribution such that* $p(h=0) = p(h=1) = \frac{1}{2}$*. If* π *is a deterministic protocol with messages* M *that computes* h *with error* δ *on the distribution* $p(XY)$*, then* $|p(M|h=0) - p(M|h=1)| \geq 1 - 2\delta$.

Proof: Since $|p(M|h=0) - p(M|h=1)| = \max_{\mathcal{Q}}(p(M \in \mathcal{Q}|h=0) - p(M \in \mathcal{Q}|h=1))$ it suffices to exhibit an event $\mathcal{Q}$ such that $p(M \in \mathcal{Q}|h=0) - p(M \in \mathcal{Q}|h=1) = 1 - 2\delta$. Let $\mathcal{M}_0$ denote the event that the protocol outputs a zero. Then, since $p(h=0) = p(h=1) = \frac{1}{2}$, writing the probability of success in terms of $\mathcal{M}_0$, we have

$$\begin{aligned}1 - \delta &= \frac{p(M \in \mathcal{M}_0|h=0)}{2} + \frac{1 - p(M \in \mathcal{M}_0|h=1)}{2} \\ &= \frac{1}{2} + \frac{p(M \in \mathcal{M}_0|h=0) - p(M \in \mathcal{M}_0|h=1)}{2}.\end{aligned}$$

On rearranging, the above gives us that $p(M \in \mathcal{M}_0|h=0) - p(M \in \mathcal{M}_0|h=1) = 1 - 2\delta$ and hence the statistical distance must be at least $1 - 2\delta$. ∎

C. Quantum Information Theory

Here we briefly state the facts we need from quantum information theory. For details, see the textbooks [Wil13] or [Wat18].

Quantum States and Measurements: Overloading the notation, we use capital letters A, B, etc. to represent registers and use $\mathcal{H}_A, \mathcal{H}_B$, etc. to denote the associated Hilbert spaces. As before, given registers $A = A_1, \ldots, A_n$ and a set $S \subseteq [n]$, we will use A_S to denote the sequence of registers $\{A_i\}_{i \in S}$. For any register A, $|A| = \lceil \log(\dim \mathcal{H}_A) \rceil$ denotes the number of qubits in A. Given a Hilbert space $\mathcal{H}_A$, we use $\{|a\rangle_A\}$ to denote a canonical orthonormal basis, and if A is a single-qubit register we use $\{|0\rangle_A, |1\rangle_A\}$ to denote the computational basis for the Hilbert space $\mathcal{H}_A$. We write U_A to denote a unitary acting on the Hilbert space $\mathcal{H}_A$ corresponding to a register A.

A *density operator* on $\mathcal{H}_A$ is a linear operator from $\mathcal{H}_A$ to $\mathcal{H}_A$ that is positive semi-definite and has a unit trace. The set of all density operators on a Hilbert space $\mathcal{H}_A$ will be denoted by $\mathcal{D}(\mathcal{H}_A)$. Since a linear operator on a finite-dimensional Hilbert space can be described equivalently with a matrix representation, we will use these notions interchangeably.

A *(quantum) state* ρ_A on a register A is a density operator on $\mathcal{H}_A$. A state ρ_A is called *pure* if it has rank 1. For a unit vector $|\psi\rangle_A \in \mathcal{H}_A$ (viewed as a column vector), we denote by $\langle\psi|_A$ its adjoint (a row vector), and by ψ_A the corresponding state $|\psi\rangle\langle\psi|_A$, but we will also sometimes use the vector $|\psi\rangle_A$ to refer to the corresponding pure state. A classical distribution $p(A)$ can be viewed as the diagonal state $\sum_a p(a)|a\rangle\langle a|_A$ and vice versa, so we will refer to any diagonal state as a classical state.

We use $\rho_A \otimes \sigma_B$ to denote the tensor product of ρ_A and σ_B on the Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. We adopt the convention of omitting Identity operators from a tensor product: instead of $U_R \otimes I_A$ or $\langle r|_R \otimes I_A$, we write U_R or $\langle r|_R$ since the subscripts will convey the necessary information.

A state ρ_{XA} is called a *classical-quantum* state with X being the classical register if it is of the form $\rho_{XA} = \sum_x p(x)|x\rangle\langle x|_X \otimes \rho_A^x$ where $p(X)$ is a classical probability distribution and ρ_A^x is a state on the register A.

Given a linear operator M_{AB} on $\mathcal{H}_A \otimes \mathcal{H}_B$, the *partial trace* of M_{AB} over A is defined as

$$\mathsf{Tr}_A(M_{AB}) = \sum_a \langle a|_A M_{AB} |a\rangle_A.$$

The partial trace operation is linear: $\mathsf{Tr}_A(M_{AB}+M'_{AB}) = \mathsf{Tr}_A(M_{AB}) + \mathsf{Tr}_A(M'_{AB})$ and satisfies the following identities: $\mathsf{Tr}_A(M_A \otimes M_B) = \mathsf{Tr}_A(M_A)M_B$ and $\mathsf{Tr}_A(U_B M_{AB}) = U_B \mathsf{Tr}_A(M_{AB})$.

With the above, we can define the notion of a marginal or reduced state: for a bipartite state ρ_{AB}, the *marginal state* ρ_B on the register B is defined as $\rho_B := \mathsf{Tr}_A(\rho_{AB})$. Note that if we have a classical-quantum state ρ_{XA}, then the marginal state ρ_X is a classical state.

Given a state ρ_A we can always consider it as a marginal of a pure state $\rho_{EA} = |\rho_{EA}\rangle\langle\rho_{EA}|_{EA}$ on a larger system. Such a state $|\rho_{EA}\rangle_{EA}$ is called a *purification* of ρ_A. We will adopt the convention of using the same Greek letters to denote the purification: if we say that $|\rho_{EA}\rangle_{EA}$ is a purification with reference register E, then it is a purification of the state ρ_A, that is, $\rho_A = \mathsf{Tr}_E(|\rho_{EA}\rangle\langle\rho_{EA}|_{EA})$. Given a classical state $\rho_X = \sum_x p(x)|x\rangle\langle x|_X$, we define $\sum_x \sqrt{p(x)}|x\rangle_X|x\rangle_X$ to be its *canonical* purification.

A *positive operator valued measurement* (POVM) is a collection $\{\Lambda_i\}_i$ of linear operators acting on a Hilbert space $\mathcal{H}_A$ such that for each i, the operator Λ_i is positive semi-definite, and $\sum_i \Lambda_i = I_A$. The probability that the outcome of applying a POVM on a quantum state $\rho_A \in \mathcal{D}(\mathcal{H})$ is j is given by $\mathsf{Tr}(\Lambda_j \rho_A)$. Given a single-qubit register A, we will specifically be interested in measurement in the computational basis, which corresponds to the POVM $\{|0\rangle\langle 0|_A, |1\rangle\langle 1|_A\}$. Given a state $\rho_A \in \mathcal{D}(\mathcal{H}_A)$, the probability that the measurement outcome is the bit $b \in \{0,1\}$ is $\mathsf{Tr}(|b\rangle\langle b|_A \rho_A)$.

We say that U_{XA} is a unitary with X as a *control* register if $U_{XA} = \sum_x |x\rangle\langle x|_X \otimes U_A^x$ for some unitary U_A^x's. Also, note that in this case $U_{XA}^\dagger$ is a unitary controlled by X as well.

Distance Measures: Recall that the trace norm $\|M\|_1$ of a matrix M is defined as $\|M\|_1 = \mathsf{Tr}\sqrt{M^\dagger M}$. Equivalently, $\|M\|_1$ is the sum of the singular values of M. Then, the *trace distance* between two quantum states ρ_A and σ_A is defined as $\|\rho_A - \sigma_A\|_1$. We say two states ρ_A and σ_A are ϵ-*close* in trace norm if $\|\rho_A - \sigma_A\|_1 \leq \epsilon$, and write this as $\rho_A \overset{\epsilon}{\approx} \sigma_A$.

The *fidelity* between two quantum states is defined as $F(\rho_A,\ \sigma_A) = \left\|\sqrt{\rho_A}\sqrt{\sigma_A}\right\|_1$ (note that some papers define fidelity as the square of our definition). If ρ_A and σ_A are pure states, then their fidelity is just the absolute value of the inner product of the corresponding vectors. The *Hellinger distance* between the states is $\mathfrak{h}(\rho_A,\ \sigma_A) = \sqrt{1 - F(\rho_A,\ \sigma_A)} = \sqrt{1 - \left\|\sqrt{\rho_A}\sqrt{\sigma_A}\right\|_1}$. If $\psi_A = |\psi\rangle\langle\psi|_A$ is a pure state, for brevity we will sometimes write $\mathfrak{h}(|\psi\rangle_A,\ |\sigma\rangle_A)$ (or $F(|\psi\rangle_A,\ |\sigma\rangle_A)$) to mean $\mathfrak{h}(\psi_A,\ \sigma_A)$ (or $F(\psi_A,\ \sigma_A)$). The Hellinger distance is a metric and in particular satisfies the triangle inequality: $\mathfrak{h}(\rho_A,\ \sigma_A) \leq \mathfrak{h}(\rho_A,\ \psi_A) + \mathfrak{h}(\psi_A,\ \sigma_A)$.

The trace distance and Hellinger distance are both invariant under applying unitaries and decrease under taking marginals:

Proposition II.11. *Given unitaries U_A and V_A, it holds that*

$$\left\|U_A(\rho_A - \sigma_A)V_A^\dagger\right\|_1 = \|\rho_A - \sigma_A\|_1 \text{ and}$$
$$\mathfrak{h}\left(U_A\rho_A V_A^\dagger,\ U_A\sigma_A V_A^\dagger\right) = \mathfrak{h}(\rho_A,\ \sigma_A).$$

Proposition II.12. $\|\rho_A - \sigma_A\|_1 \leq \|\rho_{AB} - \sigma_{AB}\|_1$ *and* $\mathfrak{h}(\rho_A,\ \sigma_A) \leq \mathfrak{h}(\rho_{AB},\ \sigma_{AB})$.

The Hellinger and trace distance are related in the following way:

Proposition II.13. *For quantum states ρ_A and σ_A, it holds that*

$$\mathfrak{h}(\rho_A,\ \sigma_A)^2 \leq \frac{1}{2}\|\rho_A - \sigma_A\|_1 \leq \sqrt{2}\,\mathfrak{h}(\rho_A,\ \sigma_A).$$

The trace distance normalized by 2 is the largest probability difference a POVM could produce between the two states, which is the quantum generalization of total variation distance:

Proposition II.14. *For states ρ_A and σ_A in $\mathcal{D}(\mathcal{H}_A)$, it holds that*

$$\frac{1}{2}\|\rho_A - \sigma_A\|_1 = \max_\Lambda \mathsf{Tr}(\Lambda(\rho_A - \sigma_A)),$$

where Λ ranges over all positive semi-definite operators over $\mathcal{H}_A$ that have eigenvalues at most one.

Proposition II.15 (Uhlmann's Theorem)**.** *Let $|\rho\rangle_{EA}$ and $|\sigma\rangle_{EA}$ be pure states. Then, we have*

$$F(\rho_A,\ \sigma_A) = \max_{U_E} F(U_E|\rho\rangle_{EA},\ |\sigma\rangle_{EA}),$$

or equivalently,

$$\mathfrak{h}(\rho_A,\ \sigma_A) = \min_{U_E} \mathfrak{h}(U_E|\rho\rangle_{EA},\ |\sigma\rangle_{EA}),$$

where U_E ranges over all unitaries acting on the register E.

The unitary U_E which minimizes the Hellinger distance in Uhlmann's theorem is the one for which $\sqrt{\rho_E}\sqrt{\sigma_E}U_E$ is positive semidefinite (such a unitary is always guaranteed to exist) but we will only need the following simple case:

Proposition II.16. *Let $p(X,Y)$ and $q(X,Y)$ be distributions such that $p(X) = q(X)$. Then for the quantum states $|\rho\rangle_{X\overline{X}Y\overline{Y}} = \sum_{xy}\sqrt{p(x,y)}|xxyy\rangle_{X\overline{X}Y\overline{Y}}$*

and $|\sigma\rangle_{X\overline{X}Y\overline{Y}} = \sum_{xy} \sqrt{q(x,y)}|xxyy\rangle_{X\overline{X}Y\overline{Y}}$*, there exists a unitary* $W_{XY\overline{Y}}$ *with* X *as a control register such that* $W_{XY\overline{Y}}|\rho\rangle_{X\overline{X}Y\overline{Y}} = |\sigma\rangle_{X\overline{X}Y\overline{Y}}$.

The above is a special case of Uhlmann's Theorem as $\rho_{\overline{X}} = \sigma_{\overline{X}}$ but one can explicitly take $W_{XY\overline{Y}} = \sum_x |x\rangle\langle x|_X \otimes U^x_{Y\overline{Y}}$ where $U^x_{Y\overline{Y}}$ is any unitary that maps the vector $\sum_y \sqrt{p(x,y)}|yy\rangle_{Y\overline{Y}}$ to $\sum_y \sqrt{q(x,y)}|yy\rangle_{Y\overline{Y}}$.

Quantum Divergence and Mutual Information: The *divergence* (or *relative entropy*) between two quantum states $\rho_A, \sigma_A \in \mathcal{D}(\mathcal{H}_A)$ is defined as

$$\mathbf{D}\left(\rho_A \,||\, \sigma_A\right) = \mathsf{Tr}(\rho_A \log \rho_A) - \mathsf{Tr}(\rho_A \log \sigma_A).$$

Note that the divergence between two states ρ_A and σ_A is always non-negative, and equal to zero iff $\rho_A = \sigma_A$. The *quantum mutual information* of the bipartite state ρ_{AB} is defined as

$$\mathbf{I}_\rho\left(A:B\right) = \mathbf{D}\left(\rho_{AB} \,||\, \rho_A \otimes \rho_B\right). \tag{2}$$

For a tripartite quantum state $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, the *conditional quantum mutual information* is defined as $\mathbf{I}_\rho\left(A:B|C\right) = \mathbf{I}_\rho\left(A:BC\right) - \mathbf{I}_\rho\left(A:C\right)$. For empty C, this equals the definition of mutual information in (2).

It follows from the non-negativity of divergence that quantum mutual information is also non-negative, but it turns out that even *conditional* mutual information is non-negative:

Proposition II.17 (Strong subadditivity)**.** $\mathbf{I}_\rho\left(A:B|C\right) \geq 0$.

Proposition II.18 (Chain Rule)**.** $\mathbf{I}_\rho\left(A:BC\right) = \mathbf{I}_\rho\left(A:C\right) + \mathbf{I}_\rho\left(A:B|C\right)$.

Proposition II.19. $\mathbf{I}_\rho\left(A:B|C\right) \leq 2\min\{|A|,|B|\}$.

Proposition II.20. *If* $\rho_{AB} = \rho_A \otimes \rho_B$*, then* $\mathbf{I}_\rho\left(A:B\right) = 0$.

Basic Lemmas about Divergence and Mutual Information: Below $\rho_{ABC}, \sigma_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ and U_B is a unitary acting on B.

Proposition II.21 (Pinsker's inequality)**.** $\frac{1}{8}\left\|\rho_A - \sigma_A\right\|_1^2 \leq \mathfrak{h}\left(\rho_A,\ \sigma_A\right)^2 \leq \mathbf{D}\left(\rho_A \,||\, \sigma_A\right)$.

The proposition below says that mutual information does not change under local operations:

Proposition II.22. *If* $\sigma_{ABC} = U_B\ \rho_{ABC}\ U_B^\dagger$*, then* $\mathbf{I}_\sigma\left(A:BC\right) = \mathbf{I}_\rho\left(A:BC\right)$.

Furthermore, (2) combined with Pinsker's inequality, gives us

Proposition II.23. *Let* $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$*, then* $\mathfrak{h}\left(\rho_{AB},\ \rho_A \otimes \rho_B\right) \leq \sqrt{\mathbf{I}_\rho\left(A:B\right)}$.

Define $\mathbf{I}_\rho\left(A_S:B|S\right) := \mathbb{E}_S[\mathbf{I}_\rho\left(A_S:B\right)]$, then we have the following quantum version of Shearer's inequality from [ATYY17]:

Lemma II.24 (Quantum Shearer's Lemma [ATYY17])**.** *Let* $A = A_1, \ldots, A_n$ *and* B *be registers. Let* $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ *be a state such that* $\rho_A = \rho_{A_1} \otimes \rho_{A_2} \otimes \cdots \otimes \rho_{A_n}$*. Let* $S \subseteq [n]$ *be a random set independent of* ρ_{AB} *such that* $\mathbb{P}[i \in S] \leq \mu$ *for every* $i \in [n]$*. Then, we have*

$$\mathbf{I}_\rho\left(A_S:B|S\right) \leq \mu \cdot \mathbf{I}_\rho\left(A:B\right).$$

D. Quantum Communication Complexity

We consider quantum protocols where Alice and Bob are allowed to exchange qubits and they share some pure entangled state in the beginning, for instance a number of EPR-pairs that they are not charged for. Any lower bound in this model also translates to a lower bound in other models of quantum communication (Yao's model [Yao93] with qubit communication without prior entanglement or the Cleve-Buhrman model [CB97] with classical communication and prior entanglement).

The total state of a quantum protocol consists of: Alice and Bob's input registers X and Y, Alice's private register A, the communication channel C, and Bob's private register B. We assume that initially Alice and Bob share some pure entangled state $\psi_{A'B'}$ where A' and B' are part of Alice's and Bob's private registers A and B respectively, while the rest of the qubits in their private workspaces are initially zero ($|0\rangle$). The channel is also initially zero. Before the start of the protocol Alice and Bob copy their inputs from the input registers to their private workspaces. Let $|\psi\rangle_{AB}$ denote the state of registers A and B at the start. This includes the initial entangled state on A' and B', a number of zero-qubits and copy of their inputs x and y.

Given an input distribution $p(X,Y)$ on the inputs, the starting state of the protocol is then

$$\rho^{(0)}_{XYABC} = \sum_{xy} p(xy)|xy\rangle\langle xy|_{XY} \otimes |\psi\rangle\langle\psi|_{AB} \otimes |0\rangle\langle 0|_C.$$

Note that the marginal state $\rho^{(0)}_{XY}$ is a classical state, but not necessarily pure if X and Y are not independent. To make the above state a pure state, we will add purifying registers $\overline{X}$ and $\overline{Y}$ and consider the canonical purification of $\rho^{(0)}_{XY}$ which is the pure state

$$|\rho^{(0)}\rangle_{X\overline{X}Y\overline{Y}} = \sum_{xy} \sqrt{p(xy)}|xxyy\rangle_{X\overline{X}Y\overline{Y}}.$$

With the above purifying registers, the initial global state of the protocol is the pure state

$$|\rho^{(0)}\rangle_{X\overline{X}Y\overline{Y}ABC} = |\rho^{(0)}\rangle_{X\overline{X}Y\overline{Y}} \otimes |\psi\rangle_{AB} \otimes |0\rangle_C.$$

At each step of the protocol, either Alice or Bob applies a unitary to a subset of the registers. We will assume that they alternate: on odd rounds Alice acts and on even rounds

Bob acts. We will also assume that the channel C consists of one qubit. These assumptions can be made without loss of generality as they only affect the communication by a constant factor (see the remark at the end of Section IV, though).

In an odd round r, Alice applies a fixed unitary transformation $U^{(r)}_{XAC} = \sum_x |x\rangle\langle x|_X \otimes U^{(r),x}_{AC}$ to her private register and the channel. This corresponds to her private computation as well as to putting a one-qubit message on the channel. Note that the unitary uses the input register only as a control and does not change its contents. In an even round, Bob proceeds similarly. Hence the content of the input registers X and Y as well as the corresponding purifying registers $\overline{X}$ and $\overline{Y}$ remain unchanged throughout the protocol.

We assume that in the last round of the protocol Bob talks. The final state of an ℓ-round protocol (for even ℓ) on input distribution $p(X,Y)$ is the following pure state:

$$|\rho^{(\ell)}\rangle_{X\overline{X}Y\overline{Y}ABC} = U^{(\ell)}_{YBC} U^{(\ell-1)}_{XAC} \cdots U^{(1)}_{XAC} |\rho^{(0)}\rangle_{X\overline{X}Y\overline{Y}ABC}.$$

For technical reasons it will be convenient to assume that at the end of the protocol, the channel contains the answer. A measurement of the channel qubit in the computational basis then determines the output bit of the protocol. We say that the protocol computes $f(x,y)$ on a distribution $p(X,Y)$ if the probability of error on the input distribution $p(X,Y)$ is at most ϵ. Note that we may consider the run of the protocol on a fixed input x,y by taking the initial distribution $p(X,Y)$ such that $p(x,y) = 1$. We say that the protocol computes $f(x,y)$ with error ϵ if for every input x,y the probability of error is at most ϵ.

For notational convenience, throughout this work we will sometimes write $\rho^{(r)}$ instead of $\rho^{(r)}_{X\overline{X}Y\overline{Y}ABC}$ to denote the global state of the protocol on all the registers after round r. When referring to the marginal states, however, we will always write the corresponding registers.

Basic Properties of Quantum Protocols: In the following preliminary lemmas $\rho^{(r)}_{X\overline{X}Y\overline{Y}ABC}$ and $\sigma^{(r)}_{X\overline{X}Y\overline{Y}ABC}$ are the states of a quantum protocol after r rounds when it is run on input distributions $p(XY)$ and $q(XY)$ respectively. Moreover, ℓ will denote the last round of the protocol. The following proposition is easily seen to be true since the protocol applies the same sequence of unitaries on every input x,y:

Proposition II.25. *There are pure states $\{|\psi^{(r),xy}\rangle_{ABC}\}_{xy}$ such that*

$$|\rho^{(r)}\rangle_{X\overline{X}Y\overline{Y}ABC} = \sum_{xy} \sqrt{p(xy)} |xxyy\rangle_{X\overline{X}Y\overline{Y}} \otimes |\psi^{(r),xy}\rangle_{ABC}, \text{ and}$$
$$|\sigma^{(r)}\rangle_{X\overline{X}Y\overline{Y}ABC} = \sum_{xy} \sqrt{q(xy)} |xxyy\rangle_{X\overline{X}Y\overline{Y}} \otimes |\psi^{(r),xy}\rangle_{ABC}$$

Note that after the first round the states $|\psi^{(1),xy}\rangle_{ABC}$ only depend on x.

The above proposition implies that if $p(X,Y)$ is a product distribution on X and Y, and if in a round r, Bob applies a unitary $U^{(r)}_{YBC}$, then the marginal states satisfy

$$\rho^{(r)}_{X\overline{X}Y\overline{Y}BC} = U^{(r)}_{YBC} \rho^{(r-1)}_{X\overline{X}Y\overline{Y}BC} \left(U^{(r)}_{YBC}\right)^\dagger \text{ and}$$
$$\rho^{(r)}_{X\overline{X}Y\overline{Y}A} = \rho^{(r-1)}_{X\overline{X}Y\overline{Y}A},$$

and a similar statement also holds when Alice acts.

The following lemma follows easily from Proposition II.25:

Lemma II.26. $\frac{1}{2}\left\|\rho^{(r)}_C - \sigma^{(r)}_C\right\|_1 \leq \|p(XY) - q(XY)\|_{\mathsf{tv}}$.

Proof: Let $\delta = \|p(XY) - q(XY)\|_{\mathsf{tv}}$. Then using Proposition II.25, we can write

$$\frac{1}{2}\left\|\rho^{(r)}_C - \sigma^{(r)}_C\right\|_1 = \frac{1}{2}\left\|\sum_{xy}(p(xy) - q(xy))\psi^{(r),xy}_C\right\|_1 \leq \frac{1}{2}\sum_{xy}|p(xy) - q(xy)|\left\|\psi^{(r),xy}_C\right\|_1 \leq \delta,$$

where the second inequality is the triangle inequality and the last one follows from Proposition II.1 and the fact that $\{\psi^{(r),xy}_C\}_{xy}$ are density operators and have unit trace. ∎

Using Proposition II.14, the above also implies that if $p(XY)$ and $q(XY)$ are δ-close, then the output distributions of the protocol for both cases are δ-close.

Lemma II.27 (Errors and Trace Norm)**.** *Given a boolean function $f(x,y)$, let $p(X,Y)$ be a distribution supported on its 0-inputs and $q(X,Y)$ be a distribution supported on its 1-inputs. If an ℓ-round quantum protocol computes $f(x,y)$ with error δ, then $\frac{1}{2}\left\|\rho^{(\ell)}_C - \sigma^{(\ell)}_C\right\|_1 \geq 1 - 2\delta$.*

Proof: Recall that the last bit of the channel contains the answer and since the output of a protocol is given by a measurement of the channel qubit in the computational basis, the probabilities that the output is 0 under ρ_C and σ_C are respectively given by $\mathsf{Tr}(|0\rangle\langle 0|_C \rho_C) \geq 1 - \delta$ and $\mathsf{Tr}(|0\rangle\langle 0|_C \sigma_C) \leq \delta$. Using Proposition II.14, we have

$$\frac{1}{2}\left\|\rho^{(\ell)}_C - \sigma^{(\ell)}_C\right\|_1 \geq \mathsf{Tr}(|0\rangle\langle 0|_C(\rho_C - \sigma_C)) \geq (1-\delta) - \delta = 1 - 2\delta.$$

∎

Quantum protocols have no notion of a transcript, but the following lemma still gives a bound on how much information is revealed by a quantum protocol in terms of the communication.

Lemma II.28 (Information Cost)**.** *Let $p(XY)$ be a product input distribution on X and Y. Then, for any round r in the communication protocol, it holds that*

$$\mathbf{I}_{\rho^{(r)}}\left(X : Y\overline{Y}BC\right) \leq 2r \text{ and } \mathbf{I}_{\rho^{(r)}}\left(Y : X\overline{X}AC\right) \leq 2r.$$

Proof: The proof is by induction on the number of rounds. We will only prove the first inequality as the second one follows analogously. When $r = 0$, no messages have been exchanged and since $p(x, y) = p(x)p(y)$ for any x, y, it follows that the initial state is of the form $\rho^{(0)}_{X\overline{X}Y\overline{Y}ABC} = \rho^{(0)}_{X\overline{X}} \otimes \rho^{(0)}_{Y\overline{Y}} \otimes \rho^{(0)}_{ABC}$. So, using Proposition II.20, it follows that $\mathbf{I}_{\rho^{(0)}}\left(X : Y\overline{Y}BC\right) = 0$.

Now, let us assume that the statement holds for $r - 1$ rounds. When r is even, Bob applies a unitary $U^{(r)}_{YBC}$. Since $p(XY)$ is a product distribution on X and Y, Proposition II.25 implies that $\rho^{(r)}_{XY\overline{Y}BC} = U^{(r)}_{YBC}\rho^{(r-1)}_{XY\overline{Y}BC}\left(U^{(r)}_{YBC}\right)^{\dagger}$. Hence, using Proposition II.22, we have

$$\mathbf{I}_{\rho^{(r)}}\left(X : Y\overline{Y}BC\right) = \mathbf{I}_{\rho^{(r-1)}}\left(X : Y\overline{Y}BC\right) \leq 2(r-1),$$

where the inequality follows from the inductive hypothesis.

When r is odd, Alice applies a unitary $U^{(r)}_{XAC}$ with X as control. Using chain rule, we can write

$$\begin{aligned}
&\mathbf{I}_{\rho^{(r)}}\left(X : Y\overline{Y}BC\right)\\
&= \mathbf{I}_{\rho^{(r)}}\left(X : Y\overline{Y}B\right) + \mathbf{I}_{\rho^{(r)}}\left(X : C|Y\overline{Y}B\right)\\
&\leq \mathbf{I}_{\rho^{(r)}}\left(X : Y\overline{Y}B\right) + 2 = \mathbf{I}_{\rho^{(r-1)}}\left(X : Y\overline{Y}B\right) + 2\\
&\leq \mathbf{I}_{\rho^{(r-1)}}\left(X : Y\overline{Y}BC\right) + 2 \leq 2(r-1) + 2 = 2r,
\end{aligned}$$

where the first inequality follows from Proposition II.19 and the fact that $|C| = 1$, the second equality follows since $\rho^{(r)}_{XY\overline{Y}B} = \rho^{(r-1)}_{XY\overline{Y}B}$ as Alice applies a unitary U_{XAC} with X as a control register, and the second inequality follows from chain rule and non-negativity of conditional mutual information. ∎

III. Classical Communication Lower Bound

In this section, we present a new proof of the classical communication lower bound that we will later generalize to the quantum setting. We will prove that any randomized protocol for the sink function that errs with probability at most $1/3$ must communicate at least $\Omega(t)$ bits.

As is standard, to prove this we use a hard distribution $p(XY)$ on the inputs.

Hard Input Distribution $p(X, Y)$: Let $p_0(X, Y)$ and $p_1(X, Y)$ denote the uniform distribution on $\mathsf{sink}^{-1}(0)$ and $\mathsf{sink}^{-1}(1)$ respectively. In the input distribution $p(X, Y)$, the input is sampled from $p_0(X, Y)$ with probability $\frac{1}{2}$ and from $p_1(X, Y)$ with probability $\frac{1}{2}$.

Since we have a distribution on the inputs, we may assume without loss of generality that the randomized protocol is deterministic. We will prove a lower bound on the communication by showing that if the length of the messages of the protocol is at most $\frac{1}{2}\epsilon^3 t$, then the distribution of the messages looks almost the same under the distributions $p_0(X, Y)$ and $p_1(X, Y)$: denoting by $p_0(M)$ and $p_1(M)$ the induced distributions on the messages under $p_0(X, Y)$ and $p_1(X, Y)$, respectively, we will show that $p_0(M)$ and $p_1(M)$ are close in statistical distance. To show this, we use the *fooling distribution* method from [RS15]. We will give another distribution $u(X, Y)$ such that the induced distribution $u(M)$ will be close to each of $p_0(M)$ and $p_1(M)$. For the sink function, this **fooling distribution** $u(X, Y)$ is the uniform distribution on $\{0, 1\}^{n+n}$. More precisely, we prove:

Theorem III.1. *Let $\epsilon > 0$ be a constant and t be large enough. Then, for any deterministic protocol for the sink function with communication at most $\frac{1}{2}\epsilon^3 t$, we have that $p_0(M) \overset{o(1)}{\approx} u(M) \overset{4\epsilon}{\approx} p_1(M)$.*

Since the input distribution $p(X, Y)$ is balanced, using Lemma II.10, the distributions $p_0(M)$ and $p_1(M)$ must have statistical distance at least $1/3$ if the protocol has error $1/3$ on $p(X, Y)$. So, it must be that $4\epsilon + o(1) \geq 1/3$, and hence $\epsilon \geq 1/12 - o(1)$, and the $\Omega(t)$ lower bound on the communication (Theorem I.1) follows.

Next, we prove Theorem III.1. Before the proof, it will be helpful to keep in mind how the distributions $p_1(X, Y)$, $p_0(X, Y)$ and $u(X, Y)$ are related. Note that by definition, $p_0(X, Y) = u(X, Y|\mathsf{sink} = 0)$ and $p_1(X, Y) = u(X, Y|\mathsf{sink} = 1)$. Also, notice that the input distributions $p_0(X, Y)$ and $u(X, Y)$ are already very close in statistical distance:

Claim III.2. *$p_0(X, Y) \overset{\gamma}{\approx} u(X, Y)$ with $\gamma = t2^{-(t-1)} = o(1)$.*

Proof: Note that under the uniform distribution $u(XY)$, the probability that the function sink takes value 1 is exactly $t2^{-(t-1)}$, because for each vertex v, the event that v is the sink has probability exactly $2^{-(t-1)}$, and these events are disjoint for the t vertices. This means that

$$u(\mathsf{sink} = 0) = 1 - t2^{-(t-1)}.$$

Since $p_0(XY) = u(XY|\mathsf{sink} = 0)$, Lemma II.3 implies $\|p_0(XY) - u(XY)\|_{\mathsf{tv}} \leq t2^{-(t-1)} = \gamma$. ∎

Furthermore, recall that we can generate the distribution $p_1(X, Y)$ from $u(X, Y)$ by conditioning on a simple col-

lision event: for any vertex v, denoting by $N(v)$ the set of edges incident on v, the distribution $p_1(X,Y)$ can be generated from $u(X,Y)$ by first picking a uniformly random vertex $V \in [t]$ as the sink, and then conditioning on the event that $X_{N(V)} = Y_{N(V)} \oplus z_{N(V)}$, where $z_{N(v)}$ is the unique string that encodes the orientations of the edges in $N(v)$ when vertex v is the sink.

To complete the proof, we use the following lemma from [RS15], which bounds the effect of conditioning on a collision event (for completeness we include a proof in Appendix A).

Lemma III.3 (Lemma 4.3 in [RS15])**.** *Given a probability space q, if $A, B \in [r]$ are uniform and independent random variables, and $A - C - B$, then $q(C) \overset{\epsilon}{\approx} q(C|A=B)$, with $\epsilon = 2\sqrt[3]{\mathbf{I}_q(C:A)} + 2\sqrt[3]{\mathbf{I}_q(C:B)}$.*

Proof of Theorem III.1: Because we have that $\|p_0(XY) - u(XY)\|_{\mathsf{tv}} \leq t2^{-(t-1)} = o(1)$ from Claim III.2, this already implies that $\|p_0(M) - u(M)\|_{\mathsf{tv}} = o(1)$ since M is a function of X, Y. So, we focus on bounding $\|p_1(M) - u(M)\|_{\mathsf{tv}}$. For this, let $V \in [t]$ and let $u(V)$ be the uniform distribution on $[t]$. Recall that

$$p_1(XY) = \mathbb{E}_{u(v)}[u(XY|X_{N(v)} = Y_{N(v)} \oplus z_{N(v)})].$$

We will show that under the fooling distribution $u(XY)$, the messages of the protocol contain little information about $X_{N(V)}$ and $Y_{N(V)}$. Lemma III.3 and concavity will then complete the proof.

Note that for any fixed edge e, it holds that $u(e \in N(V)) = \frac{2}{t}$. Since under $u(XY)$, the binary random variables X_e (resp. Y_e) and $X_{e'}$ (resp. $Y_{e'}$) are mutually independent for any two edges e and e', applying Shearer's inequality (Lemma II.7), we get that

$$\mathbf{I}_u\left(X_{N(V)} : M|V\right) \leq \frac{2}{t} \cdot \mathbf{I}_u\left(X : M\right) \leq \frac{2}{t} \cdot |M| \leq \epsilon^3 \text{ and}$$
$$\mathbf{I}_u\left(Y_{N(V)} : M|V\right) \leq \frac{2}{t} \cdot \mathbf{I}_u\left(Y : M\right) \leq \frac{2}{t} \cdot |M| \leq \epsilon^3. \quad (3)$$

Note that for any v, shifting $Y_{N(v)}$ by a fixed string $z_{N(v)}$ does not change the mutual information $\mathbf{I}_u\left(Y_{N(v)} : M\right)$. Furthermore, since X and Y are independent $X - M - Y$ holds. Hence, using Proposition II.2 and Lemma III.3 (with $A = X_{N(v)}$, $B = Y_{N(v)} \oplus z_{N(v)}$, $C = M$), it holds that

$$\begin{aligned}
&\|p_1(M) - u(M)\|_{\mathsf{tv}} \\
&= \left\|\mathbb{E}_{u(v)}[u(M|X_{N(v)} = Y_{N(v)} \oplus z_{N(v)})] - u(M)\right\|_{\mathsf{tv}} \\
&= \mathbb{E}_{u(v)}\left[\left\| u(M|X_{N(v)} = Y_{N(v)} \oplus z_{N(v)}) - u(M) \right\|_{\mathsf{tv}}\right] \\
&\leq 2\mathbb{E}_{u(v)}\sqrt[3]{\mathbf{I}_u\left(X_{N(v)} : M\right)} + 2\mathbb{E}_{u(v)}\sqrt[3]{\mathbf{I}_u\left(Y_{N(v)} : M\right)}.
\end{aligned}$$

Further, using concavity of the cube root function over non-negative reals and (3), we get that

$$\begin{aligned}
&\|p_1(M) - u(M)\|_{\mathsf{tv}} \\
&\leq 2\sqrt[3]{\mathbb{E}_{u(v)}\mathbf{I}_u\left(X_{N(v)} : M\right)} + 2\sqrt[3]{\mathbb{E}_{u(v)}\mathbf{I}_u\left(Y_{N(v)} : M\right)} \\
&= 2\sqrt[3]{\mathbf{I}_u\left(X_{N(V)} : M|V\right)} + 2\sqrt[3]{\mathbf{I}_u\left(Y_{N(V)} : M|V\right)} \overset{(3)}{\leq} 4\epsilon.
\end{aligned}$$

Hence for t large enough, $\|p_0(M) - u(M)\|_{\mathsf{tv}} = o(1)$ and $\|p_1(M) - u(M)\|_{\mathsf{tv}} \leq 4\epsilon$, concluding the proof. ∎

IV. Quantum Communication Lower Bound

The proof for the quantum case proceeds similarly to the classical case with some minor but technical differences. Let $p_0(XY)$, and $u(XY)$ be as before: $p_0(XY)$ is uniform on $\mathsf{sink}^{-1}(0)$ and $u(XY)$ is the uniform distribution. Fix an ℓ-qubit protocol where per our convention ℓ is even as Bob sends the last message. Let $o^{(\ell)}$ and $\mu^{(\ell)}$ be the final pure states of the protocol on distributions $p_0(XY)$ and $u(XY)$, respectively. Let $V \in [t]$, let $u(V)$ denote the uniform distribution on $[t]$ and let $\iota^{v,(\ell)}$ denote the final pure state of the protocol when run on distribution $u(XY|X_{N(v)} = Y_{N(v)} \oplus z_{N(v)})$, that is, when vertex v is the sink. Note that the distribution $u(XY|X_{N(v)} = Y_{N(v)} \oplus z_{N(v)})$ is supported on only the 1-inputs to the sink function. If the protocol computes the sink function with error at most $1/3$ on every input, then Lemma II.27 implies

$$\mathbb{E}_{u(v)}\left[\left\|\iota_C^{v,(\ell)} - o_C^{(\ell)}\right\|_1\right] \geq 2/3. \quad (4)$$

We are going to argue that if $\ell \ll t^{1/3}$, then the distribution $u(XY)$ is also a fooling distribution for quantum protocols. That is, it must be the case that both $o_C^{(\ell)} \approx \mu_C^{(\ell)}$ and, for a typical vertex v, $\iota_C^{v,(\ell)} \approx \mu_C^{(\ell)}$ (and hence $\iota_C^{v,(\ell)} \approx o_C^{(\ell)}$ for a typical v).

Theorem IV.1. *Let $\epsilon > 0$ be a constant and t be large enough. Then, for any quantum protocol for the sink function with communication complexity at most $\ell = \frac{1}{8}\epsilon^{2/3}t^{1/3}$, we have that*

$$\mathbb{E}_{u(v)}\left[\left\|\iota_C^{v,(\ell)} - o_C^{(\ell)}\right\|_1\right] \leq \epsilon.$$

Combining this theorem with (4) immediately implies the quantum communication lower bound of $\Omega(t^{1/3})$ promised by Theorem I.2.

First, $o_C^{(\ell)} \approx \mu_C^{(\ell)}$ is clear because $p_0(X,Y) \approx u(X,Y)$ (see Lemma II.26). To prove that $\iota_C^{v,(\ell)} \approx \mu_C^{(\ell)}$ for a typical v, we will use Lemma 3.6 from [ATYY17] (we state it a bit differently here to make it easier for our application). This allows us to relate the fooling distribution with the input distribution similar to the role of Lemma III.3 in the proof for the classical case. The proof of this lemma is an

involved round-by-round induction; we include a proof in Appendix A for completeness.

Lemma IV.2 (Lemma 3.6 in [ATYY17])**.** *Let $X = X_1X_2$ and $Y = Y_1Y_2$ be random variables where $X, Y \in \{0,1\}^n$. Let $u'(XY)$ be the uniform distribution on XY and let $q(XY) = u'(XY|X_1 = Y_1)$ be another distribution. For every $s \le r$, let $\rho^{(s)}$ and $\sigma^{(s)}$ denote the state of a quantum protocol after s rounds on distributions $u'(XY)$ and $q(XY)$ respectively. If for every $s \le r$, we have*

$$\mathbf{I}_{\rho^{(s)}}\left(X_1 : Y\overline{Y}BC\right) \le \epsilon_s \text{ for odd } s, \text{ and}$$
$$\mathbf{I}_{\rho^{(s)}}\left(Y_1 : X\overline{X}AC\right) \le \epsilon_s \text{ for even } s,$$

then it holds that

$$\left\|\sigma^{(r)}_{X_1Y_1C} - \sigma^{(r)}_{X_1Y_1} \otimes \rho^{(r)}_C\right\|_1 \le 4\sqrt{2}\sum_{s=1}^{r}\sqrt{\epsilon_s}.$$

Let us fix a vertex $v \in [t]$ and for notational convenience, define $\epsilon_{v,s} = \mathbf{I}_{\mu^{(s)}}\left(X_{N(v)} : Y\overline{Y}BC\right)$ for odd rounds s, and $\epsilon_{v,s} = \mathbf{I}_{\mu^{(s)}}\left(Y_{N(v)} : X\overline{X}AC\right)$ for even rounds s. If these $\epsilon_{v,s}$'s are mostly small, then $\iota_C^{v,(\ell)} \approx \mu_C^{(\ell)}$ as the following lemma shows.

Lemma IV.3. $\left\|\iota_C^{v,(\ell)} - \mu_C^{(\ell)}\right\|_1 \le 4\sqrt{2}\sum_{s=1}^{\ell}\sqrt{\epsilon_{v,s}}$.

Proof: To apply Lemma IV.2, we will choose $X_1 = X_{N(v)}$, $X_2 = X_{N(v)^c}$ and $Y_1 = Y_{N(v)} \oplus z_{N(v)}, Y_2 = Y_{N(v)^c}$ and $u'(XY) = u(XY)$. Note that $u'(XY)$ is still the uniform distribution. Furthermore, using Proposition II.25, for every s, the state $\rho^{(s)}$ in Lemma IV.2 is the same as $\mu^{(s)}$ after a suitable relabeling. Hence, it follows that $\mathbf{I}_{\rho^{(s)}}\left(X_1 : Y\overline{Y}BC\right) = \mathbf{I}_{\mu^{(s)}}\left(X_{N(v)} : Y\overline{Y}BC\right) = \epsilon_{v,s}$ for odd s, and $\mathbf{I}_{\rho^{(s)}}\left(Y_1 : X\overline{X}AC\right) = \mathbf{I}_{\mu^{(s)}}\left(Y_{N(v)} : X\overline{X}AC\right) = \epsilon_{v,s}$ for even s.

Now, we apply Lemma IV.2. Since $\mathsf{Tr}_{X_1Y_1}\left(\iota^{v,(\ell)}_{X_1Y_1C}\right) = \iota_C^{v,(\ell)}$ and $\mathsf{Tr}_{X_1Y_1}\left(\iota^{v}_{X_1Y_1} \otimes \mu_C^{(\ell)}\right) = \mu_C^{(\ell)}$, we get that $\left\|\iota_C^{v,(\ell)} - \mu_C^{(\ell)}\right\|_1 \le \left\|\iota^{v,(\ell)}_{X_1Y_1C} - \iota^{v}_{X_1Y_1} \otimes \mu_C^{(\ell)}\right\|_1 \le 4\sqrt{2}\sum_{s=1}^{\ell}\sqrt{\epsilon_{v,s}}$. ∎

We move on to the proof of the theorem now.

Proof of Theorem IV.1: Recall that $\|p_0(XY) - u(XY)\|_{\mathsf{tv}} \le t2^{-(t-1)} = o(1)$ from Claim III.2, and using Lemma II.26, this already implies that $\left\|o_C^{(\ell)} - \mu_C^{(\ell)}\right\|_1 = o(1)$.

Let us turn to bounding $\mathbb{E}_{u(v)}\left[\left\|\iota_C^{v,(\ell)} - \mu_C^{(\ell)}\right\|_1\right]$. We first show that under the fooling distribution $u(XY)$, the states of the quantum protocol contain little information about $X_{N(V)}$ and $Y_{N(V)}$. Then, applying Lemma IV.3 and appealing to concavity will complete the proof similar to the classical case.

Note that for any fixed edge e, it holds that $u(e \in N(V)) = \frac{2}{t}$, and also recall that under $u(XY)$, the random variables X_e (resp. Y_e) and $X_{e'}$ (resp. $Y_{e'}$) are mutually independent for any two edges e and e'. Therefore, using Proposition II.25 the state $\mu_X^{(\ell)} = \underset{e}{\otimes}\mu^{(\ell)}_{X_e}$ (and similarly $\mu_Y^{(\ell)} = \underset{e}{\otimes}\mu^{(\ell)}_{Y_e}$). Hence, applying the quantum version of Shearer's inequality (Lemma II.24) and using Lemma II.28, for every round $s \le \ell$ we get that

$$\begin{aligned}
&\mathbf{I}_{\mu^{(s)}}\left(X_{N(V)} : Y\overline{Y}BC|V\right) \\
&\qquad \le \frac{2}{t}\cdot\mathbf{I}_{\mu^{(s)}}\left(X : Y\overline{Y}BC\right) \le \frac{4\ell}{t} \text{ and} \\
&\mathbf{I}_{\mu^{(s)}}\left(Y_{N(V)} : X\overline{X}AC|V\right) \\
&\qquad \le \frac{2}{t}\cdot\mathbf{I}_{\mu^{(s)}}\left(Y : X\overline{X}AC\right) \le \frac{4\ell}{t}. \qquad (5)
\end{aligned}$$

Further using Lemma IV.3, concavity, and (5) we get

$$\begin{aligned}
&\mathbb{E}_{u(v)}\left[\left\|\iota_C^{v,(\ell)} - \mu_C^{(\ell)}\right\|_1\right] \\
&\le 4\sqrt{2}\ \mathbb{E}_{u(v)}\left[\sum_{s=1}^{\ell}\sqrt{\epsilon_{v,s}}\right] \\
&\le 4\sqrt{2}\ \sum_{s=1}^{\ell}\sqrt{\mathbb{E}_{u(v)}[\epsilon_{v,s}]} \\
&= 4\sqrt{2}\sum_{\substack{s=1\\ s \text{ odd}}}^{\ell}\sqrt{\mathbf{I}_{\mu^{(s)}}\left(X_{N(V)} : Y\overline{Y}BC|V\right)} \\
&\qquad + 4\sqrt{2}\sum_{\substack{s=1\\ s \text{ even}}}^{\ell}\sqrt{\mathbf{I}_{\mu^{(s)}}\left(Y_{N(V)} : X\overline{X}AC|V\right)} \\
&\le 4\sqrt{2}\cdot\frac{\ell}{2}\sqrt{\frac{4\ell}{t}} + 4\sqrt{2}\cdot\frac{\ell}{2}\sqrt{\frac{4\ell}{t}} = \sqrt{\frac{128\ell^3}{t}} \le \frac{\epsilon}{2}.
\end{aligned}$$

Using the triangle inequality, we get that for large enough t, the following holds

$$\begin{aligned}
&\mathbb{E}_{u(v)}\left[\left\|\iota_C^{v,(\ell)} - o_C^{(\ell)}\right\|_1\right] \\
&\le \mathbb{E}_{u(v)}\left[\left\|\iota_C^{v,(\ell)} - \mu_C^{(\ell)}\right\|_1\right] + \left\|o_C^{(\ell)} - \mu_C^{(\ell)}\right\|_1 \\
&\le \frac{\epsilon}{2} + o(1) \le \epsilon.
\end{aligned}$$

∎

We remark that the proof given above can be modified to show the stronger result that the r-round quantum communication complexity of the sink function is $\Omega(\max\{t/r^2, r\})$. We briefly sketch the modifications necessary, but do not attempt a formal proof here. First note that for ease of presentation, so far we used the communication model where the size of the channel remains fixed; one can do this without loss of generality if one does not care about rounds, but to properly define bounded-round quantum protocols one has to consider a different communication model where the size of the channel register can be different in different rounds. In this model, one can still show that the corresponding information quantities in Lemma II.28 remain upper bounded by

twice the total communication until that point. Moreover, Lemma IV.2 also remains valid in this model (this is the model considered in [ATYY17]). Then, proceeding as in the proof of Theorem IV.1, one upper bounds the trace distance between the corresponding final states by $O(r\sqrt{\ell/t})$, where ℓ is the total communication of the protocol. Because that distance is $\Omega(1)$ for a good protocol, we obtain $\ell = \Omega(t/r^2)$. Since at least one qubit must be communicated per round, we also have $\ell = \Omega(r)$.

V. Future Work

One obvious remaining open problem is to close the gap between the current lower bound of $\Omega(t^{1/3})$ on the quantum communication complexity of the sink function, and the best known upper bound of $\widetilde{O}(\sqrt{t})$. We conjecture the upper bound is essentially tight. One way to improve our lower bound would be to improve Lemma IV.3, maybe with a different distance measure.

The main question left open by this work, as well as by [CMS19], [ABT19], is of course the status of the (non-approximate) log-rank conjecture itself. The proof that the sink function has low approximate rank crucially uses the fact that the identity matrix has low approximate rank (which follows from the fact that the equality function has low randomized communication complexity). In contrast, the actual (non-approximate) rank of the identity matrix is as large as its dimension. Accordingly, it is not so clear what examples like the sink function suggest for the status of the log-rank conjecture itself. We are not sure what to conjecture about that conjecture.

Acknowledgments.

Thanks to Sander Gribling for useful discussions and a heartfelt gratitude to him for careful proofreading of our first draft. Thanks to Siva Ramamoorthy for helpful comments. We also thank Arkadev Chattopadhyay, Nikhil Mande and Suhail Sherif for answering questions about their paper [CMS19], as well as Anurag Anshu, Naresh Goud Boddu and Dave Touchette for helpful correspondence about their independent work [ABT19] and helpful comments on our draft. RdW thanks Srinivasan Arunachalam for pointing him to [CMS19] when it had just appeared on ECCC, and for helpful comments on the draft. We thank Peter van Emde Boas for pointing us to the unpublished report [BEL74].

References

[AA05] Scott Aaronson and Andris Ambainis. Quantum search of spatial regions. *Theory of Computing*, 1(1):47–79, 2005. Earlier version in FOCS'03. quant-ph/0303041.

[ABBD+16] Anurag Anshu, Aleksandrs Belovs, Shalev Ben-David, Mika Göös, Rahul Jain, Robin Kothari, Troy Lee, and Miklos Santha. Separations in communication complexity using cheat sheets and information complexity. In *Proceedings of 57th IEEE FOCS*, pages 555–564, 2016. arXiv:1605.01142.

[ABDG+17] Anurag Anshu, Shalev Ben-David, Ankit Garg, Rahul Jain, Robin Kothari, and Troy Lee. Separating quantum communication and approximate rank. In *Proceedings of Computational Complexity Conference (CCC'17)*, pages 24:1–24:33, 2017. arXiv:1611.05754.

[ABT19] Aurag Anshu, Naresh Goud Boddu, and Dave Touchette. Quantum Log-Approximate-Rank conjecture is also false. In *Proceedings of 60th FOCS*, 2019. arXiv:1811.10525.

[ATYY17] Anurag Anshu, Dave Touchette, Penghui Yao, and Nengkun Yu. Exponential separation of quantum communication and classical information. In *Proceedings of 49th ACM STOC*, pages 277–288, 2017.

[BCW98] Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *Proceedings of 30th ACM STOC*, pages 63–68, 1998. quant-ph/9802040.

[BEL74] Marc R. Best, Peter van Emde Boas, and Hendrik W. Lenstra, Jr. A sharpened version of the Aanderaa-Rosenberg conjecture. Technical Report Mathematisch Centrum ZW 30/74, January 1974. https://ir.cwi.nl/pub/6921.

[BW01] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of 16th IEEE Conference on Computational Complexity*, pages 120–130, 2001. cs.CC/9910010.

[CB97] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997. quant-ph/9704026.

[CMS19] Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. In *Proceedings of 51th ACM STOC*, pages 42–53, 2019. ECCC 25:176 (2018).

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[GJPW17] Mika Göös, T. S. Jayram, Toniann Pitassi, and Thomas Watson. Randomized communication vs. partition number. In *Proceedings of 44th ICALP*, pages 52:1–52:15, 2017.

[GKR16] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *Journal of the ACM*, 63(5):46:1–46:31, 2016.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *Proceedings of 56th FOCS*, pages 1077–1088, 2015.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996. quant-ph/9605043.

[HHL16] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In *Proceedings of 57th IEEE FOCS*, pages 282–288, 2016.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, New York, NY, USA, 1997.

[KR11] Boaz Klartag and Oded Regev. Quantum one-way communication is exponentially stronger than classical communication. In *Proceedings of 43rd ACM STOC*, 2011. arXiv:1009.3640.

[Lov14] Shachar Lovett. Recent advances on the log-rank conjecture in communication complexity. *Bulletin of the EATCS*, 112, 2014.

[Lov16] Shachar Lovett. Communication is bounded by root of rank. *Journal of the ACM*, 63(1):1:1–1:9, 2016. Earlier version in STOC'14.

[LS93] Lászlo Lovász and Michael Saks. Communication complexity and combinatorial lattice theory. *Journal of Computer and System Sciences*, 47:322–349, 1993. Earlier version in FOCS'88.

[LS09] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Foundations and Trends in Theoretical Computer Science*, 3(4):263–398, 2009.

[MS82] Kurt Mehlhorn and Erik Schmidt. Las Vegas is better than determinism in VLSI and distributed computing. In *Proceedings of 14th ACM STOC*, pages 330–337, 1982.

[Raz99] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of 31st ACM STOC*, pages 358–367, 1999.

[Rot14] Thomas Rothvoß. A direct proof for Lovett's bound on the communication complexity of low rank matrices. *CoRR*, abs/1409.6366, 2014.

[RS15] Anup Rao and Makrand Sinha. Simplified Separation of Information and Communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:57, 2015.

[RY18] Anup Rao and Amir Yehudayoff. *Communication Complexity*. Textbook (Draft), 2018.

[TWXZ13] Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang:. Fourier sparsity, spectral norm, and the log-rank conjecture. In *Proceedings of 54th IEEE FOCS*, pages 658–667, 2013. arXiv:1304.1245.

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[Wil13] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.

[Yao93] Andrew C-C. Yao. Quantum circuit complexity. In *Proceedings of 34th IEEE FOCS*, pages 352–360, 1993.

[Zha14] Shengyu Zhang. Efficient quantum protocols for XOR functions. In *Proceedings of 25th SODA*, pages 1878–1885, 2014. arXiv:1307.6738.

APPENDIX

Lemma III.3 (Lemma 4.3 in [RS15])**.** *Given a probability space q, if $A, B \in [r]$ are uniform and independent random variables, and $A - C - B$, then $q(C) \overset{\epsilon}{\approx} q(C|A=B)$, with $\epsilon = 2\sqrt[3]{\mathbf{I}_q(C:A)} + 2\sqrt[3]{\mathbf{I}_q(C:B)}$.*

Proof: We assume $\mathbf{I}_q(C:A), \mathbf{I}_q(C:B) \leq 1$, since otherwise the lemma is trivially true. For brevity, set

$$\alpha^3 = \mathbf{I}_q(C:A) = \mathop{\mathbb{E}}_{q(c)}\left[\mathbf{D}\left(q(A|c) \,||\, q(A)\right)\right] \text{ and}$$
$$\beta^3 = \mathbf{I}_q(C:B) = \mathop{\mathbb{E}}_{q(c)}\left[\mathbf{D}\left(q(B|c) \,||\, q(B)\right)\right].$$

Call c *bad* if $\mathbf{D}\left(q(A|c) \,||\, q(A)\right) \geq \alpha^2$ or $\mathbf{D}\left(q(B|c) \,||\, q(B)\right) \geq \beta^2$, and *good* otherwise. By Markov's inequality, the probability that C is bad is at most $\alpha + \beta$. To prove Lemma III.3, we need the following claim proved in [GKR16]. For completeness, we include the short proof after finishing the proof of Lemma III.3.

Claim A.1. *Given independent random variables $A^*, B^* \in [r]$ in a probability space q, if A^* is γ_1-close to uniform, and B^* is γ_2-close to uniform, then $q(A^* = B^*) \geq \dfrac{1-\gamma_1-\gamma_2}{r}$.*

When c is good, Pinsker's inequality (Proposition II.6) implies that conditioned on c, A is α-close to uniform and B is β-close to uniform. Then, since $A-C-B$, using Claim A.1 (with $A^* = A$ and $B^* = B$ in the probability space q conditioned on c) implies that $q(A = B|c) \geq \frac{1-\alpha-\beta}{r}$. Since $q(A=B) = \frac{1}{r}$, we have that for a good c,

$$q(c|A=B) = \frac{q(c)\cdot q(A=B|c)}{q(A=B)} \geq (1-\alpha-\beta)\cdot q(c). \tag{6}$$

For any event $\mathcal{Q}$, (6) implies that

$$\begin{aligned}
&q(C \in \mathcal{Q}) - q(C \in \mathcal{Q}|A=B)\\
&\leq \sum_{c\in\mathcal{Q}, c \text{ bad}} q(c) + \sum_{c\in\mathcal{Q}, c \text{ good}} \left(q(c) - q(c|A=B)\right)\\
&\leq q(C \text{ is bad}) + \sum_c q(c)(\alpha+\beta)\\
&\leq \alpha+\beta+\sum_c q(c)(\alpha+\beta) \leq 2\alpha+2\beta,
\end{aligned}$$

and since $\|q(C) - q(C|A=B)\|_{\mathsf{tv}} = \max_{\mathcal{Q}}(q(C\in\mathcal{Q}) - q(C\in\mathcal{Q}|A=B))$ we get the required upper bound on statistical distance. ∎

Proof of Claim A.1: For each i, let $q(A^* = i) = \frac{1}{r} + \alpha_i$ and $q(B^* = i) = \frac{1}{r} + \beta_i$. Then, $\sum_i \alpha_i = \sum_i \beta_i = 0$, and $\alpha_i, \beta_i \geq -\frac{1}{r}$. Using these facts,

$$\begin{aligned} q(A^* = B^*) &= \sum_i \left(\frac{1}{r} + \alpha_i\right)\left(\frac{1}{r} + \beta_i\right) \\ &= \frac{1}{r} + \frac{\sum_i \alpha_i}{r} + \frac{\sum_i \beta_i}{r} + \sum_i \alpha_i\beta_i \\ &= \frac{1}{r} + \sum_i \alpha_i\beta_i. \end{aligned}$$

To lower bound the above, we will only consider the negative terms in the summation:

$$\begin{aligned} q(A^* = B^*) &\geq \frac{1}{r} + \sum_{i:\alpha_i>0,\beta_i<0} \alpha_i\beta_i + \sum_{i:\alpha_i<0,\beta_i>0} \alpha_i\beta_i \\ &\geq \frac{1}{r} - \frac{1}{r}\sum_{i:\alpha_i>0} \alpha_i - \frac{1}{r}\sum_{i:\beta_i>0} \beta_i. \end{aligned}$$

From Proposition II.1, it follows that $\sum_{i:\alpha_i>0} \alpha_i$ is the statistical distance γ_1 between A^* and the uniform distribution on $[r]$ and likewise for B^*. So we get

$$q(A^* = B^*) \geq \frac{1 - \gamma_1 - \gamma_2}{r}.$$

■

Lemma IV.2 (Lemma 3.6 in [ATYY17])**.** *Let $X = X_1X_2$ and $Y = Y_1Y_2$ be random variables where $X, Y \in \{0,1\}^n$. Let $u'(XY)$ be the uniform distribution on XY and let $q(XY) = u'(XY|X_1 = Y_1)$ be another distribution. For every $s \leq r$, let $\rho^{(s)}$ and $\sigma^{(s)}$ denote the state of a quantum protocol after s rounds on distributions $u'(XY)$ and $q(XY)$ respectively. If for every $s \leq r$, we have*

$$\begin{aligned} &\mathbf{I}_{\rho^{(s)}}\left(X_1 : Y\overline{Y}BC\right) \leq \epsilon_s \text{ for odd } s, \text{ and} \\ &\mathbf{I}_{\rho^{(s)}}\left(Y_1 : X\overline{X}AC\right) \leq \epsilon_s \text{ for even } s, \end{aligned}$$

then it holds that

$$\left\|\sigma^{(r)}_{X_1Y_1C} - \sigma^{(r)}_{X_1Y_1} \otimes \rho^{(r)}_C\right\|_1 \leq 4\sqrt{2}\sum_{s=1}^{r}\sqrt{\epsilon_s}.$$

To simplify the notation in the proof, define $R = X_2\overline{X_2}Y_2\overline{Y_2}AB$, and $X_1' = X_1\overline{X_1}$, $Y_1' = Y_1\overline{Y_1}$, $X_2' = X_2\overline{X_2}$ and $Y_2' = Y_2\overline{Y_2}$. Furthermore, we will use boldface letters to denote different classical registers with the same dimensions, for example $\mathbf{X_1'} = \mathbf{X_1}\overline{\mathbf{X_1}}$ will denote an independent register of the same dimension as X_1'. One should think of the boldface registers as a relabeling of the original registers but they will be needed since we will consider states like $|\sigma^{(r)}\rangle_{X_1'Y_1'} \otimes |\rho^{(r)}\rangle_{\mathbf{X_1'Y_1'}RC}$.

Also, note that if we have two unitaries U_{XA} and V_{XB} that both have a classical register X as control, then U_{XA} and V_{XB} commute (recall our convention that we omit to write tensor product with the identity operator on the remaining spaces).

Proof of Lemma IV.2:

We will bound

$$\begin{aligned} &\left\|\sigma^{(r)}_{X_1Y_1C} - \sigma^{(r)}_{X_1Y_1} \otimes \rho^{(r)}_C\right\|_1 \\ &\leq 2\sqrt{2}\,\mathfrak{h}\left(\sigma^{(r)}_{X_1Y_1C},\ \sigma^{(r)}_{X_1Y_1} \otimes \rho^{(r)}_C\right) \qquad (7) \\ &= 2\sqrt{2}\,\min_{\widetilde{U}} \mathfrak{h}\big(\widetilde{U}|\sigma^{(r)}\rangle_{X_1'Y_1'RC} \otimes |\rho\rangle_{\mathbf{X_1'Y_1'}}, \\ &\qquad\qquad |\sigma^{(r)}\rangle_{X_1'Y_1'} \otimes |\rho^{(r)}\rangle_{\mathbf{X_1'Y_1'}RC}\big) \qquad (8) \end{aligned}$$

where the inequality used Proposition II.13, and the equality follows from Uhlmann's theorem (Proposition II.15) with $\widetilde{U}$ ranging over all unitaries acting on $\overline{X}_1\overline{Y}_1\mathbf{X_1'Y_1'}R$. Notice that apart from $\overline{X}_1\overline{Y}_1R$, we also need the boldface registers to make the state $\sigma^{(r)}_{X_1Y_1} \otimes \rho^{(r)}_C$ a pure state. Also, note that $|\rho^{(r)}\rangle_{X_1'Y_1'}$ and $|\sigma^{(r)}\rangle_{X_1'Y_1'}$ remain the same throughout all rounds, so we will drop the superscript r for these states.

To upper bound the right-hand side in (7), we will exhibit a unitary $\widetilde{U}$ so that the Hellinger distance is small. Let us first note that since $u'(X_1) = q(X_1)$, using Proposition II.16, there exists a unitary $W_{X_1Y_1'}$ with X_1 as a control register such that $W_{X_1Y_1'}|\rho\rangle_{X_1'}|\rho\rangle_{Y_1'} = |\sigma\rangle_{X_1'Y_1'}$. Similarly, since $u'(Y_1) = q(Y_1)$ there exists a similar unitary $W_{X_1'Y_1}$ with Y_1 as a control (we will use the same letter to denote them since the subscripts will make the registers clear).

We first claim that

Claim A.2. *There exist unitaries $V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}$ for odd s, and $V^{(s)}_{\overline{Y}_1\mathbf{Y_1'}Y_2'B}$ for even s with $V^{(0)}_{\overline{Y}_1\mathbf{Y_1'}Y_2'B} = I_{\overline{Y}_1\mathbf{Y_1'}Y_2'B}$, such that*

$$\begin{aligned} &\mathfrak{h}\big(V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}W_{X_1Y_1'}|\rho^{(s)}\rangle_{X_1'\mathbf{Y_1'}RC} \otimes |\rho\rangle_{\mathbf{X_1'}Y_1'}, \\ &\quad |\rho^{(s)}\rangle_{\mathbf{X_1'Y_1'}RC} \otimes |\sigma\rangle_{X_1'Y_1'}\big) \leq \sqrt{\epsilon_s} \text{ for odd } s, \quad \text{and} \\ &\mathfrak{h}\big(V^{(s)}_{\overline{Y}_1\mathbf{Y_1'}Y_2'B}W_{X_1'Y_1}|\rho^{(s)}\rangle_{\mathbf{X_1'}Y_1'RC} \otimes |\rho\rangle_{X_1'\mathbf{Y_1'}}, \\ &\quad |\rho^{(s)}\rangle_{\mathbf{X_1'Y_1'}RC} \otimes |\sigma\rangle_{X_1'Y_1'}\big) \leq \sqrt{\epsilon_s} \text{ for even } s. \end{aligned}$$

Below we will drop the registers when we are writing states over all the registers $X_1'Y_1'RC\mathbf{X_1'Y_1'}$. We will also drop the registers from the unitaries $V^{(s)}$ since their indices (whether odd or even) will describe the corresponding registers they act on, unless we need to emphasize it.

Let us define

$$\begin{aligned} &|\theta^{(s)}\rangle = V^{(s)}V^{(s-1)}|\sigma^{(s)}\rangle_{X_1'Y_1'RC} \otimes |\rho\rangle_{\mathbf{X_1'Y_1'}} \text{ and} \\ &|\lambda^{(s)}\rangle = |\rho^{(s)}\rangle_{\mathbf{X_1'Y_1'}RC} \otimes |\sigma\rangle_{X_1'Y_1'}. \end{aligned} \qquad (9)$$

Then, we will prove by induction that for every round s, the following holds:

Claim A.3. $\mathfrak{h}\left(|\theta^{(s)}\rangle,\ |\lambda^{(s)}\rangle\right) \leq \delta_s$ *where* $\delta_s = \sqrt{\epsilon_s} + \sqrt{\epsilon_{s-1}} + 2\sum_{i=1}^{s-2}\sqrt{\epsilon_i}$.

For $s = r$, Claim A.3 implies that the unitary $V^{(r)}V^{(r-1)}$ is a particular unitary acting on $\overline{X}_1\overline{Y}_1\mathbf{X_1'Y_1'}R$ for which the

right-hand side in (7) is small, so taking $\widetilde{U}$ to be $V^{(r)}V^{(r-1)}$ in (7),

$$\left\|\sigma^{(r)}_{X_1Y_1C} - \sigma^{(r)}_{X_1Y_1}\otimes\rho^{(r)}_C\right\|_1 \leq 2\sqrt{2}\left(\sqrt{\epsilon_r}+\sqrt{\epsilon_{r-1}}+2\sum_{s=1}^{r-2}\sqrt{\epsilon_s}\right)\leq 4\sqrt{2}\left(\sum_{s=1}^{r}\sqrt{\epsilon_s}\right).$$

This completes the proof of Lemma IV.2 assuming the claims. ∎

We next prove Claims A.2 and A.3 in order.

Proof of Claim A.2: We will only prove the first inequality as the second one is analogous. From the assumption that $\mathbf{I}_{\rho^{(s)}}\left(X_1 : Y\overline{Y}BC\right)\leq\epsilon_s$ it also follows that $\mathbf{I}_{\rho^{(s)}}\left(X_1 : \mathbf{Y_1'}Y_2'BC\right)\leq\epsilon_s$ since we are just relabeling the Y_1' registers to $\mathbf{Y_1'}$ (recall $Y_1'=Y_1\overline{Y}_1$). Using Proposition II.23,

$$\mathfrak{h}\left(\rho^{(s)}_{X_1\mathbf{Y_1'}Y_2'BC},\ \rho^{(s)}_{\mathbf{Y_1'}Y_2'BC}\otimes\rho_{X_1}\right)\leq\sqrt{\mathbf{I}_{\rho^{(s)}}\left(X_1:\mathbf{Y_1'}Y_2'BC\right)}\leq\sqrt{\epsilon_s}.$$

Recalling that $R=X_2\overline{X_2}Y_2\overline{Y_2}AB$ and using Uhlmann's Theorem (Proposition II.15), there exists a unitary $V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}$ such that

$$\begin{aligned}\sqrt{\epsilon_s}&\geq\mathfrak{h}\big(V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}|\rho^{(s)}\rangle_{X_1'\mathbf{Y_1'}RC}\otimes|\rho\rangle_{\mathbf{X_1'}},\\&\qquad|\rho^{(s)}\rangle_{\mathbf{X_1'}\mathbf{Y_1'}RC}\otimes|\rho\rangle_{X_1'}\big)\\&=\mathfrak{h}\big(V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}|\rho^{(s)}\rangle_{X_1'\mathbf{Y_1'}RC}\otimes|\rho\rangle_{\mathbf{X_1'}}\otimes|\rho\rangle_{Y_1'},\\&\qquad|\rho^{(s)}\rangle_{\mathbf{X_1'}\mathbf{Y_1'}RC}\otimes|\rho\rangle_{X_1'}\otimes|\rho\rangle_{Y_1'}\big)\\&=\mathfrak{h}\big(V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}W_{X_1Y_1'}|\rho^{(s)}\rangle_{X_1'\mathbf{Y_1'}RC}\otimes|\rho\rangle_{\mathbf{X_1'}Y_1'},\\&\qquad|\rho^{(s)}\rangle_{\mathbf{X_1'}\mathbf{Y_1'}RC}\otimes|\sigma\rangle_{X_1'Y_1'}\big),\end{aligned}$$

where in the last equality we multiplied both states by the unitary $W_{X_1Y_1'}$ and used that $W_{X_1Y_1'}|\rho\rangle_{X_1'}\otimes|\rho\rangle_{Y_1'}=|\sigma\rangle_{X_1'Y_1'}$ as well as the fact that $W_{X_1Y_1'}$ and $V^{(s)}_{\overline{X}_1\mathbf{X_1'}X_2'A}$ commute (disjoint registers). ∎

Proof of Claim A.3: Base case $s=1$: Recall that $V^{(0)}$ is the identity. Let $W_{X_1Y_1'}$ be the unitary that satisfies $W_{X_1Y_1'}|\rho\rangle_{X_1'Y_1'}=|\sigma\rangle_{X_1'Y_1'}$ as before. Then, since $u'(X)=q(X)$ and $q(Y_1Y_2)=q(Y_1)q(Y_2)$ and $q(Y_2)=u'(Y_2)$, it follows from Proposition II.25 that $W_{X_1Y_1'}|\rho^{(1)}\rangle_{X_1'Y_1'RC}=|\sigma^{(1)}\rangle_{X_1'Y_1'RC}$ (recall that $R=X_2\overline{X_2}Y_2\overline{Y_2}AB$). Using this and the fact that $|\rho^{(1)}\rangle_{X_1'\mathbf{Y_1'}RC}=|\rho^{(1)}\rangle_{X_1'RC}\otimes|\rho\rangle_{\mathbf{Y_1'}}$ and $|\rho\rangle_{\mathbf{X_1'}Y_1'}=|\rho\rangle_{\mathbf{X_1'}}\otimes|\rho\rangle_{Y_1'}$, we get

$$\begin{aligned}&W_{X_1Y_1'}|\rho^{(1)}\rangle_{X_1'\mathbf{Y_1'}RC}\otimes|\rho\rangle_{\mathbf{X_1'}Y_1'}\\&\qquad=W_{X_1Y_1'}|\rho^{(1)}\rangle_{X_1'Y_1'RC}\otimes|\rho\rangle_{\mathbf{X_1'}\mathbf{Y_1'}}\\&\qquad=|\sigma^{(1)}\rangle_{X_1'Y_1'RC}\otimes|\rho\rangle_{\mathbf{X_1'}\mathbf{Y_1'}}.\end{aligned}\tag{10}$$

Furthermore, by the definition of $|\theta^{(1)}\rangle$ and $|\lambda^{(1)}\rangle$ with equations (9) and (10) above, it follows that

$$\begin{aligned}&\mathfrak{h}\left(|\theta^{(1)}\rangle,\ |\lambda^{(1)}\rangle\right)\\&=\mathfrak{h}\big(V^{(1)}_{\overline{X}_1\mathbf{X_1'}X_2'A}W_{X_1Y_1'}|\rho^{(1)}\rangle_{X_1'\mathbf{Y_1'}RC}\otimes|\rho\rangle_{\mathbf{X_1'}Y_1'},\\&\qquad|\rho^{(1)}\rangle_{\mathbf{X_1'}\mathbf{Y_1'}RC}\otimes|\sigma\rangle_{X_1'Y_1'}\leq\sqrt{\epsilon_1}\big),\end{aligned}$$

where we used (10) to show that $|\theta^{(1)}\rangle$ equals the first state in the middle expression and the inequality follows from Claim A.2. This proves the base case.

Induction: For the induction let us assume that s is even (since the case for odd s is similar) and that $\mathfrak{h}\left(|\theta^{(s-1)}\rangle,\ |\lambda^{(s-1)}\rangle\right)\leq\delta_{s-1}$. Using the triangle inequality we bound

$$\begin{aligned}\mathfrak{h}\left(|\theta^{(s)}\rangle,\ |\lambda^{(s)}\rangle\right)&\leq\mathfrak{h}\left(|\theta^{(s)}\rangle,\ |\omega^{(s)}\rangle\right)\\&\quad+\mathfrak{h}\left(|\omega^{(s)}\rangle,\ |\pi^{(s)}\rangle\right)\\&\quad+\mathfrak{h}\left(|\pi^{(s)}\rangle,\ |\lambda^{(s)}\rangle\right),\end{aligned}\tag{11}$$

where

$$\begin{aligned}|\omega^{(s)}\rangle&=V^{(s)}U^{(s)}_{YBC}\Big(V^{(s-2)}\Big)^\dagger|\lambda^{(s-1)}\rangle,\text{ and}\\|\pi^{(s)}\rangle&=V^{(s)}W_{X_1'Y_1}|\rho^{(s)}\rangle_{\mathbf{X_1'}Y_1'RC}\otimes|\rho\rangle_{X_1'\mathbf{Y_1'}},\end{aligned}\tag{12}$$

with $U^{(s)}_{YBC}$ being the protocol unitary with Y as control that Bob applies in round s. Note that from the definition of the protocol we have that

$$|\sigma^{(s)}\rangle_{X_1'Y_1'RC}=U^{(s)}_{YBC}|\sigma^{(s-1)}\rangle_{X_1'Y_1'RC}.\tag{13}$$

Let us consider the first term in (11). Since Hellinger distance is unitarily invariant, we multiply both states with $V^{(s-2)}\Big(U^{(s)}_{YBC}\Big)^\dagger\big(V^{(s)}\big)^\dagger$ and using (9), (12) and (13), we get that

$$\begin{aligned}&\mathfrak{h}\left(|\theta^{(s)}\rangle,\ |\omega^{(s)}\rangle\right)\\&\overset{(12)}{=}\mathfrak{h}\left(V^{(s-2)}\Big(U^{(s)}_{YBC}\Big)^\dagger\Big(V^{(s)}\Big)^\dagger|\theta^{(s)}\rangle,\ |\lambda^{(s-1)}\rangle\right)\\&=\mathfrak{h}\left(|\theta^{(s-1)}\rangle,\ |\lambda^{(s-1)}\rangle\right)\leq\delta_{s-1},\end{aligned}$$

where we used that the first state in the middle expression equals $|\theta^{(s-1)}\rangle$:

$$\begin{aligned}&V^{(s-2)}\Big(U^{(s)}_{YBC}\Big)^\dagger\Big(V^{(s)}\Big)^\dagger|\theta^{(s)}\rangle\\&\overset{(9)}{=}V^{(s-2)}\Big(U^{(s)}_{YBC}\Big)^\dagger V^{(s-1)}|\sigma^{(s)}\rangle_{X_1'Y_1'RC}\otimes|\rho\rangle_{\mathbf{X_1'}\mathbf{Y_1'}}\\&\overset{(13)}{=}V^{(s-2)}V^{(s-1)}|\sigma^{(s-1)}\rangle_{X_1'Y_1'RC}\otimes|\rho\rangle_{\mathbf{X_1'}\mathbf{Y_1'}}\\&=V^{(s-1)}V^{(s-2)}|\sigma^{(s-1)}\rangle_{X_1'Y_1'RC}\otimes|\rho\rangle_{\mathbf{X_1'}\mathbf{Y_1'}}\overset{(9)}{=}|\theta^{(s-1)}\rangle,\end{aligned}$$

with the second equality using (13) and the fact that $U^{(s)}_{YBC}$ and $V^{(s-1)}_{\overline{X}_1\mathbf{X}'_1X'_2A}$ commute, and the third equality using that $V^{(s-1)}_{\overline{X}_1\mathbf{X}'_1X'_2A}$ and $V^{(s-2)}_{\overline{Y}_1\mathbf{Y}'_1Y'_2B}$ commute.

To bound the second term, notice that by the definition of the protocol $|\rho^{(s-1)}\rangle_{\mathbf{X}'_1\mathbf{Y}'_1RC} = U^{(s-1)}_{\mathbf{X}_1X_2AC}|\rho^{(s-2)}\rangle_{\mathbf{X}'_1\mathbf{Y}'_1RC}$ (recall $R = X_2\overline{X_2}Y_2\overline{Y_2}AB$) and therefore using (12) and (9), it follows that

$$|\omega^{(s)}\rangle = V^{(s)}U^{(s)}_{YBC}(V^{(s-2)})^\dagger U^{(s-1)}_{\mathbf{X}'_1X_2AC}|\rho^{(s-2)}\rangle_{\mathbf{X}'_1\mathbf{Y}'_1RC} \otimes |\sigma\rangle_{X'_1Y'_1}. \tag{14}$$

Now multiplying both states by the unitary $Q = \left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger V^{(s-2)}\left(U^{(s)}_{YBC}\right)^\dagger\left(V^{(s)}\right)^\dagger$ and again using unitary invariance of Hellinger, we get that

$$\begin{aligned}
&\mathfrak{h}\left(|\omega^{(s)}\rangle,\ |\pi^{(s)}\rangle\right)\\
&= \mathfrak{h}\left(Q|\omega^{(s)}\rangle,\ Q|\pi^{(s)}\rangle\right)\\
&\overset{(14)}{=} \mathfrak{h}\big(|\rho^{(s-2)}\rangle_{\mathbf{X}'_1\mathbf{Y}'_1RC}\otimes|\sigma\rangle_{X'_1Y'_1},\\
&\qquad V^{(s-2)}|\rho^{(s-2)}\rangle_{\mathbf{X}'_1Y'_1RC}\otimes|\sigma\rangle_{X'_1\mathbf{Y}'_1}\big)\\
&\leq \sqrt{\epsilon_{s-2}},
\end{aligned}$$

where the inequality follows from Claim A.2 and we simplified the second state $Q|\pi^{(s)}\rangle$ using commutativity of the pairs $\left\{\left(V^{(s-2)}_{\overline{Y}_1\mathbf{Y}'_1Y'_2B}\right)^\dagger, \left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger\right\}$ (disjoint registers), $\left\{\left(U^{(s)}_{YBC}\right)^\dagger, W_{X'_1Y_1}\right\}$ (disjoint registers except for both being controlled on the shared register Y_1), and $\left\{\left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger, W_{X'_1Y_1}\right\}$ (disjoint registers) as follows:

$$\begin{aligned}
Q|\pi^{(s)}\rangle &= \left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger V^{(s-2)}\left(U^{(s)}_{YBC}\right)^\dagger\left(V^{(s)}\right)^\dagger|\pi^{(s)}\rangle\\
&\overset{(12)}{=} V^{(s-2)}\left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger\left(U^{(s)}_{YBC}\right)^\dagger W_{X'_1Y_1}|\rho^{(s)}\rangle_{\mathbf{X}'_1Y'_1RC}\\
&\quad\otimes|\rho\rangle_{X'_1\mathbf{Y}'_1}\\
&= V^{(s-2)}\left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger W_{X'_1Y_1}\left(U^{(s)}_{YBC}\right)^\dagger|\rho^{(s)}\rangle_{\mathbf{X}'_1Y'_1RC}\\
&\quad\otimes|\rho\rangle_{X'_1\mathbf{Y}'_1}\\
&= V^{(s-2)}W_{X'_1Y_1}\left(U^{(s-1)}_{\mathbf{X}_1X_2AC}\right)^\dagger|\rho^{(s-1)}\rangle_{\mathbf{X}'_1Y'_1RC}\\
&\quad\otimes|\rho\rangle_{X'_1\mathbf{Y}'_1}\\
&= V^{(s-2)}W_{X'_1Y_1}|\rho^{(s-2)}\rangle_{\mathbf{X}'_1Y'_1RC}\otimes|\rho\rangle_{X'_1\mathbf{Y}'_1}.
\end{aligned}$$

To upper bound the third term of (11), by the definition of states $|\pi^{(s)}\rangle$ and $|\lambda^{(s)}\rangle$ (equations (12) and (9)) and Claim A.2, we get

$$\begin{aligned}
&\mathfrak{h}\left(|\pi^{(s)}\rangle,\ |\lambda^{(s)}\rangle\right)\\
&= \mathfrak{h}\big(V^{(s)}W_{X'_1Y_1}|\rho^{(s)}\rangle_{\mathbf{X}'_1Y'_1RC}\otimes|\rho\rangle_{X'_1\mathbf{Y}'_1},\\
&\qquad|\rho^{(s)}\rangle_{\mathbf{X}'_1\mathbf{Y}'_1RC}\otimes|\sigma\rangle_{X'_1Y'_1}\big)\\
&\leq \sqrt{\epsilon_s}.
\end{aligned}$$

Plugging the bounds for each of the terms back in (11), we get that

$$\begin{aligned}
\mathfrak{h}\left(|\theta^{(s)}\rangle,\ |\omega^{(s)}\rangle\right) &\leq \delta_{s-1} + \sqrt{\epsilon_{s-2}} + \sqrt{\epsilon_s}\\
&= \sqrt{\epsilon_s} + \sqrt{\epsilon_{s-1}} + 2\sum_{i=1}^{s-2}\sqrt{\epsilon_i} = \delta_s.
\end{aligned}$$

■