Near-optimal bounds on bounded-round quantum communication complexity of disjointness

Mark Braverman [*]   Ankit Garg [†]   Young Kun Ko [‡]   Jieming Mao [§]

Dave Touchette [¶]

May 11, 2015

Abstract

We prove a near optimal round-communication tradeoff for the two-party quantum communication complexity of disjointness. For protocols with r rounds, we prove a lower bound of $\tilde{\Omega}(n/r)$ on the communication required for computing disjointness of input size n, which is optimal up to logarithmic factors. The previous best lower bound was $\Omega(n/r^2)$ due to Jain, Radhakrishnan and Sen [JRS03]. Along the way, we develop several tools for quantum information complexity, one of which is a lower bound for quantum information complexity in terms of the generalized discrepancy method. As a corollary, we get that the quantum communication complexity of any boolean function f is at most $2^{O(QIC(f))}$, where $QIC(f)$ is the prior-free quantum information complexity of f (with error 1/3).

Contents

1   Introduction

We prove near-optimal bounds on the bounded-round quantum communication complexity of disjointness. Quantum communication complexity, introduced by Yao [Yao93], studies the amount of quantum communication that two parties, Alice and Bob, need to exchange in order to compute a function (usually boolean) of their private inputs. It is the natural quantum extension of classical communication complexity [Yao79]. While the inputs are classical and the end result is classical, the players are allowed to use quantum resources while communicating. The motivation for the introduction of quantum communication was to study questions in quantum computation. For example, in [Yao93], Yao used it to prove that the majority function does not have any linear size quantum formulas.

While quantum communication (with entanglement) offers only a factor of 2 savings when transmitting n bits of classical information [Hol73, BW92, CvDNT98], it can still offer super-constant savings (and sometimes exponential) in communication if the goal is just to compute a boolean function of the inputs. For total boolean functions, the best-known separation between classical and quantum communication is quadratic, for the disjointness function [KS92, Raz92, Gro96, BCW98, AA03]. It is, in fact, a major open problem whether classical and quantum communication are polynomially related for all total boolean functions. For partial functions, exponential separations are known even between one-way quantum communication and arbitrary classical communication [Raz99, KR11].

For disjointness with input size n, Grover's search [Gro96, BBHT98] can be used to obtain a quantum communication protocol (with probability of error $1/3$) with communication cost $O(\sqrt{n}\log n)$ [BCW98]. The bound was later improved to $O(\sqrt{n})$ in [AA03]. The protocols attaining this upper bound are very interactive and require $\Theta(\sqrt{n})$ rounds of interaction. The $O(\sqrt{n})$ upper bound on the quantum communication complexity of disjointness has been shown to be tight in [Raz02].

If we restrict the players to allow only r rounds of interaction, then it is not hard to use the $O(\sqrt{n})$ protocol discussed above as a black-box to obtain an $O(n/r)$ communication protocol for $n \geq r^2$. The best known lower bound was $\Omega(n/r^2)$ [JRS03]. We prove a lower bound of $\tilde{\Omega}(n/r)$, which is optimal up to logarithmic factors:

Theorem A. (Theorem 7.3, rephrased) *The r-round quantum communication complexity of $DISJ_n$ is $\Omega\left(\frac{n}{r\log^8(r)}\right)$.*

The analogous result for query complexity of quantum search, an $\Omega(n/r)$ lower bound for the number of queries when r sets of nonadaptive queries are allowed, was known before [Zal99]. Our lower bound does not give a new proof of the $\Omega(\sqrt{n})$ bound on the quantum communication complexity of disjointness [Raz02] since our proof uses that lower bound (in fact we use something much stronger, a strengthening of the strong direct product theorem for disjointness [KSDW04] due to [She12]).

There is a rich history of papers studying lower bounds on bounded-round communication complexity, for example for the pointer jumping problem [NW93, PRV01, Kla98, KNTSZ01], for sparse set disjointness [ST13], for equality [BCK14] and several other examples. Most of these lower bounds are proven via a round elimination strategy: show that an r-round protocol can be converted into an $(r-1)$-round protocol without too much increase in communication cost and error; arrive at contradition by obtaining a too-good-to-be-true 1-round or 0-round protocol. Even the result of [JRS03] can be viewed as round elimination on quantum information complexity of the

3

2-bit AND. Despite substantial effort, obtaining the optimal $\Omega(1/r)$ lower bound on the r-round quantum information complexity of AND via round elimination has remained elusive. We prove:

Theorem B. (Corollary 7.2, rephrased) *The r-round quantum information complexity of AND with prior $1/3, 1/3, 1/3, 0$ is $\Omega\left(\frac{1}{r \log^8(r)}\right)$.*

As discussed below, we obtain this result by using existing lower bounds for the communication complexity of quantum disjointness. A direct proof of a quantum information complexity lower bound for the 2-bit AND remains an intriguing open problem. In light of the fact that disjointness has a sub-linear quantum communication complexity, it is not surprising that the quantum information complexity of AND vanishes with the number of rounds. This phenomenon is closely related to the Elitzur-Vaidman bomb tester [EV93, KWHZ95], which gives a sequence of quantum measurements that allows one to test whether a bomb is loaded without detonating it. The loss of the protocol (i.e. the probability that the bomb will explode — which loosely corresponds to the amount of information revealed about the bomb) behaves like $1/r$, where r is the number of measurements performed.

Our proof relies on the notion of quantum information complexity, defined recently in [Tou15], where it is used to prove a direct sum theorem for constant round quantum communication. It is harder to manipulate quantum information than in the classical case, and tools that are standard in the classical setting are yet to be developed for the quantum case. However, it could still be useful in proving partial direct sum and direct product theorems, which we know in the classical world [BBCR10], [BRWY13]. Moreover, a model similar to that of quantum communication complexity is connected to proving SDP extension complexity lower bounds [JSWZ13]. Although the recent breakthrough for SDP lower bounds [LRS15] does not follow this direction, it is likely that a quantum information complexity viewpoint will provide further insights as information complexity has provided in the classical case (LP extension complexity) [BM13, BP13]. Further development of tools for quantum communication and information complexity is likely to further the SDP extension complexity program.

We also prove that for all boolean functions, prior-free quantum information complexity is lower bounded by the generalized discrepancy method:

Theorem C. (Theorem 5.7, rephrased) *For any boolean function f and a sufficiently small constant error $\eta > 0$, the prior-free quantum information complexity of f with error η is lower bounded by the generalized discrepancy bound for f.*

Previously no lower bounds were known on the quantum information complexity of general boolean functions. Our proof relies on the strong direct product theorem for quantum communication complexity in terms of the generalized discrepancy method [She12]. Note that in the classical setting such a result can be proven directly using zero-communication protocols [KLL+12]. It remains to be seen whether such a direct proof can be obtained in the quantum setting.

As a corollary we also get that the quantum communication complexity of any boolean function is at most exponential in the prior-free quantum information complexity.

Theorem D. (Corollary 5.8, rephrased) *For any boolean function f, quantum communication complexity of f with error $1/3$ is at most $2^{O(QIC(f,1/3)+1)}$, where $QIC(f, 1/3)$ is the prior-free quantum information complexity of f with error $1/3$.*

Note that the classical analogue of this is proven via a compression argument [Bra12], but we prove this via an indirect argument. It would be interesting to prove this directly via a quantum compression argument.

2 Proof overview and discussion

High-level strategy. At a high-level, the proof builds on the connection between quantum information complexity and quantum communication complexity of the disjointness function $DISJ_m$ with various values of m. There are two parts to the proof:

1. Suppose there is a r-round quantum protocol for disjointness of input size $n \geq r^2$ with communication cost $\frac{n}{r \cdot \text{polylog}(r)}$. Then there exists a protocol for disjointness of input size r^2 with quantum information cost $\leq o(r)$.

2. Lower bound on quantum information complexity of disjointness: we prove that the (prior-free) quantum information complexity of any boolean function is lower bounded by the generalized discrepancy method, which by results in [She07] implies that quantum information complexity of disjointness with input size r^2 is $\Omega(r)$.

Note that these two steps imply a lower bound on the bounded round quantum communication complexity of disjointness. Also the above statements are about computation with some constant error (say 1/3).

Both directions are proven via a connection between the information complexity of a problem and its communication complexity. In one direction, a protocol for a large sized disjointness can be converted into a low-information protocol for a smaller size disjointness. Using the converse direction of the connection, a low-information protocol for $DISJ_{r^2}$ leads to a protocol for many copies of the problem that violate known direct product results. The former connection has been at the heart of many classical lower bounds involving information complexity [BYJKS04, BGPW13a]. The latter connection (deriving information complexity lower bound from known communication lower bound on an "amortized" version of the problem) has been previously explored in the classical setting by [BGPW13b].

Let us start by giving a high level overview of the first step. If there is a r-round quantum protocol for disjointness of input size n with communication cost $\frac{n}{r \cdot \text{polylog}(r)}$ and 1/3 probability of error, then by a direct sum argument in [Tou15], there exists a r-round quantum protocol π for AND with 1/3 probability of error (for a worst case input) and quantum information cost $\leq \frac{1}{r \cdot \text{polylog}(r)}$ w.r.t **any distribution μ s.t. $\mu(1,1) = 0$**. Now we want to use π to obtain a low information protocol for disjointness of size r^2. One can imagine if we run π on each coordinate of the disjointness instance, we get an r-round protocol τ of information cost $\leq \frac{r}{\text{polylog}(r)}$ and also it solves disjointness with small error (assuming we first amplify the error of π to $1/r^3$ losing a log factor in information cost). However, the issue is that information cost of τ is low only w.r.t. **distributions ν supported on disjoint pairs of sets**. The information cost of τ may increase dramatically when it is run on a pair of sets with many intersections. To deal with this we use a trick used in [BGPW13a].

Note that if there are too many intersections in a disjointness instance, then the players can just subsample some of the coordinates and check for an intersection in those coordinates. Hence

we can assume wlog that the intersection size in a typical input distributed according to ν is small. This means that if we look at a typical coordinate i, the marginal distribution ν_i has small mass on $(1, 1)$. And in this case, we can run π on each coordinate. The only thing left to understand is: how does the information cost of π change if we place a small mass, say w, on $(1, 1)$? The answer to this turns out to be $r \cdot H(w)$, where π has r-rounds. Note that this is in contrast to the classical case, where the answer would be just $H(w)$. Later we will give an example of a quantum protocol for AND whose information cost does go up by $r \cdot H(w)$. Also **this is the only place where we use the fact that the protocol we started with had only r rounds**. Such a dependence is necessary here, since an $\Omega(n/r)$ lower bound for general (non-r-round) protocols would violate the $O(\sqrt{n})$ upper bound.

For the second step, we use compression along with a strong direct product theorem for quantum communication complexity of f in terms of the generalized discrepancy lower bound $GDM_{1/5}(f)$ due to Sherstov [She12]. It says that to compute k copies of a boolean function f with success probability $2^{-\Omega(k)}$, it requires at least $k \cdot GDM_{1/5}(f)$ qubits of communication (with arbitrary amount of entanglement). Note that a strong direct product theorem for quantum communication complexity of disjointness was already known [KSDW04], but we need a stronger version for our proof which shows that even computing a large fraction of the copies is hard and Sherstov's result also holds in this case[1].

Suppose there is a protocol π for a function f with quantum information cost $\leq I$ w.r.t a distribution μ and probability of error $\leq \epsilon$, then by quantum information equals amortized communication [Tou15], we get a protocol π_k for f^k which computes at least $(1 - 2\epsilon)k$ coordinates correctly with probability ≥ 0.99 (w.r.t. μ^k) and $QCC(\pi_k) \leq k \cdot I + o(k)$. To apply Sherstov's theorem, we need such a protocol which works for worst case inputs. We show how to obtain such a worst case to average case reduction, whence applying Sherstov's result gives us the lower bound on information complexity.

Discussion and open problems

In its entirety our proof shows how from a r-round protocol for disjointness, one can obtain a protocol for k copies of disjointness of size r^2. But to achieve this reduction, we have to move to information complexity, since the number of rounds r only comes up in an information theoretic context in our proof.

Thus the reduction structure of the proof is communication$\rightarrow$information$\rightarrow$communication, with the latter communication problem having a known lower bound. Lower bounds for disjointness in the classical setting [BYJKS04, BGPW13a] only do a reduction of the form communication $\rightarrow$ information, with an information complexity lower bound on the resulting problem proven directly.

Open Problem 2.1. *Give a direct proof of a lower bound for the information complexity of* $DISJ_{r^2}$.

One possible attack route would be along the lines of the proof for the classical case using zero-communication protocols [KLL$^+$12]. In the past, techniques developed for two-party quantum communication, e.g. the pattern matrix method [She07], turned out to be useful for multiparty

---

[1]We could probably base our result off the lower bound of [KSDW04], but the reduction would be considerably more complicated.

number-on-forehead communication [CA08, She14]. It could be that techniques developed for quantum information also result in similar progress.

Another natural question is whether the lower bound on the information complexity of AND can be proved using a direct argument:

Open Problem 2.2. *Give a direct proof of Theorem B.*

Even though efforts since [JRS03] to-date have been unsuccessful, it still could be possible to directly obtain Theorem B via round elimination or other techniques and that would be really interesting, since it would also yield a new proof of the lower bound for quantum communication complexity of disjointness [Raz02, She07]. The recent breakthrough results in lower bounding conditional quantum mutual information [FR14, BHOS14, BT15] should be relevant.

Remark 2.3. Our proofs can be adapted to show that the (unbounded round) zero-error quantum information complexity of AND w.r.t the prior $(1-\epsilon)/3, (1-\epsilon)/3, (1-\epsilon)/3, \epsilon$ is $\tilde{\Omega}(\sqrt{\epsilon})$. It is another intriguing question whether it is possible to have a direct proof for this. Note that this requires a global view of quantum information complexity, even though it is defined round by round. By a continuity argument this would also resolve open problem 2.2.

More generally, our understanding of the relationship between quantum information and communication complexity is in its early stages of development. Questions of interactive protocol compression occupy a central position in understanding the connection between classical information and communication complexity [BBCR10, Bra12, GKR14]. In particular, [BBCR10] shows that a protocol π with information cost I and communication cost C can be compressed into a protocol with communication cost $\tilde{O}(\sqrt{I \cdot C})$. It remains open whether this (or an analagous) fact is true in the quantum setting:

Open Problem 2.4. *Given a quantum protocol π over a distribution μ of inputs whose communication cost is C and whose quantum information cost is I, can π be simulated (with a small error) using a quantum protocol π' whose communication cost is $\tilde{O}(\sqrt{I \cdot C})$?*

3 Preliminaries

3.1 Quantum Information Theory

We use the following notation for quantum theory; see [Wat13, Wil13] for more details. We associate a quantum register A with a corresponding vector space, also denoted by A. We only consider finite-dimensional vector spaces. A state of quantum register A is represented by a density operator $\rho \in \mathcal{D}(A)$, with $\mathcal{D}(A)$ the set of all unit trace, positive semi-definite linear operators mapping A into itself. We say that a state ρ is pure if it is a projection operator, i.e. $(\rho^A)^2 = \rho^A$. For a pure state ρ, we might use the pure state formalism, and represent ρ by the vector $|\rho\rangle$ it projects upon, i.e. $\rho = |\rho\rangle\langle\rho|$; this is well-defined up to an irrelevant phase factor.

A quantum channel from quantum register A into quantum register B is represented by a super-operator $\mathcal{N}^{A \to B} \in \mathcal{C}(A, B)$, with $\mathcal{C}(A, B)$ the set of all completely positive, trace-preserving linear operators from $\mathcal{D}(A)$ into $\mathcal{D}(B)$. If $A = B$, we might simply write $\mathcal{N}^A$, and when systems are clear from context, we might drop the superscripts. For channels $\mathcal{N}_1 \in \mathcal{C}(A, B), \mathcal{N}_2 \in \mathcal{C}(B, C)$, we denote their composition as $\mathcal{N}_2 \circ \mathcal{N}_1 \in \mathcal{C}(A, C)$, with action $(\mathcal{N}_2 \circ \mathcal{N}_1)(\rho) = \mathcal{N}_2(\mathcal{N}_1(\rho))$ on any state $\rho \in \mathcal{D}(A)$. We might drop the $\circ$ symbol if the composition is clear from context. For A and B

isomorphic, we denote the identity mapping as $I^{A\to B}$, with some implicit choice for the change of basis. For $\mathcal{N}^{A_1\to B_1} \otimes I^{A_2\to B_2} \in \mathcal{C}(A_1\otimes A_2, B_1\otimes B_2)$, we might abbreviate this as $\mathcal{N}$ and leave the identity channel implicit when the meaning is clear from context.

An important subset of $\mathcal{C}(A,B)$ when A and B are isomorphic spaces is the set of unitary channels $\mathcal{U}(A,B)$, the set of all maps $U \in \mathcal{C}(A,B)$ with an adjoint map $U^\dagger \in \mathcal{C}(B,A)$ such that $U^\dagger \circ U = I^A$ and $U \circ U^\dagger = I^B$. More generally, if $\dim(B) \geq \dim(A)$, we denote by $\mathcal{U}(A,B)$ the set of isometric channels, i.e. the set of all maps $V \in \mathcal{C}(A,B)$ with an adjoint map $V^\dagger \in \mathcal{C}(B,A)$ such that $V^\dagger \circ V = I^A$. Another important example of channel that we use is the partial trace $\mathrm{Tr}_B(\cdot) \in C(A\otimes B, A)$ which effectively gets rid of the B subsystem to obtain the marginal state on subsystem A. Fixing an orthonormal basis $\{|b\rangle\}$ for B, we can write the action of Tr_B on any $\rho^{AB} \in \mathcal{D}(A\otimes B)$ as $\mathrm{Tr}_B(\rho^{AB}) = \sum_b \langle b\,|\,\rho^{AB}\,|b\rangle$. Note that the action of Tr_B is independent of the choice of basis chosen to represent it, so we unambiguously write $\rho^A = \mathrm{Tr}_B(\rho^{AB})$. We also use the notation $\mathrm{Tr}_{\neg A} = \mathrm{Tr}_B$ to express that we want to keep only the A register.

Fixing a basis also allows us to talk about classical states and joint states: $\rho \in \mathcal{D}(B)$ is classical (with respect to this basis) if it is diagonal in basis $\{|b\rangle\}$, i.e. $\rho = \sum_b p_B(b) \cdot |b\rangle\langle b|$ for some probability distribution p_B. More generally, subsystem B of ρ^{AB} is said to be classical if we can write $\rho^{AB} = \sum_b p_B(b) \cdot |b\rangle\langle b|^B \otimes \rho_b^A$ for some $\rho_b^A \in \mathcal{D}(A)$. An important example of a channel mapping a quantum system to a classical one is the measurement channel Δ_B, defined as $\Delta_B(\rho) = \sum_b \langle b\,|\,\rho\,|b\rangle \cdot |b\rangle\langle b|^B$ for any $\rho \in \mathcal{D}(B)$. Note that for any state $\rho \in \mathcal{D}(B_1\otimes B_2\otimes C\otimes R)$ of the form

$$|\rho\rangle^{B_1 B_2 CR} = \sum_b \sqrt{p_B(b)} \cdot |b\rangle^{B_1} |b\rangle^{B_2} |\rho_b\rangle^{CR},$$

we have $\mathrm{Tr}_{B_2}(\rho^{B_1 B_2 CR}) = \sum_b p_B(b) \cdot |b\rangle\langle b|^{B_1} \otimes \rho_b^{CR}$ and $\mathrm{Tr}_{B_2 R}(\rho^{B_1 B_2 CR}) = \sum_b p_B(b) \cdot |b\rangle\langle b|^{B_1} \otimes \rho_b^C$, with the state on B_1 classical in both cases. Often, $A, B, C, \cdots$ will be used to discuss general systems, while $X, Y, Z, \cdots$ will be reserved for classical systems, or quantum systems like B_1 and B_2 above that are classical once one of them is traced out, and can be thought of as containing a quantum copy of the classical content of one another.

For a state $\rho^A \in \mathcal{D}(A)$, a purification is a pure state $\rho^{AR} \in \mathcal{D}(A\otimes R)$ satisfying $\mathrm{Tr}_R(\rho^{AR}) = \rho^A$. If R has dimension at least that of A, then such a purification always exists. For a given R, all purifications are equivalent up to a unitary on R, and more generally, if $\dim(R') \geq \dim(R)$ and $\rho_1^{AR}, \rho_2^{AR'}$ are two purifications of ρ^A, then there exists an isometry $V_\rho^{R\to R'}$ such that $\rho_2^{AR'} = V_\rho(\rho_1^{AR})$. For a channel $\mathcal{N} \in C(A,B)$, an isometric extension is a unitary $U_\mathcal{N} \in U(A, A'\otimes B)$ with $\mathrm{Tr}_{A'}(U_\mathcal{N}(\rho^A)) = \mathcal{N}(\rho^A)$ for all ρ^A. Such an extension always exists provided A' is of dimension at least $\dim(A)^2$. For the measurement channel Δ_B, an isometric extension is given by $U_\Delta = \sum_b |b\rangle^{B'} |b\rangle^B \langle b|^B$.

The notion of distance we use is the trace distance, defined for two states $\rho_1, \rho_2 \in \mathcal{D}(A)$ as the sum of the absolute values of the eigenvalues of their difference:

$$\|\rho_1 - \rho_2\|_A = \mathrm{Tr}(|\rho_1 - \rho_2|).$$

It has an operational interpretation as four times the best bias possible in a state discrimination test between ρ_1 and ρ_2. The subscript tells on which subsystems the trace distance is evaluated, and remaining subsystems might need to be traced out. We use the following results about trace distance. For proofs of these and other standard results in quantum information theory that we

8

use, see [Wil13]. The trace distance is monotone under noisy channels: for any $\rho_1, \rho_2 \in \mathcal{D}(A)$ and $\mathcal{N} \in \mathcal{C}(A, B)$,

$$\|\mathcal{N}(\rho_1) - \mathcal{N}(\rho_2)\|_B \leq \|\rho_1 - \rho_2\|_A. \tag{1}$$

For isometries, the inequality becomes an equality, a property called isometric invariance of the trace distance. Hence, for any $\rho_1, \rho_2 \in \mathcal{D}(A)$ and any $U \in \mathcal{U}(A, B)$, we have

$$\|U(\rho_1) - U(\rho_2)\|_B = \|\rho_1 - \rho_2\|_A. \tag{2}$$

Also, the trace distance cannot be increased by adjoining an uncorrelated system: for any $\rho_1, \rho_2 \in \mathcal{D}(A), \sigma \in \mathcal{D}(B)$

$$\|\rho_1 \otimes \sigma - \rho_2 \otimes \sigma\|_{AB} = \|\rho_1 - \rho_2\|_A. \tag{3}$$

The trace distance obeys a property that we call joint linearity: for a classical system X and two states $\rho_1^{XA} = p_X(x) \cdot |x\rangle\langle x|^X \otimes \rho_{1,x}^A$ and $\rho_2^{XA} = p_X(x) \cdot |x\rangle\langle x|^X \otimes \rho_{2,x}^A$,

$$\|\rho_1 - \rho_2\|_{XA} = \sum_x p_X(x) \|\rho_{1,x} - \rho_{2,x}\|_A. \tag{4}$$

The measure of information that we use is the von Neumann entropy, defined for any state $\rho \in \mathcal{D}(A)$ as

$$H(A)_\rho = -\operatorname{Tr}(\rho \log \rho),$$

in which we take the convention that $0 \log 0 = 0$, justified by a continuity argument. The logarithm $\log$ is taken in base 2, while the natural logarithm is denoted $\ln$. Note that H is invariant under isometries applied on ρ. If the state to be evaluated is clear from context, we might drop the subscript. Conditional entropy for a state $\rho^{ABC} \in \mathcal{D}(A \otimes B \otimes C)$ is then defined as

$$H(A|B) = H(AB) - H(B),$$

mutual information as

$$I(A; B) = H(A) - H(A|B),$$

and conditional mutual information as

$$I(A; B|C) = H(A|C) - H(A|BC).$$

Note that mutual information and conditional mutual information are symmetric in interchange of A, B, and invariant under a local isometry applied to A, B or C. For any pure bipartite state $\rho^{AB} \in \mathcal{D}(A \otimes B)$, the entropy on each subsystem is the same:

$$H(A) = H(B). \tag{5}$$

Since all purifications are equivalent up to an isometry on the purification registers, we get that for any two pure states $|\phi\rangle^{ABCR'}$ and $|\psi\rangle^{ABCR}$ such that $\phi^{ABC} = \psi^{ABC}$,

$$I(C; R'|B)_\phi = I(C; R|B)_\psi. \tag{6}$$

9

For isomorphic A, A', a maximally entangled state $\psi \in \mathcal{D}(A \otimes A')$ is a pure state satisfying $H(A) = H(A') = \log \dim(A) = \log \dim(A')$. For a system A of dimension $\dim(A)$ and any $\rho \in \mathcal{D}(A \otimes B \otimes C)$, we have the bounds

$$0 \leq H(A) \leq \log \dim(A), \tag{7}$$
$$-H(A) \leq H(A|B) \leq H(A), \tag{8}$$
$$0 \leq I(A;B) \leq 2H(A), \tag{9}$$
$$0 \leq I(A;B|C) \leq 2H(A). \tag{10}$$

If A or B is a classical system, we get the tighter bounds

$$0 \leq H(A|B), \tag{11}$$
$$I(A;B) \leq H(A), \tag{12}$$
$$I(A;B|C) \leq H(A). \tag{13}$$

The conditional mutual information satisfies a chain rule: for any $\rho \in \mathcal{D}(A \otimes B \otimes C \otimes D)$,

$$I(AB;C|D) = I(A;C|D) + I(B;C|AD). \tag{14}$$

For product states $\rho^{A_1 B_1 C_1 A_2 B_2 C_2} = \rho_1^{A_1 B_1 C_1} \otimes \rho_2^{A_2 B_2 C_2}$, entropy is additive,

$$H(A_1 A_2) = H(A_1) + H(A_2), \tag{15}$$

and so there is no conditional mutual information between product system,

$$I(A_1; A_2|B_1 B_2) = 0, \tag{16}$$

and conditioning on a product system is useless,

$$I(A_1; B_1|C_1 A_2) = I(A_1; B_1|C_1). \tag{17}$$

More generally,

$$I(A_1 A_2; B_1 B_2|C_1 C_2) = I(A_1; B_1|C_1) + I(A_2; B_2|C_2). \tag{18}$$

Two important properties of the conditional mutual information are non-negativity, equivalent to strong subadditivity, and the data processing inequality. For any $\rho \in \mathcal{D}(A \otimes B \otimes C)$ and $\mathcal{N} \in \mathcal{C}(B, B')$, with $\sigma = \mathcal{N}(\rho)$,

$$I(A;B|C)_\rho \geq 0, \tag{19}$$
$$I(A;B|C)_\rho \geq I(A;B'|C)_\sigma. \tag{20}$$

For classical systems, conditioning is equivalent to taking an average: for any $\rho^{ABCX} = \sum_x p_X(x) \cdot |x\rangle\langle x|^X \otimes \rho_x^{ABC}$, for a classical system X and some appropriate $\rho_x \in \mathcal{D}(A \otimes B \otimes C)$,

$$H(A|BX)_\rho = \sum_x p_X(x) \cdot H(A|B)_{\rho_x}, \tag{21}$$

$$I(A;B|CX)_\rho = \sum_x p_X(x) \cdot I(A;B|C)_{\rho_x}. \tag{22}$$

3.2 Quantum Communication Model

The model for communication complexity that we consider is the following. For a given bipartite relation $T \subset X \times Y \times Z_A \times Z_B$ and input distribution μ on $X \times Y$, Alice and Bob are given input registers A_{in}, B_{in} containing their classical input $x \in X, y \in Y$ at the outset of the protocol, respectively, and they output registers A_{out}, B_{out} containing their classical output $z_A \in Z_A, z_B \in Z_B$ at the end of the protocol, respectively, which should satisfy the relation T. We generally allow for some small error ϵ in the output, which will be formalized below. In this distributional communication complexity setting, the input is a classical state $\rho = \sum_{x \in X, y \in Y} \mu(x, y) \cdot |x\rangle\langle x|^{A_{in}} \otimes |y\rangle\langle y|^{B_{in}}$, similarly for the output $\Pi(\rho) = \sum_{z_A \in Z_A, z_B \in Z_B} p_{Z_A Z_B}(z_A, z_B) \cdot |z_A\rangle\langle z_A|^{A_{out}} \otimes |z_B\rangle\langle z_B|^{B_{out}}$ of the protocol Π implementing the relation, and the error parameter corresponds to the average probability of failure $\sum_{x,y} \mu(x, y) \cdot [(x, y, \Pi(x, y)) \notin R] \leq \epsilon$.

A r-round protocol Π for implementing relation T on input $\rho^{A_{in} B_{in}}$ is defined by a sequence of isometries $U_1, \cdots, U_{r+1}$ along with a pure state $\psi \in \mathcal{D}(T_A \otimes T_B)$ shared between Alice and Bob, for arbitrary finite dimensional registers T_A, T_B. For appropriate finite dimensional memory registers $A_1, A_3, \cdots A_{r-1}, A'$ held by Alice, $B_2, B_4, \cdots B_{r-2}, B'$ held by Bob, and communication registers $C_1, C_2, C_3, \cdots C_r$ exchanged by Alice and Bob, we have $U_1 \in \mathcal{U}(A_{in} \otimes T_A, A_1 \otimes C_1), U_2 \in \mathcal{U}(B_{in} \otimes T_B \otimes C_1, B_2 \otimes C_2), U_3 \in \mathcal{U}(A_1 \otimes C_2, A_3 \otimes C_3), U_4 \in \mathcal{U}(B_2 \otimes C_3, B_4 \otimes C_4), \cdots, U_r \in \mathcal{U}(B_{r-2} \otimes C_{r-1}, B_{out} \otimes B' \otimes C_r), U_{r+1} \in \mathcal{U}(A_{r-1} \otimes C_r, A_{out} \otimes A')$. We adopt the convention that, in the first round, $B_1 = B_0 = B_{in} \otimes T_B$, in even rounds $B_i = B_{i-1}$, and in odd rounds $A_i = A_{i-1}$. In this way, in round i, after application of U_i, Alice holds register A_i, Bob holds register B_i and the communication register is C_i. We slightly abuse notation and also write Π to denote the channel implemented by the protocol, i.e.

$$\Pi(\rho) = \operatorname{Tr}_{A'B'}(U_{r+1} U_r \cdots U_2 U_1 (\rho \otimes \psi)). \tag{23}$$

To formally define the error, we introduce a purification register R. For a classical input $\rho^{A_{in} B_{in}} = \sum_{x \in X, y \in Y} \mu(x, y) \cdot |x\rangle\langle x|^{A_{in}} \otimes |y\rangle\langle y|^{B_{in}}$ like we consider here, we can always take this purification to be of the form $|\rho\rangle^{A_{in} B_{in} R} = \sum_{x \in X, y \in Y} \sqrt{\mu(x, y)} |x\rangle^{A_{in}} |y\rangle^{B_{in}} |xy\rangle^{R_1} |xy\rangle^{R_2}$, for an appropriately chosen partition of R into R_1, R_2. Note that if we trace out the the R_2 register, then we are left with a classical state such that R_1 contains a copy of the joint input. Then we say that a protocol Π for implementing relation T on input $\rho^{A_{in} B_{in}}$, with purification $\rho^{A_{in} B_{in} R}$, has average error $\epsilon \in [0, 1]$ if $P_e^{\mu} = \operatorname{Pr}_{\mu, \Pi}[\Pi(\rho^{A_{in} B_{in} R_1}) \notin T] \leq \epsilon$. We denote the set of all such protocols as $\mathcal{T}(T, \mu, \epsilon)$. If we want to restrict this set to bounded round protocols with r rounds, we write $\mathcal{T}^r(T, \mu, \epsilon)$. The worst case error of a protocol is $P_e^w = \max_\mu P_e^\mu$, in which it is sufficient to optimize over all atomic distributions μ. We denote by $\mathcal{T}(T, \epsilon)$ the set of all protocols implementing relation T with worst case error at most ϵ, and by $\mathcal{T}^r(T, \epsilon)$ if we restrict this set to r-round protocols.

Let us formally define the different quantities that we work with.

Definition 3.1. For a protocol Π as defined above, we define the *quantum communication cost* of Π as

$$QCC(\Pi) = \sum_i \log \dim(C_i).$$

Note that we do not require that $\dim(C_i) = 2^k$ for some $k \in \mathbb{N}$, as is usually done. This will not affect our definition on information cost and complexity, but might affect the quantum

communication complexity by at most a factor of two, without affecting the round complexity. The corresponding notions of quantum communication complexity of a relation are:

Definition 3.2. For a relation $T \subset X \times Y \times Z_A \times Z_B$, an input distribution μ on $X \times Y$ and an error parameter $\epsilon \in [0, 1]$, we define the ϵ-error *quantum communication complexity* of T on input μ as

$$QCC(T, \mu, \epsilon) = \min_{\Pi \in \mathcal{T}(T, \mu, \epsilon)} QCC(\Pi),$$

and the worst-case ϵ-error *quantum communication complexity* of T as

$$QCC(T, \epsilon) = \min_{\Pi \in \mathcal{T}(T, \epsilon)} QCC(\Pi),$$

Remark 3.3. For any $T, \mu, 0 \le \epsilon_1 \le \epsilon_2 \le 1$, the following holds:

$$QCC(T, \mu, \epsilon_2) \le QCC(T, \mu, \epsilon_1),$$
$$QCC(T, \epsilon_2) \le QCC(T, \epsilon_1).$$

We have the following definitions for bounded round quantum communication complexity, and a similar remark holds.

Definition 3.4. For a relation $T \subset X \times Y \times Z_A \times Z_B$, an input distribution μ on $X \times Y$, an error parameter $\epsilon \in [0, 1]$ and a bound $r \in \mathbb{N}$ on the number of rounds, we define the r-round, ϵ-error *quantum communication complexity* of T on input μ as

$$QCC^r(T, \mu, \epsilon) = \min_{\Pi \in \mathcal{T}^r(T, \mu, \epsilon)} QCC(\Pi),$$

and r-round, worst-case ϵ-error *quantum communication complexity* of T as

$$QCC^r(T, \epsilon) = \min_{\Pi \in \mathcal{T}^r(T, \epsilon)} QCC(\Pi),$$

3.3 Quantum Information Complexity

We use the notion of quantum information complexity as defined in [Tou15]. The register R is the purification register, invariant throughout the protocol since we consider local isometric processing. Note that, as noted before when considering a $R_1 R_2$ partition for R, for classical input distributions, the purification register can be thought of as containing a (quantum) copy of the classical input. The definition is however invariant under the choice of R and corresponding purification.

Definition 3.5. For a protocol Π and a state ρ with purification held in system R, we define the *quantum information cost* of Π on input ρ as

$$QIC(\Pi, \rho) = \sum_{i>0, odd} \frac{1}{2} I(C_i; R|B_i) + \sum_{i>0, even} \frac{1}{2} I(C_i; R|A_i).$$

Definition 3.6. For a relation $T \subset X \times Y \times Z_A \times Z_B$, an input distribution μ on $X \times Y$, an error parameter $\epsilon \in [0,1]$ and a number of round r, we define the ϵ-error *quantum information complexity* of T on input μ as

$$QIC(T, \mu, \epsilon) = \inf_{\Pi \in \mathcal{T}(T, \mu, \epsilon)} QIC(\Pi, \mu),$$

and the r-round, ϵ-error *quantum information complexity* of T on input μ as

$$QIC^r(T, \mu, \epsilon) = \inf_{\Pi \in \mathcal{T}^r(T, \mu, \epsilon)} QIC(\Pi, \mu),$$

The following properties of quantum information cost and complexity were proved in Ref. [Tou15].

Lemma 3.7. *For any protocol Π and input distribution μ, the following holds:*

$$0 \leq QIC(\Pi, \mu) \leq QCC(\Pi).$$

Lemma 3.8. *For a relation $T \subset X \times Y \times Z_A \times Z_B$, an input distribution μ on $X \times Y$, an error parameter $\epsilon \in [0,1]$ and a number of round r, the following holds:*

$$0 \leq QIC(T, \mu, \epsilon) \leq QCC(T, \mu, \epsilon),$$
$$0 \leq QIC^r(T, \mu, \epsilon) \leq QCC^r(T, \mu, \epsilon).$$

Lemma 3.9. *For any two protocols Π^1 and Π^2 with r_1 and r_2 rounds, respectively, there exists a r-round protocol Π_2, satisfying $\Pi_2 = \Pi^1 \otimes \Pi^2, r = \max(r_1, r_2)$, such that the following holds for any corresponding input states ρ^1, ρ^2:*

$$QIC(\Pi_2, \rho^1 \otimes \rho^2) = QIC(\Pi^1, \rho^1) + QIC(\Pi^2, \rho^2).$$

Lemma 3.10. *For any r-round protocol Π_2 and any input states $\rho^1 \in \mathcal{D}(A_{in}^1 \otimes B_{in}^1), \rho_2 \in \mathcal{D}(A_{in}^2 \otimes B_{in}^2)$, there exist r-round protocols Π^1, Π^2 satisfying $\Pi^1(\cdot) = \mathrm{Tr}_{A_{out}^2 B_{out}^2} \circ \Pi_2(\cdot \otimes \rho^2)$, $\Pi^2(\cdot) = \mathrm{Tr}_{A_{out}^1 B_{out}^1} \circ \Pi_2(\rho^1 \otimes \cdot)$, and the following holds:*

$$QIC(\Pi^1, \rho^1) + QIC(\Pi^2, \rho^2) = QIC(\Pi_2, \rho^1 \otimes \rho^2).$$

Lemma 3.11. *For any $p \in [0,1]$, any two protocols Π^1, Π^2 with r_1, r_2 rounds, respectively, there exists a r-round protocol Π satisfying $\Pi = p\Pi^1 + (1-p)\Pi^2, r = \max(r_1, r_2)$, such that the following holds for any state ρ:*

$$QIC(\Pi, \rho) = pQIC(\Pi^1, \rho) + (1-p)QIC(\Pi^2, \rho).$$

Corollary 3.12. *For any $p \in [0,1], T$ and $\epsilon, \epsilon_1, \epsilon_2 \in [0,1]$ satisfying $\epsilon = p\epsilon_1 + (1-p)\epsilon_2$, for any bound $r = \max(r_1, r_2), r_1, r_2 \in \mathbb{N}$ on the number of rounds and for any input distribution μ on $X \times Y$, the following holds:*

$$QIC(T, \mu, \epsilon) \leq pQIC(T, \mu, \epsilon_1) + (1-p)QIC(T, \mu, \epsilon_2),$$
$$QIC^r(T, \mu, \epsilon) \leq pQIC^{r_1}(T, \mu, \epsilon_1) + (1-p)QIC^{r_2}(T, \mu, \epsilon_2).$$

Lemma 3.13. *Let ν be a distribution over input states ρ and denote $\overline{\rho} := \mathbf{E}_{\rho \sim \nu} \rho$. Then for any protocol π,*

$$\mathbb{E}_{\rho \sim \nu}[QIC(\pi, \rho)] \leq QIC(\pi, \overline{\rho})$$

Lemma 3.14. *For any r-round protocol Π, any input distribution μ with copies of x, y in R_1, and any $\epsilon \in (0, 2], \delta > 0$, there exists a large enough $n_0(\Pi, \rho, \epsilon, \delta)$ such that for any $n \geq n_0$, there exists a r-round protocol Π_n satisfying*

$$\|\Pi_n((\rho^{A_{in}B_{in}R_1})^{\otimes n}) - \Pi^{\otimes n}((\rho^{A_{in}B_{in}R_1})^{\otimes n})\|_{(A_{out}B_{out}R_1)^{\otimes n}} \leq \epsilon,$$

$$\frac{1}{n}QCC(\Pi_n) \leq QIC(\Pi, \rho) + \delta.$$

3.4 Generalized Discrepancy Method

Generalized discrepancy method, also known as smooth discrepancy method, is one of the strongest methods for proving lower bounds for quantum communication.

Definition 3.15. *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean function. The δ-generalized discrepancy bound of f, denoted by $GDM_\delta(f)$, is defined as:*

$$GDM_\delta(f) = \max\{GDM_\delta^\mu(f) : \mu \text{ a distribution over } \mathcal{X} \times \mathcal{Y}\}$$

$$GDM_\delta^\mu(f) = \max\{\log\left(\frac{1}{\text{disc}^\mu(g)}\right), g : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}, \Pr_{(x,y)\sim\mu}[f(x, y) \neq g(x, y)] \leq \delta\}$$

$$\text{disc}^\mu(g) = \max\left\{|\sum_{(x,y)\in R}(-1)^{g(x,y)} \cdot \mu(x, y)| : R \in \mathcal{R}\right\}$$

Here $\mathcal{R}$ is the set of combinatorial rectangles $\mathcal{A} \times \mathcal{B}$, $\mathcal{A} \subseteq \mathcal{X}, \mathcal{B} \subseteq \mathcal{Y}$. We state two results on the generalized discrepancy method, both due to Sherstov [She07, She12], which we will use to lower bound the quantum information complexity of disjointness. The first is a threshold direct product result that will be useful to prove that the generalized discrepancy method is a lower bound on the quantum information complexity of boolean functions, and the second is a lower bound on the generalized discrepancy for the disjointness function.

Theorem 3.16 ([She12]). *Let $\epsilon_{sh} > 0$ be a small enough absolute constant. Then for any boolean function f, the following communication problem requires $\Omega(nGDM_{1/5}(f))$ qubits of communication (with arbitrary entanglement): Solving with probability $2^{-\epsilon_{sh}n}$, at least $(1 - \epsilon_{sh})n$ among n instances of f.*

The disjointness function is defined as follows: for $x, y \in \{0, 1\}^n \times \{0, 1\}^n$, $DISJ_n(x, y) = 1$ if for all $i \in [n]$, $x_i \wedge y_i = 0$, and 0 otherwise. We will need the following theorem.

Theorem 3.17 ([She07]). $GDM_{1/5}(DISJ_n) \geq \Omega(\sqrt{n})$

4 Properties of Quantum Information Complexity

In this section, we prove general results about quantum information complexity that we use to obtain the main results. These may be of independent interest.

4.1 Prior-free Quantum Information Complexity

We want to define a sensible notion of quantum information complexity for classical tasks. Like in the classical setting [Bra12], there are two sensible orderings for the optimization over inputs and protocols. We provide the two corresponding definitions and then investigate the link between them. We denote by $\mathcal{D}_{XY}$ the set of all distributions μ on input space $X \times Y$.

Definition 4.1. The *max-distributional quantum information complexity* of a relation T with error $\epsilon \in [0, 1]$ is

$$QIC_D(T, \epsilon) = \max_{\mu \in \mathcal{D}_{XY}} QIC(T, \mu, \epsilon).$$

When restricting to r-round protocols, it is

$$QIC_D^r(T, \epsilon) = \max_{\mu \in \mathcal{D}_{XY}} QIC^r(T, \mu, \epsilon).$$

Definition 4.2. The *quantum information complexity* of a relation T with error $\epsilon \in [0, 1]$ is

$$QIC(T, \epsilon) = \inf_{\Pi \in \mathcal{T}(T, \epsilon)} \max_{\mu \in \mathcal{D}_{XY}} QIC(\Pi, \mu).$$

When restricting to r-round protocols, it is

$$QIC^r(T, \epsilon) = \inf_{\Pi \in \mathcal{T}^r(T, \epsilon)} \max_{\mu \in \mathcal{D}_{XY}} QIC(\Pi, \mu).$$

Lemma 4.3 (Information lower bounds communication)**.** *For any relation T, error parameter $\epsilon \in [0, 1]$, and number of rounds $r \in \mathbb{N}$, the following holds:*

$$QIC^r(T, \epsilon) \leq QCC^r(T, \epsilon),$$
$$QIC(T, \epsilon) \leq QCC(T, \epsilon).$$

Proof. Let Π be a protocol computing T correctly except with probability ϵ on all input and satisfying $QCC(\Pi) = QCC(T, \epsilon)$. We get the result by noting that $QIC(T, \epsilon) \leq \max_\mu QIC(\Pi, \mu) \leq QCC(\Pi)$. □

Clearly, $QIC_D(T, \epsilon) \leq QIC(T, \epsilon)$, and $QIC_D^r(T, \epsilon) \leq QIC^r(T, \epsilon)$. We prove that we can almost reverse the quantifiers. The proof idea follows the lines of the proof of Theorem 3.5 in Ref. [Bra12], but special care must be taken for quantum protocols. The idea we use is to take an ϵ-net over $\mathcal{D}_{XY}$, and then take a δ-optimal protocol for each distribution in the net. To extend this result to the unbounded round quantum setting, we adapt a compactness argument from Ref. [BGPW13a], itself adapted from Ref. [Ter72]. The following results will be used.

Lemma 4.4 (Continuity in average error)**.** *Quantum information complexity is continuous in the error. This holds uniformly in the input. That is, for all T, r and $\epsilon, \delta > 0$, there exists $\epsilon' \in (0, \epsilon)$ such that for all $\epsilon'' \in (\epsilon', \epsilon)$ and for all μ,*

$$|QIC(T, \mu, \epsilon - \epsilon'') - QIC(T, \mu, \epsilon)| \leq \delta,$$
$$|QIC^r(T, \mu, \epsilon - \epsilon'') - QIC^r(T, \mu, \epsilon)| \leq \delta.$$

Proof. Note that we can drop the absolute values and also work at ϵ' since quantum information complexity is non-increasing in the error, i.e. $QIC(T, \mu, \epsilon) \leq QIC(T, \mu, \epsilon - \epsilon'') \leq QIC(T, \mu, \epsilon - \epsilon')$. Let $0 < p < \frac{1}{2}$ and use Corollary 3.12 with $\epsilon_1 = 0, \epsilon_2 = \epsilon, \epsilon' = p\epsilon$ for the current ϵ. We get

$$QIC(T, \mu, \epsilon - \epsilon') \leq pQIC(T, \mu, 0) + (1 - p)QIC(T, \mu, \epsilon)$$
$$\leq pQCC(T, 0) + QIC(T, \mu, \epsilon).$$

Rearranging terms, we get

$$|QIC(T, \mu, \epsilon - \epsilon') - QIC(T, \mu, \epsilon)| \leq \frac{\epsilon'}{\epsilon} QCC(T, 0).$$

This bound is independent of μ, and goes to zero as p and ϵ' do, so the result follows. The bounded round result is proved in the same way, obtaining $QCC^r(T, 0)$ in the final bound instead. $\qquad\square$

Lemma 4.5 (Convexity in error). *For any $p \in [0, 1], T$ and $\epsilon, \epsilon_1, \epsilon_2 \in [0, 1]$ satisfying $\epsilon = p\epsilon_1 + (1 - p)\epsilon_2$ and for any bound $r = \max(r_1, r_2), r_1, r_2 \in \mathbb{N}$ on the number of rounds, the following holds:*

$$QIC(T, \epsilon) \leq pQIC(T, \epsilon_1) + (1 - p)QIC(T, \epsilon_2),$$
$$QIC^r(T, \epsilon) \leq pQIC^{r_1}(T, \epsilon_1) + (1 - p)QIC^{r_2}(T, \epsilon_2).$$

Proof. The proof is similar to the one for the analogous result with fixed input. Given $\delta > 0$, let Π^1 and Π^2 be protocols satisfying, for all μ, for $i \in \{1, 2\}, \Pi^i \in \mathcal{T}(T, \epsilon_i), QIC(\Pi^i, \mu) \leq QIC(T, \epsilon_i) + \delta$, and take the corresponding protocol Π of Lemma 3.11. First, it holds that protocol Π successfully accomplish its task, i.e. it implements task T on all inputs with error bounded by $\epsilon = p\epsilon_1 + (1 - p)\epsilon_2$. We must now verify that the quantum information cost satisfies the convexity property:

$$QIC(T, \epsilon) \leq \max_\mu QIC(\Pi, \mu)$$
$$= \max_\mu \left(pQIC(\Pi^1, \mu) + (1 - p)QIC(\Pi^2, \mu) \right)$$
$$\leq p \max_\mu QIC(\Pi^1, \mu) + (1 - p) \max_\mu QIC(\Pi^2, \mu)$$
$$\leq pQIC(T, \epsilon_1) + (1 - p)QIC(T, \epsilon_2) + 2\delta.$$

Keeping track of rounds, we get the bounded round result. $\qquad\square$

Corollary 4.6 (Continuity in error). *Quantum information complexity is continuous in the error. That is, for all T, r and $\epsilon, \delta > 0$, there exists $\epsilon' \in (0, \epsilon)$ such that for all $\epsilon'' \in (\epsilon', \epsilon)$*

$$|QIC(T, \epsilon - \epsilon'') - QIC(T, \epsilon)| \leq \delta,$$
$$|QIC^r(T, \epsilon - \epsilon'') - QIC^r(T, \epsilon)| \leq \delta.$$

Lemma 4.7 (Quasi-convexity in input). *For any $p \in [0, 1]$, define $\rho = p\rho_1 + (1 - p)\rho_2$ for any two input states ρ_1, ρ_2. Then the following holds for any r-round protocol Π:*

$$QIC(\Pi, \rho) \geq pQIC(\Pi, \rho_1) + (1 - p)QIC(\Pi, \rho_2)$$
$$QIC(\Pi, \rho) \leq pQIC(\Pi, \rho_1) + (1 - p)QIC(\Pi, \rho_2) + rH(p).$$

16

Proof. The first inequality is Lemma 3.13, and the second is obtained by keeping track of the remainder terms discarded in its proof. Let R be a register holding a purification of ρ_1 and ρ_2, then we can purify ρ with two copies S_1, S_2 of a selector reference register, such that $|\rho\rangle^{A_{in}B_{in}RS_1S_2} = \sqrt{p}\,|\rho_1\rangle^{A_{in}B_{in}R}\,|1\rangle^{S_1}\,|1\rangle^{S_2} + \sqrt{1-p}\,|\rho_2\rangle^{A_{in}B_{in}R}\,|2\rangle^{S_1}\,|2\rangle^{S_2}$. We can then expand each term as

$$I(C_i; RS_1S_2|B_i)_\rho = I(C_i; S_1|B_i)_\rho + I(C_i; R|B_iS_1)_\rho + I(C_i; S_2|B_iRS_1)_\rho,$$

and similarly for terms conditioning on Alice's systems A_i. The result follows by summing over all rounds since

$$I(C_i; R|B_iS_1)_\rho = pI(C_i; R|B_i)_{\rho_1} + (1-p) \cdot I(C_i; R|B_i)_{\rho_2},$$

and then $H(S) = H(p)$ upper bounds the two remainder terms in each of the r rounds. $\square$

Lemma 4.8 (Continuity in input). *Quantum information cost for r-round protocols is uniformly continuous in the input distribution. This holds uniformly over all r-round protocols over input $X \times Y$. That is, for all $r, |X|, |Y|,$ and $\epsilon > 0$, there exists $\delta > 0$ such that for all μ_1 and μ_2 that are δ-close and all r-round protocols Π,*

$$|QIC(\Pi, \mu_1) - QIC(\Pi, \mu_2)| \leq \epsilon.$$

Proof. Let $\delta > 0$ and fix μ_1 and μ_2 that are δ-close. We can then write, for some common part μ_0 and remainder parts μ_1', μ_2',

$$\mu_1 = (1 - \delta)\mu_0 + \delta\mu_1',$$
$$\mu_2 = (1 - \delta)\mu_0 + \delta\mu_2',$$
$$\mu_0(x, y) = \frac{\min(\mu_1(x, y), \mu_2(x, y))}{\sum_{x', y'} \min(\mu_1(x', y'), \mu_2(x', y'))}.$$

Using the bounds in the lemma above once on each of μ_1 and μ_2, we get

$$\begin{aligned} QIC(\Pi, \mu_1) &\leq (1 - \delta)QIC(\Pi, \mu_0) + \delta QIC(\Pi, \mu_1') + rH(\delta) \\ &\leq (1 - \delta)QIC(\Pi, \mu_0) + \delta QIC(\Pi, \mu_2') + \delta QIC(\Pi, \mu_1') + rH(\delta) \\ &\leq QIC(\Pi, \mu_2) + \delta \cdot r(\log|X| + \log|Y|) + rH(\delta). \end{aligned}$$

Similarly, we get a bound on $QIC(\Pi, \mu_2)$ in terms of $QIC(\Pi, \mu_1)$, so the following holds:

$$|QIC(\Pi, \mu_1) - QIC(\Pi, \mu_2)| \leq \delta \cdot r(\log|X| + \log|Y|) + rH(\delta).$$

This bound is independent of μ_1, μ_2, depends on Π only through r and $|X|, |Y|$, and goes to zero as δ does, so the result follows. $\square$

Corollary 4.9. *Suppose we have a r-round protocol Π for AND. Then,*

$$QIC(\Pi, \mu) \leq QIC(\Pi, \mu_0) + O(rH(w)) \tag{24}$$

where $w = \mu(1, 1) \leq 1/2$, $\mu_0(1, 1) = 0$, and $\mu_0(x_i, y_i) = \frac{1}{1-w}\mu(x_i, y_i)$ otherwise.

Proof. This just follows from the proof of lemma 4.8, since the input size is constant. $\square$

Theorem 4.10. *For a relation $T \subset X \times Y \times Z_A \times Z_B$, an error parameter $\epsilon \in (0,1)$, a number of rounds r and each value $\alpha \in (0,1)$,*

$$QIC^r(T, \frac{\epsilon}{\alpha}) \leq \frac{QIC_D^r(T, \epsilon)}{1 - \alpha}.$$

Proof. Fix T, r, ϵ, α and denote $I = QIC_D^r(T, \epsilon)$. For any $\delta_1 \in (0,1)$, we want to prove the existence of a protocol $\Pi \in \mathcal{T}^r(T, \frac{\epsilon}{\alpha} \cdot (1 + 2\delta_1))$ satisfying $QIC(\Pi, \mu) \leq \frac{I \cdot (1+2\delta_1)}{1-\alpha}$ for all $\mu \in \mathcal{D}_{XY}$. This shows that $QIC^r(T, \frac{\epsilon}{\alpha} \cdot (1 + 2\delta_1)) \leq \frac{I}{1-\alpha} \cdot (1 + 2\delta_1)$, and then by continuity of quantum information complexity in the error, we get the result by taking δ_1 to 0. The proof follows along the lines of the one for the analogous result for classical information complexity [Bra12], using a minimax argument. We take extra care to account for the continuum of quantum protocols, the round-by-round definition of quantum information cost, and the fact that we do not have a bound on the size of the entanglement. Let $\delta_2 \in (0, \epsilon\delta_1)$ satisfy the following two properties for all μ_1, μ_2 that are δ_2-close, and for all r-round protocols Π:

$$|QIC(\Pi, \mu_1) - QIC(\Pi, \mu_2)| \leq I \cdot \frac{\delta_1}{10}, \tag{25}$$

$$|QIC^r(T, \mu_1, \epsilon - \delta_2) - QIC^r(T, \mu_1, \epsilon)| \leq I \cdot \frac{\delta_1}{10}. \tag{26}$$

The first inequality is possible by Lemma 4.8, i.e. by the uniform continuity of quantum information cost in the input, uniformly over all r-rounds protocols, and the second is possible by Lemma 4.4, i.e. the continuity of quantum information complexity in the error, uniformly over all inputs. Fix a finite δ_2-net for $\mathcal{D}_{XY}$, that we denote N_{XY}. For each $\mu \in N_{XY}$, fix a protocol $\Pi_\mu \in \mathcal{T}^r(T, \mu, \epsilon - \delta_2)$ such that $QIC(\Pi_\mu, \mu) \leq QIC^r(T, \mu, \epsilon - \delta_2) \cdot (1 + \frac{\delta_1}{10})$ and denote the set of all such protocols P_N. We then have $|P_N| = |N_{XY}| < \infty$, and we get using (26) that

$$
\begin{aligned}
QIC(\Pi_\mu, \mu) &\leq QIC^r(T, \mu, \epsilon - \delta_2) \cdot (1 + \frac{\delta_1}{10}) \\
&\leq \left(QIC^r(T, \mu, \epsilon) + I \cdot \frac{\delta_1}{10} \right)(1 + \frac{\delta_1}{10}) \\
&\leq I(1 + \frac{\delta_1}{10})^2 \\
&\leq I(1 + \frac{\delta_1}{2}).
\end{aligned}
\tag{27}
$$

We define the following two-player zero-sum game over these two sets. Player A comes up with a quantum protocol $\Pi \in P_N$. Player B comes up with a distribution $\mu \in N_{XY}$. Player B's payoff is given by

$$P_B(\Pi, \mu) = (1 - \alpha) \cdot \frac{QIC(\Pi, \mu)}{I} + \alpha \cdot \frac{Pr_\mu[\Pi \notin T]}{\epsilon},$$

and then player A's is given by $P_A(\Pi, \mu) = -P_B(\Pi, \mu)$. We first show the following.

Claim 4.11. *The value of the game for player B is bounded by $1 + \delta_1$.*

Proof. Let ν_B be a probability distribution over N_{XY} representing a mixed strategy for player B. To prove the claim, it suffices to show that there is a protocol $\Pi \in P_N$ such that $\mathbb{E}_{\nu_B}[P_B(\Pi, \mu)] < 1 + \delta_1$. Let $\bar{\mu}$ be the distribution corresponding to averaging over ν_B, that is

$$\bar{\mu}(x, y) = \mathbb{E}_{\nu_B} \mu(x, y).$$

Let $\mu' \in N_{XY}$ be a distribution that is δ_2-close to $\bar{\mu}$, and $\Pi' \in P_N$ the corresponding protocol. We will show that Π' is also good for $\bar{\mu}$. We first have

$$
\begin{aligned}
Pr_{\bar{\mu}}[\Pi' \notin T] &\le Pr_{\mu'}[\Pi' \notin T] + \delta_2 \\
&\le \epsilon - \delta_2 + \delta_2 \\
&= \epsilon,
\end{aligned}
$$

in which the first inequality follows from the fact that $\bar{\mu}$ and μ' are δ_2-close and the second inequality from the fact that $\Pi' \in P_N$ is the protocol corresponding to $\mu' \in N_{XY}$, i.e. $\Pi' \in \mathcal{T}^r(T, \mu', \epsilon - \delta_2)$. We also have

$$
\begin{aligned}
QIC(\Pi', \bar{\mu}) &\le QIC(\Pi', \mu') + I \cdot \frac{\delta_1}{2} \\
&\le I \cdot (1 + \delta_1),
\end{aligned}
$$

in which the first inequality follows from (25) and the second from the fact that $\Pi' \in P_N$ is the protocol corresponding to $\mu' \in N_{XY}$ along with (27). We obtain

$$
\begin{aligned}
\mathbb{E}_{\nu_B}[P_B(\Pi', \mu)] &= \mathbb{E}_{\nu_B}\Big[(1 - \alpha) \cdot \frac{QIC(\Pi', \mu)}{I} + \alpha \cdot \frac{Pr_\mu[\Pi' \notin T]}{\epsilon}\Big] \\
&= (1 - \alpha) \cdot \mathbb{E}_{\nu_B}\Big[\frac{QIC(\Pi', \mu)}{I}\Big] + \alpha \cdot \frac{Pr_{\bar{\mu}}[\Pi' \notin T]}{\epsilon} \\
&\le (1 - \alpha) \cdot \Big[\frac{QIC(\Pi', \bar{\mu})}{I}\Big] + \alpha \cdot \frac{Pr_{\bar{\mu}}[\Pi' \notin T]}{\epsilon} \\
&< (1 - \alpha) \cdot (1 + \delta_1) + \alpha \\
&< 1 + \delta_1,
\end{aligned}
$$

in which the first equality is by definition, the second by linearity of expectation, the first inequality is by Lemma 3.13, i.e. concavity of quantum information cost in the input state, and the second inequality is by the above results about Π'. This concludes the proof of the claim. $\qquad\square$

By the minimax theorem for zero-sum games, the above claim implies that there exists a probability distribution ν_A over P_N representing a mixed strategy for player A and such that the value of the game for player B is at most $1 + \delta_1$. That is, for all $\mu \in N_{XY}$,

$$\mathbb{E}_{\nu_A}(P_B(\Pi, \mu)) < 1 + \delta_1.$$

Let $\bar{\Pi} = \mathbb{E}_{\nu_A}(\Pi)$ be the r-round protocol obtained by publicly averaging over ν_A, as per Lemma 3.11. This is the protocol we are looking for. The following claim holds.

Claim 4.12. *For all $\mu \in \mathcal{D}_{XY}$, $(1 - \alpha) \cdot \frac{QIC(\bar{\Pi}, \mu)}{I} + \alpha \cdot \frac{Pr_\mu[\bar{\Pi} \notin T]}{\epsilon} < 1 + 2\delta_1$.*

Proof. Fix any $\mu \in \mathcal{D}_{XY}$, and let $\mu' \in N_{XY}$ be a distribution that is δ_2-close to μ. Then we obtain

$$
\begin{aligned}
(1-\alpha) \cdot \frac{QIC(\bar{\Pi}, \mu)}{I} + \alpha \cdot \frac{Pr_\mu[\bar{\Pi} \notin T]}{\epsilon} &\leq (1-\alpha) \cdot \frac{QIC(\bar{\Pi}, \mu') + I\delta_1}{I} + \alpha \cdot \frac{Pr_{\mu'}[\bar{\Pi} \notin T] + \delta_2}{\epsilon} \\
&= (1-\alpha) \cdot \frac{QIC(\bar{\Pi}, \mu')}{I} + \alpha \cdot \mathbb{E}_{\nu_A} \frac{Pr_{\mu'}[\Pi \notin T]}{\epsilon} \\
&\quad + (1-\alpha) \cdot \delta_1 + \alpha \cdot \frac{\delta_2}{\epsilon} \\
&\leq (1-\alpha) \cdot \mathbb{E}_{\nu_A}\Big[\frac{QIC(\Pi, \mu')}{I}\Big] + \alpha \cdot \mathbb{E}_{\nu_A}\Big[\frac{Pr_{\mu'}[\Pi \notin T]}{\epsilon}\Big] + \delta_1 \\
&= \mathbb{E}_{\nu_A}[P_B(\Pi, \mu')] + \delta_1 \\
&< 1 + 2\delta_1,
\end{aligned}
$$

in which the first inequality follows from (25) and the fact that μ, μ' are δ_2-close, the first equality is because we take expectation over a probability, the second inequality is because $\delta_2 \leq \epsilon \cdot \delta_1$ and by Lemma 3.11, i.e. by the convexity of quantum information cost in the protocol, the second equality is by linearity of expectation and the definition of $P_B(\Pi, \mu')$, and the last inequality is because ν_A represents the mixed strategy obtained by the minimax theorem. Since this holds for all $\mu \in \mathcal{D}_{XY}$, this conclude the proof of the claim. □

To conclude the proof of the theorem, we first note that the above claim implies that for all $\mu \in \mathcal{D}_{XY}$,

$$
QIC(\bar{\Pi}, \mu) \leq \frac{I}{1-\alpha}(1 + 2\delta_1),
$$

so $\bar{\Pi}$ satisfies the quantum information cost property we are looking for. Is left to verify that it also has low error on all inputs. The above claim also implies that for all μ,

$$
Pr_\mu[\bar{\Pi} \notin T] \leq \frac{\epsilon}{\alpha} \cdot (1 + 2\delta_1).
$$

Letting μ run over all atomic distributions, we get the desired error property, and so

$$
QIC^r(T, \frac{\epsilon}{\alpha} \cdot (1 + 2\delta_1)) \leq \frac{I}{1-\alpha}(1 + 2\delta_1),
$$

as desired. □

Theorem 4.13. *For a relation $T \subset X \times Y \times Z_A \times Z_B$, an error parameter $\epsilon \in (0,1)$ and each value $\alpha \in (0,1)$,*

$$
QIC(T, \frac{\epsilon}{\alpha}) \leq \frac{QIC_D(T, \epsilon)}{1-\alpha}.
$$

Proof. Let $I = QIC_D(T, \epsilon)$, and denote by $P_e^\mu(\Pi)$ the average error of Π for computing T on μ, and by P_T the set of all protocols over the same input and output spaces as T. Then for any Π, $P_e^\mu(\Pi)$ is continuous in μ by properties of the statistical distance. Given $\delta > 0$, define

$$
A(\Pi) = \{\mu \in \mathcal{D}_{XY} : QIC(\Pi, \mu) \geq I + 2 \cdot \delta \text{ or } P_e^\mu(\Pi) \geq \epsilon + \delta\}.
$$

20

By continuity of $QIC(\Pi, \mu)$ and $P_e^\mu(\Pi)$ in μ, these sets are closed for all $\Pi \in P_T$. Then, by definition of I, for all μ there exists $\Pi_\mu \in \mathcal{T}(T, \mu, \epsilon)$ such that $QIC(\Pi_\mu, \mu) \leq I + \delta$, and so $\cap_{\Pi \in P_T} A(\Pi) = \emptyset$. Since $\mathcal{D}_{XY}$ is compact and the sets $A(\Pi)$ are closed, we get that there exists a finite set $Q \subset P_T$ such that $\cap_{\Pi \in Q} A(\Pi) = \emptyset$. We get that for all μ, there exists $\Pi_\mu \in Q$ such that $QIC(\Pi_\mu, \mu) < I + 2\delta$ and $P_e^\mu(\Pi_\mu) < \epsilon + \delta$. Let $r_M = \max\{r : \text{there is } \Pi \in Q \text{ with } r \text{ rounds }\}$, then

$$
\begin{aligned}
I + 2\delta &\geq \max_\mu \min_{\Pi \in Q \cap \mathcal{T}(T, \mu, \epsilon + \delta)} QIC(\Pi, \mu) \\
&\geq QIC_D^{r_M}(T, \epsilon + \delta) \\
&\geq (1 - \alpha) \cdot QIC^{r_M}(T, \frac{\epsilon}{\alpha} + \frac{\delta}{\alpha}) \\
&\geq (1 - \alpha) \cdot QIC(T, \frac{\epsilon}{\alpha} + \frac{\delta}{\alpha}).
\end{aligned}
$$

The result follows by continuity of QIC and by taking δ to zero. $\qquad\square$

4.2 Subadditivity

Lemma 4.14. *For any two protocols Π^1, Π^2 with r_1, r_2 rounds, respectively, there exists a r-round protocol Π_2, satisfying $\Pi_2 = \Pi^1 \otimes \Pi^2, r = \max(r_1, r_2)$, such that the following holds for any joint input state $\rho_{12} \in \mathcal{D}(A_{in}^1 \otimes B_{in}^1 \otimes A_{in}^2 \otimes B_{in}^2)$:*

$$
QIC(\Pi_2, \rho_{12}) \leq QIC(\Pi^1, \rho_1) + QIC(\Pi^2, \rho_2),
$$

with $\rho_1 = \mathrm{Tr}_{A_{in}^2 B_{in}^2}(\rho_{12})$ and $\rho_2 = \mathrm{Tr}_{A_{in}^1 B_{in}^1}(\rho_{12})$.

Proof. Given protocols Π^1 and Π^2, we assume without loss of generality that $r_1 \geq r_2$, and we define the protocol Π_2 in the following way.

1. Run protocols Π^1, Π^2 in parallel for r_2 rounds, on corresponding input registers $A_{in}^1, B_{in}^1, A_{in}^2, B_{in}^2$ until Π^2 has finished.

2. Finish running protocol Π^1

3. Take as output the output registers $A_{out}^1, B_{out}^1, A_{out}^2, B_{out}^2$ of both Π^1 and Π^2.

It is clear that the channel that Π_2 implements is $\Pi_2 = \Pi^1 \otimes \Pi^2$, and the number of rounds satisfies $r = \max(r_1, r_2)$, so is left to analyze its quantum information cost on input ρ_{12}. Let R_{12} be a purifying register such that $\rho_{12}^{A_{in}^1 B_{in}^1 A_{in}^2 B_{in}^2 R_{12}}$ is a pure state. Also, denote the purified joint state in round i as $(\rho_{12}^i)^{A_i^1 B_i^1 C_i^1 A_i^2 B_i^2 C_i^2 R_{12}}$, and the local state for protocol Π^1 as

$$
(\rho_1^i)^{A_i^1 B_i^1 C_i^1} = \mathrm{Tr}_{A_i^2 B_i^2 C_i^2 R_{12}}((\rho_{12}^i)^{A_i^1 B_i^1 C_i^1 A_i^2 B_i^2 C_i^2 R_{12}}), \tag{28}
$$

and similarly for that of protocol Π^2. Notice that for all i, $(\rho_1^i)^{A_i^1 B_i^1 C_i^1}$ is purified by $(\rho_1^i)^{A_i^1 B_i^1 C_i^1 A_{in}^2 B_{in}^2 R_{12}} \otimes \phi_2^{T_A^2 T_B^2}$, with $A_{in}^2 B_{in}^2 R_{12}$ the registers of state ρ_{12} before application of the unitaries corresponding to Π^1, and ϕ_2 is the pure entangled state used in Π_2. If we denote, for $i \geq r_2 + 1$, $A_i^2 = A_{out}^2 \otimes (A')^2, B_i^2 = B_{out}^2 \otimes (B')^2$, then by the definition of QIC and application of chain rule,

$$
2 \cdot QIC(\Pi_2, \rho_{12}) = \sum_{i=1, i \text{ odd}}^{r_2} I(C_i^1 C_i^2 ; R_{12} | B_i^1 B_i^2)_{\rho_{12}} + \sum_{i=1, i \text{ even}}^{r_2} I(C_i^1 C_i^2 ; R_{12} | A_i^1 A_i^2)_{\rho_{12}}
$$

21

$$+ \sum_{i=r_2+1, i \ odd}^{r_1} I(C_i^1; R_{12} | B_i^1 B_i^2)_{\rho_{12}} + \sum_{i=r_2+1, i \ even}^{r_1} I(C_i^1; R_{12} | A_i^1 A_i^2)_{\rho_{12}}$$

$$= \sum_{i=1, i \ odd}^{r_2} I(C_i^2; R_{12} | B_i^1 B_i^2 C_i^1)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_2} I(C_i^2; R_{12} | A_i^1 A_i^2 C_i^1)_{\rho_{12}}$$

$$+ \sum_{i=1, i \ odd}^{r_1} I(C_i^1; R_{12} | B_i^1 B_i^2)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; R_{12} | A_i^1 A_i^2)_{\rho_{12}}.$$

Now for protocol Π^1, as noted above, the registers $A_{in}^2 B_{in}^2 R_{12} T_A^2 T_B^2$ purify $(\rho_1^i)^{A_i^1 B_i^1 C_i^1}$ for all i, so

$$2 \cdot QIC(\Pi^1, \rho_1) = \sum_{i=1, i \ odd}^{r_1} I(C_i^1; A_{in}^2 B_{in}^2 R_{12} T_A^2 T_B^2 | B_i^1)_{\rho_1} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; A_{in}^2 B_{in}^2 R_{12} T_A^2 T_B^2 | A_i^1)_{\rho_1}$$

$$= \sum_{i=1, i \ odd}^{r_1} I(C_i^1; A_i^2 B_i^2 C_i^2 R_{12} | B_i^1)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; A_i^2 B_i^2 C_i^2 R_{12} | A_i^1)_{\rho_{12}}$$

$$= \sum_{i=1, i \ odd}^{r_1} I(C_i^1; B_i^2 | B_i^1)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; A_i^2 | A_i^1)_{\rho_{12}}$$

$$+ \sum_{i=1, i \ odd}^{r_1} I(C_i^1; R_{12} | B_i^1 B_i^2)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; R_{12} | A_i^1 A_i^2)_{\rho_{12}}$$

$$+ \sum_{i=1, i \ odd}^{r_1} I(C_i^1; A_i^2 C_i^2 | B_i^1 B_i^2 R_{12})_{\rho_{12}} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; B_i^2 C_i^2 | A_i^1 A_i^2 R_{12})_{\rho_{12}}$$

$$\geq \sum_{i=1, i \ odd}^{r_1} I(C_i^1; R_{12} | B_i^1 B_i^2)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_1} I(C_i^1; R_{12} | A_i^1 A_i^2)_{\rho_{12}},$$

in which the first equality is by definition, the second is by isometric invariance of the conditional quantum mutual information (CQMI), the third by the chain rule for CQMI, and the inequality is by non-negativity of CQMI. Similarly for protocol Π^2, with a slightly different application of the chain rule, we get

$$2 \cdot QIC(\Pi^2, \rho_2) = \sum_{i=1, i \ odd}^{r_2} I(C_i^2; A_{in}^1 B_{in}^1 R_{12} T_A^1 T_B^1 | B_i^2)_{\rho_2} + \sum_{i=1, i \ even}^{r_2} I(C_i^2; A_{in}^1 B_{in}^1 R_{12} T_A^1 T_B^1 | A_i^2)_{\rho_2}$$

$$= \sum_{i=1, i \ odd}^{r_2} I(C_i^2; A_i^1 B_i^1 C_i^1 R_{12} | B_i^2)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_2} I(C_i^2; A_i^1 B_i^1 C_i^1 R_{12} | A_i^2)_{\rho_{12}}$$

$$= \sum_{i=1, i \ odd}^{r_2} I(C_i^2; B_i^1 C_i^1 | B_i^2)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_2} I(C_i^2; A_i^1 C_i^1 | A_i^2)_{\rho_{12}}$$

$$+ \sum_{i=1, i \ odd}^{r_2} I(C_i^2; R_{12} | B_i^1 B_i^2 C_i^1)_{\rho_{12}} + \sum_{i=1, i \ even}^{r_2} I(C_i^2; R_{12} | A_i^1 A_i^2 C_i^1)_{\rho_{12}}$$

$$+ \sum_{i=1,i \ odd}^{r_2} I(C_i^2; A_i^1 | B_i^1 B_i^2 C_i^1 R_{12})_{\rho_{12}} + \sum_{i=1,i \ even}^{r_2} I(C_i^2; B_i^2 | A_i^1 A_i^2 C_i^1 R_{12})_{\rho_{12}}$$

$$\geq \sum_{i=1,i \ odd}^{r_2} I(C_i^2; R_{12} | B_i^1 B_i^2 C_i^1)_{\rho_{12}} + \sum_{i=1,i \ even}^{r_2} I(C_i^2; R_{12} | A_i^1 A_i^2 C_i^1)_{\rho_{12}}.$$

The result then follows by comparing terms. $\qquad\square$

4.3 Reducing the Error for Functions

Similarly to communication, it is possible to reduce the error when computing functions without increasing too much the information.

Lemma 4.15. *For any function f and error parameter $\epsilon > 0$, the following holds:*

$$QIC(f, \epsilon) \leq O\big(\log 1/\epsilon \cdot QIC(f, 1/3)\big).$$

Proof. Given $\delta > 0$, let Π be a protocol computing f correctly except with probability $1/3$ on every input and satisfying $QIC(\Pi, \mu) \leq QIC(f, 1/3) + \delta$ for all μ. Let $n \in O(\log 1/\epsilon)$ be given by the Chernoff bound such that protocol Π_n running Π n times in parallel as per Lemma 4.14, with each input being a copy of the instance to f, and taking a majority vote (with arbitrary tie-breaking) computes f correctly except with probability ϵ on every input. This n can be chosen independently of δ. We now argue on the quantum information cost of Π_n. Consider an arbitrary distribution μ for f, and let μ_n be the distribution once the n copies have been made. If we denote the marginal for the i-th copy by μ^i, then $\mu^i = \mu$. By Lemma 4.14 and an easy induction, we then get that

$$\begin{aligned} QIC(f, \epsilon) &\leq QIC(\Pi_n, \mu_n) \\ &\leq nQIC(\Pi, \mu) \\ &\leq n(QIC(f, 1/3) + \delta). \end{aligned}$$

The result follows by taking δ to 0. $\qquad\square$

4.4 Reduction from DISJ to AND

With the following definition, the above proof also establishes the following corollary.

Definition 4.16. *For all $r \in \mathbb{N}, \epsilon \in [0, 1]$,*

$$QIC_0^r(AND, \epsilon) = \inf_{\Pi \in \mathcal{T}^r(AND, \epsilon)} \max_{\mu_0} QIC(\Pi, \mu_0),$$

in which the maximum ranges over all μ_0 satisfying $\mu_0(1, 1) = 0$.

Corollary 4.17. *For any $\epsilon > 0$ and $r \in \mathbb{N}$,*

$$QIC_0^r(AND, \epsilon) \leq O\big(\log 1/\epsilon \cdot QIC_0^r(AND, 1/3)\big).$$

We provide a slight variant of the argument of [Tou15] to obtain a low information protocol for AND from a protocol for disjointness.

Lemma 4.18. *For any n, r, ϵ and μ_0 such that $\mu_0(1,1) = 0$,*

$$\inf_{\Pi_A \in \mathcal{T}^r(AND, \epsilon)} QIC(\Pi_A, \mu_0) \leq \inf_{\Pi_D \in \mathcal{T}^r(DISJ_n, \epsilon)} \frac{1}{n} QIC(\Pi_D, \mu_0^{\otimes n}).$$

Proof. Let $I_n = \inf_{\Pi_D \in \mathcal{T}^r(DISJ_n, \epsilon)} QIC(\Pi_D, \mu_0^{\otimes n})$. We prove the result by induction on n. The base case is trivial since $DISJ_1 = \neg AND$, and so a protocol to compute $DISJ_1$ with error ϵ can be used to compute AND with error ϵ and vice-versa. In particular, we get $I_1 = \inf_{\Pi_A \in \mathcal{T}^r(AND, \epsilon)} QIC(\Pi_A, \mu_0)$. For the induction, suppose the result holds for $DISJ_{n-1}$, we will use Lemma 3.10 to go from $DISJ_n$ to $DISJ_1$ and $DISJ_{n-1}$. Indeed, given $\delta > 0$ and Π_D computing $DISJ_n$ with error ϵ and satisfying $QIC(\Pi_D, \mu_0^{\otimes n}) \leq I_n + \delta$, we can use Lemma 3.10 with $\rho_1 = \mu_0, \rho_2 = \mu_0^{\otimes n-1}$ and then it is clear that Π^1 computes $DISJ_1$ with error ϵ and Π^2 computes $DISJ_{n-1}$ with error ϵ. We get

$$\begin{aligned}
I_n + \delta &\geq QIC(\Pi_D, \mu_0^{\otimes n}) \\
&= QIC(\Pi^1, \mu_0) + QIC(\Pi^2, \mu_0^{\otimes n-1}) \\
&\geq I_1 + I_{n-1} \\
&\geq n I_1.
\end{aligned}$$

$\square$

The following lemma is very similar to Theorem 4.10. The only difference is that the distributions we consider are restricted and on the right hand side the error of the protocol is measured in the worst case. Since the error is worst case, there is no loss in the error, and the payoff function would be simply $P_B(\Pi, \mu) = QIC(\Pi, \mu)/I$.

Lemma 4.19.
$$QIC_0^r(AND, \epsilon) = \max_{\mu_0, \mu_0(1,1)=0} \inf_{\Pi \in \mathcal{T}^r(AND, \epsilon)} QIC(\Pi, \mu_0)$$

Lemma 4.20. *For all $r, n \in \mathbb{N}$,*

$$QCC^r(DISJ_n, 1/3) \geq n \cdot QIC_0^r(AND, 1/3)$$

Proof. The result follows from the following chain of inequality:

$$\begin{aligned}
QCC^r(DISJ_n, 1/3) &\geq QIC^r(DISJ_n, 1/3) \\
&\geq \max_{\mu_0} \inf_{\Pi_D \in \mathcal{T}^r(DISJ_n, 1/3)} QIC(\Pi_D, \mu_0^{\otimes n}) \\
&\geq \max_{\mu_0} \inf_{\Pi_A \in \mathcal{T}^r(AND, 1/3)} n \cdot QIC(\Pi_A, \mu_0) \\
&\geq n \cdot QIC_0^r(AND, 1/3).
\end{aligned}$$

The first inequality is by Lemma 4.3, the second since, on the r.h.s., the maximization is over a smaller set of product distributions with $\mu_0(1,1) = 0$ and the minimization over a larger set of protocols, the third is by Lemma 4.18, and the last is by Lemma 4.19. $\square$

5 Lower bound on QIC by generalized discrepancy method

5.1 Compression

Definition 5.1. We say that $QCC(f^k, \mu^k, \eta_1 k, \eta_2) \leq C$ if there exists a protocol π for f^k s.t. $QCC(\pi) \leq C$ and

$$Pr[\pi \text{ computes } \geq \eta_1 k \text{ coordinates correctly}] \geq 1 - \eta_2$$

Here the probability is both over the distribution μ^k and the randomness of protocol (which includes the randomness due to quantum measurements). We don't require the protocol to declare which coordinates were computed correctly.

Lemma 5.2. *If there exists a protocol Π for f with error $\leq \epsilon$ w.r.t μ s.t. $QIC(\Pi, \mu) = I$, then for all $\epsilon', \delta > 0$, there exists $k_0(\Pi, \mu, \epsilon', \delta)$ such that for all $k \geq k_0$, $QCC(f^k, \mu^k, (1-2\epsilon)k, e^{-2\epsilon^2 k} + \epsilon') \leq k(I + \delta)$.*

Proof. Suppose $(E_1, \ldots, E_k)$ is the vector of indicator random variables of the errors in various coordinates of $\Pi^{\otimes k}$ i.e. $E_i = 1$ if error occurred on the ith coordinate. Also look at Π_k obtained from lemma 3.14 for large enough k with parameters $2\epsilon', \delta$ and where ρ is μ. Suppose $(E_1', \ldots, E_k')$ is the vector of errors for Π_k. According to lemma 3.14, Π_k satisfies the following:

$$\mathbb{E}_{((x_1,\ldots,x_k),(y_1,\ldots,y_k))\sim\mu^k} ||\Pi_k((x_1, \ldots, x_k), (y_1, \ldots, y_k)) - \Pi^{\otimes k}((x_1, \ldots, x_k), (y_1, \ldots, y_k))||_1 \leq 2\epsilon'$$

Hence it follows that

$$||(E_1, \ldots, E_k) - (E_1', \ldots, E_k')||_{\text{TV}} \leq \epsilon'$$

Here $||P - Q||_{\text{TV}}$ is the total variation distance between the distributions P and Q (we are not distinguishing between random variables and their distributions). Since $Pr[\sum_i E_i \geq 2\epsilon k] \leq e^{-2\epsilon^2 k}$ by Chernoff bounds, it follows that

$$\Pr\left[\sum_i E_i' \geq 2\epsilon k\right] \leq e^{-2\epsilon^2 k} + \epsilon'$$

which implies the lemma along with the fact that $QCC(\Pi_k) \leq (I + \delta)k$. $\square$

5.2 Average case to worst case

In this section, we prove the following lemma which turns a protocol for average case input to a protocol for worst case input.

Lemma 5.3. *Suppose $f_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is an arbitrary boolean function. Let $k \geq 2^{5n}$ and $\epsilon > 10k^{-0.005}$. Assume for any product input distribution μ^k, there exists a protocol π_{μ^k} with $QCC\left(\pi_{\mu^k}\right) \leq l$ that computes at least $(1-\alpha)k$ coordinates of f_n^k correctly with probability at least γ. Then there exists a protocol τ s.t. for any input $((x_1, \cdots, x_k), (y_1, \cdots, y_k))$, for any integer $c \geq 3$ and constant $\epsilon > 0$, τ computes at least $\left(1 - 2^{-c/2} - c\alpha\right)k$ coordinates of f_n^k correctly with probability at least $\frac{1}{2}\left(\left(\frac{\gamma}{(1+\epsilon)^k}\right)^c - 2^{-2^{2-2ck}}\right)$. Also $QCC(\tau) \leq c \cdot l + o(k)$.*

Proof. In this lemma, we want to construct a protocol τ which works for an arbitrary input based on protocols which work on product input distributions (product across coordinates). The main idea of the proof is that corresponding to any input $((x_1, ..., x_k), (y_1, ..., y_k))$ (x_i and y_i are inputs of a f_n instance and have n bits), we can associate a μ, which is the empirical distribution:

$$\mu(x, y) = \frac{\# \text{ of } i, (x_i, y_i) = (x, y)}{k}.$$

So it makes sense to construct τ from π_{μ^k}. The players can simulate μ^k by sampling independent coordinates from their input (with replacement). However the issue is that the players don't know μ, so they have no idea what π_{μ^k} is. So in the actual protocol Alice and Bob will first sample some coordinates to get an estimate $\tilde{\mu}$ of μ and then run protocol $\pi_{\tilde{\mu}^k}$. The protocol τ is described in Protocol 1.

Inputs: $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_k)$

1. Get an estimate $\tilde{\mu}$ of μ.

2. Alice and Bob use shared randomness to obtain random independent samples from $[k]$, $j_1, \ldots, j_{ck}$. Run the protocol $\pi_{\tilde{\mu}^k}$ c times. In the tth iteration, the protocol is run on inputs $(x_{j_{(t-1)k+1}}, \ldots, x_{j_{tk}}), (y_{j_{(t-1)k+1}}, \ldots, y_{j_{tk}})$. In the process we obtain answers for various coordinates (some of the coordinates will be sampled multiple times and we will obtain multiple answers for them).

3. If a coordinate was sampled in the previous step, output the answer $\pi_{\tilde{\mu}^k}$ gave for it. If they got multiple results on one coordinate, they will output the first one. If a coordinate was not sampled, output 0 on that coordinate.

Protocol 1: Protocol τ

Now let's analyze this protocol. We first need the following two lemmas to show how to get an estimate $\tilde{\mu}$ of μ.

Lemma 5.4. *After communicating $O(k^{0.52} \log k)$ bits, for some specific input (x, y), with success probability at least $1 - 1/k$, Alice and Bob know $\mu(x, y)$ exactly if $\mu(x, y) \cdot k < k^{0.02}$, otherwise Alice and Bob know that $\mu(x, y) \cdot k \geq k^{0.02}$.*

Proof. In [BCW98], they showed that to compute the disjointness between two inputs of length k, the quantum communication complexity is $O(\sqrt{k} \log k)$. The corresponding protocol has constant error rate and will find one intersection place. We will use this protocol to solve our problem by the following reduction. For each input (x_i, y_i), we set $a_i = 1_{x_i = x}$ and $b_i = 1_{y_i = y}$. Then finding (x, y) in the input is just like finding intersection between $a = (a_1, ..., a_k)$ and $b = (b_1, ..., b_k)$. Protocol 2 shows how to finish the task described in the lemma.

Let's analyze this protocol. First its quantum communication cost is clear to be $O(k^{0.52} \log k)$ as the DISJ protocol has quantum communication cost $O(\sqrt{k} \log k)$. Then for each repeat of step 3, if the DISJ protocol gives wrong answer, we will not do anything. And if the DISJ protocol gives the correct intersection, the counter will be increased by one and the intersection place will be removed and we can find other intersections. Thus we only have to show with probability at least $1 - 1/k$, DISJ protocol gives a correct answer for at least $k^{0.02}$ times. Assume the DISJ protocol

---

1. Set a and b as we just described. Set $cnt = 0$.

2. Do the following step $c_1 \cdot k^{0.02}$ times, c_1 is some constant to be figured out in the proof:

3. Use protocol for DISJ in [BCW98] to find the intersection between a and b, let it be at place j, Alice and Bob communicate 2 bits to check if $a_j = b_j = 1$. If it is true, then set $cnt = cnt + 1$, $a_j = 0$, $b_j = 0$.

Protocol 2: Protocol count

succeeds with some constant probability p. Let Cr denote the random variable for the number of correct answers DISJ protocol gives. We know $\mathbb{E}[Cr] = p \cdot c_1 \cdot k^{0.02}$. By the additive Chernoff bound, the probability that DISJ protocol give a correct answer for at least $k^{0.02}$ times is

$$\Pr[Cr \geq k^{0.02}] = 1 - \Pr[Cr < k^{0.02}] \geq 1 - e^{-2(p \cdot c_1 \cdot k^{0.02} - k^{0.02})^2 / (c_1 \cdot k^{0.02})}.$$

By picking c_1 properly, for example $c_1 = 2/p$, we get $\Pr[Cr \geq k^{0.02}] \geq 1 - 1/k$. $\qquad\square$

Lemma 5.5. *Let $\epsilon > 10k^{-0.005}$ be some constant. After communicating $O\left(k^{0.99} \cdot n + 2^{2n} \cdot k^{0.52} \log k\right)$ bits, with probability at least $1/2$, Alice and Bob agree on some $\tilde{\mu}$, such that for any (x, y), $\frac{\tilde{\mu}(x,y)}{\mu(x,y)} < 1 + \epsilon$.*

Proof. We use the following protocol to estimate μ:

Inputs: $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_k)$

1. Sample the coordinates randomly $k^{0.99}$ times using public randomness (with replacement). Alice and Bob exchange their input for these coordinates. For each $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$, count the number of times it appears in these coordinates and denote the count by $c_1(x, y)$.

2. For all (x, y), use Lemma 5.4 to count the number of times (x, y) appears in the input and denote the count obtained by $c_2(x, y)$.

3. We combine c_1 and c_2 as c_3. For each (x, y), if $c_2(x, y) \geq k^{0.02}$, let $c_3(x, y) = c_1(x, y) \cdot k^{0.01}$ otherwise $c_3(x, y) = c_2(x, y)$.

4. $\tilde{\mu}(x, y) = \frac{c_3(x,y)}{\sum_{x',y'} c_3(x',y')}$.

Protocol 3: Estimate μ

Let's first analyze the communication cost of this part. It's clear that the first step needs at most $O\left(k^{0.99}n\right)$ communication. For second step, by Lemma 5.4, it needs at most $O\left(2^{2n} \cdot k^{0.52} \log k\right)$ communication. Sum them up, this protocol needs $O\left(k^{0.99} \cdot n + 2^{2n} \cdot k^{0.52} \log k\right)$ bits of communication.

Then let's consider the following events:

1. For all (x, y) such that $\mu(x, y) \cdot k \geq k^{0.02}$, $|c_1(x, y) \cdot k^{0.01} - \mu(x, y) \cdot k| < \frac{\epsilon}{3} \mu(x, y) \cdot k$.

2. For any (x, y), the protocol described in Lemma 5.4 does not fail.

If these two events happen, then we know that $|c_3(x,y) - \mu(x,y) \cdot k| < \frac{\epsilon}{3}\mu(x,y) \cdot k$, therefore as desired,

$$\tilde{\mu}(x,y) = \frac{c_3(x,y)}{\sum_{x',y'} c_3(x',y')} \le \frac{(1 + \frac{\epsilon}{3})\mu(x,y) \cdot k}{(1 - \frac{\epsilon}{3}) \cdot k} < (1 + \epsilon)\mu(x,y).$$

Finally, we only have to make sure that these two events happen with probability at least $1/2$. For the first event, by the multiplicative Chernoff bound and union bound, it does not happen with probability

$$2^{2n} \cdot \Pr[|c_3(x,y)/k^{0.99} - \mu(x,y)| > \frac{\epsilon}{3}\mu(x,y)] < 2ke^{-\frac{(\epsilon/3)^2\mu(x,y)k^{0.99}}{3}} \le 2ke^{-\epsilon^2 k^{0.01}/27} < 1/4.$$

For the second event, by Lemma 5.4 and the union bound, it does not happen with probability at most $2^{2n} \cdot \frac{1}{k} < 1/4$. Thus these two events happen with probability at least $1/2$. $\qquad\square$

Let's consider the communication cost of τ. For the first step, the cost is $O\left(k^{0.99} \cdot n + 2^{2n} \cdot k^{0.52}\log k\right) = o(k)$. For the second step, the quantum communication complexity is at most $c \cdot l$. For the third step, the cost is 0. Therefore $QCC(\tau) \le c \cdot l + o(k)$.

Let's say that the protocol τ succeeds when the following things happen:

1. For all (x,y), $\frac{\tilde{\mu}(x,y)}{\mu(x,y)} < 1 + \epsilon$.

2. The c runs of protocol $\pi_{\tilde{\mu}^k}$ in step 2 of protocol τ all compute at least $(1 - \alpha)k$ coordinates correctly.

3. Number of $i \in [k]$ such that the coordinate i is not sampled in step 2 of protocol τ is at most $2^{-c/2}k$.

If τ succeeds, then it computes at least $\left(1 - 2^{-c/2} - c\alpha\right)k$ coordinates correctly. This is because errors come from two possible ways:

1. Some coordinates are not sampled. When τ succeeds, the number of coordinates that are not sampled is at most $2^{-c/2}k$.

2. Some coordinates' results are wrong in step 2. When τ succeeds, the number of errors from step 2 is at most αck.

Finally, let's analyze the success probability of protocol τ. Let's analyze step by step:

1. For step one, by Lemma 5.5, it is clear that we succeed with probability $1/2$.

2. For step two, first we know that when running $\pi_{\tilde{\mu}^k}$ on distribution $\tilde{\mu}^k$, we succeed with probability at least γ. And since we have for any (x,y), $\frac{\tilde{\mu}(x,y)}{\mu(x,y)} < 1 + \epsilon$, if we run $\pi_{\tilde{\mu}^k}$ on distribution μ^k, the success probability will be at least $\frac{\gamma}{(1+\epsilon)^k}$. When running this protocol c times independently, the success probability will be at least $\left(\frac{\gamma}{(1+\epsilon)^k}\right)^c$. Note that when we sample coordinates independently at random, the distribution we induce is μ^k.

28

3. It is only left to analyze the probability that number of coordinates not sampled in step 2 of protocol τ is at least $2^{-c/2}k$. For each coordinate i, define s_i to be the random variable that indicates whether coordinate i is sampled or not (1 means not sampled and 0 means sampled). Then we have $\mathbb{E}[s_i] = \left(1 - \frac{1}{k}\right)^{ck} < 2^{-c}$. In order to show the failure probability small by Chernoff bound, we will show that all the s_i's are negatively correlated. To show they are negatively correlated, we only have to show

$$\forall I \subseteq [k], \Pr\left[\prod_{i \in I} s_i = 1\right] \leq \prod_{i \in I} \Pr[s_i = 1].$$

Notice that $\Pr\left[\prod_{i \in I} s_i = 1\right] = \left(1 - \frac{|I|}{k}\right)^{kc}$ and $\Pr[s_i = 1] = \left(1 - \frac{1}{k}\right)^{kc}$. So we have,

$$\forall I \subseteq [k], \Pr\left[\prod_{i \in I} s_i = 1\right] = \left(1 - \frac{|I|}{k}\right)^{kc} \leq \left(\left(1 - \frac{1}{k}\right)^{|I|}\right)^{kc} = \prod_{i \in I} \Pr[s_i = 1].$$

Since all the s_i's are negatively correlated, by Chernoff bound for negatively correlated random variables, for example see [DP], we have that the failure probability

$$\Pr\left[\sum_{i=1}^{k} s_i \geq 2^{-c/2}k\right] < e^{-2k\left(2^{-c/2} - 2^{-c}\right)^2} < e^{-2^{2-2c}k} < 2^{-2^{2-2c}k}.$$

The second inequality holds for all $c \geq 3$. Notice that the event that we err in the first step is independent from the event that we err in the second step. So the success probability of τ is at least $\frac{1}{2}\left(\left(\frac{\gamma}{(1+\epsilon)^k}\right)^c - 2^{-2^{2-2c}k}\right)$.

$\square$

5.3 Lower bound on QIC

Definition 5.6. We say that $QCC(f^k, \eta_1 k, \eta_2) \leq C$ if there exists a protocol π for f^k s.t. $QCC(\pi) \leq C$ and

$$Pr[\pi \text{ computes} \geq \eta_1 k \text{ coordinates correctly}] \geq 1 - \eta_2$$

Here the probability is over randomness of protocol (which includes the randomness due to quantum measurements). We don't require the protocol to declare which coordinates were computed correctly.

Theorem 5.7. *There exists an absolute constant $\eta > 0$ s.t. for any boolean function f, $QIC_D(f, \eta) \geq \Omega(GDM_{1/5}(f) - O(1))$.*

Proof. Let $\eta > 0$ be a sufficiently small constant to be fixed later. Suppose $\max_\mu QIC(f, \mu, \eta) = I$. We will show that for sufficiently large k, it holds that

$$QCC(f^k, (1 - \epsilon_{\text{sh}})k, 1 - 2^{-\epsilon_{\text{sh}}k}) \leq O(k \cdot (I + 2)) + o(k)$$

from which the theorem follows from Theorem 3.16.

29

By definition, for all μ, there exists a protocol Π_μ for f s.t. $QIC(\Pi_\mu, \mu) \leq I+1$ and error $\leq \eta$ w.r.t μ. By lemma 5.2, for sufficiently large k, there exists a protocol $\Pi_{k,\mu,\epsilon'}$ s.t. $QCC(\Pi_{k,\mu,\epsilon'}) \leq k(I+2)$ and

$$\Pr[\Pi_{k,\mu,\epsilon'} \text{ computes} \geq (1-2\eta)k \text{ coordinates of } f^k \text{ correctly}] \geq 1 - e^{-2\eta^2 k} - \epsilon'$$

Here the probability is over the distribution μ^k and the randomness of the protocol. Choose k large enough and ϵ' small enough so that $1 - e^{-2\eta^2 k} - \epsilon' \geq 0.9$. Then by lemma 5.3, for any integer $c > 0$, any constant $\epsilon > 0$, there exists a protocol τ s.t.

$$\Pr[\tau \text{ computes} \geq (1 - 2^{-c/2} - 2c\eta)k \text{ coordinates correctly (on any input } (x_1, \ldots, x_k, y_1, \ldots, y_k))]$$
$$\geq \frac{1}{2}\left(\left(\frac{0.9}{(1+\epsilon)^k}\right)^c - 2^{-2^{2-2c}k}\right)$$

Here the randomness is only over the randomness of the protocol. Also $QCC(\tau) \leq c \cdot k \cdot (I+2) + o(k)$. Choose $c = \lceil 2\log\left(\frac{2}{\epsilon_{\text{sh}}}\right)\rceil$. Also choose $\eta = \frac{\epsilon_{\text{sh}}}{4c}$. Then

$$1 - 2^{-c/2} - 2c\eta \geq 1 - \epsilon_{\text{sh}}$$

Since $2^{2x} \geq 1 + x$ for all $x > 0$, it follows that

$$\left(\frac{0.9}{(1+\epsilon)^k}\right)^c \geq 0.9^c \cdot 2^{-2 \cdot \epsilon \cdot c \cdot k} \geq 2^{-(2\epsilon k + 1) \cdot c} \geq 2^{-4 \cdot \epsilon \cdot c \cdot k}$$

The last inequality is true for sufficiently large k. Now choose $\epsilon = \epsilon_{\text{sh}}^4/100c$. Then since

$$2^{-2^{2-2c}k} \leq 2^{-\epsilon_{\text{sh}}^4 k/16}$$

we get that

$$\frac{1}{2}\left(\left(\frac{0.9}{(1+\epsilon)^k}\right)^c - 2^{-2^{2-2c}k}\right) \geq \frac{1}{2}\left(2^{-\epsilon_{\text{sh}}^4 k/25} - 2^{-\epsilon_{\text{sh}}^4 k/16}\right)$$
$$\geq 2^{-\epsilon_{\text{sh}}^4 k/16}$$
$$\geq 2^{-\epsilon_{\text{sh}} k}$$

The second inequality holds for sufficiently large k. Hence $QCC(\tau) \leq c \cdot k \cdot (I+2) + o(k)$ and

$$\Pr[\tau \text{ computes} \geq (1 - \epsilon_{\text{sh}})k \text{ coordinates correctly (on any input } (x_1, \ldots, x_k, y_1, \ldots, y_k))]$$
$$\geq 2^{-\epsilon_{\text{sh}} k}$$

which implies that $QCC(f^k, (1-\epsilon_{\text{sh}})k, 1 - 2^{-\epsilon_{\text{sh}} k}) \leq O(k \cdot (I+2)) + o(k)$. $\qquad \square$

Corollary 5.8. *For all boolean functions f, $QCC(f, 1/3) \leq 2^{O(QIC(f,1/3)+1)}$.*

Proof. We will use the following folklore result:

$$R(f, 1/3) \leq \left(\frac{1}{\text{disc}(f)}\right)^{O(1)}$$

where $R(f, 1/3)$ is the (public-coin) randomized communication complexity of f with error $1/3$ and $\text{disc}(f) = \min_{\mu} \text{disc}^{\mu}(f)$. See, for example, exercise 3.32 in [KN97]. This implies

$$QCC(f, 1/3) \leq R(f, 1/3) \leq \left(\frac{1}{\text{disc}(f)} \right)^{O(1)} \leq 2^{O(GDM_{1/5}(f))} \tag{29}$$

Now, by theorem 5.7 and theorem 4.13, we get that $QIC(f, \eta) \geq \Omega(GDM_{1/5}(f) - O(1))$ for some small constant η. By lemma 4.15, we also get that $QIC(f, 1/3) \geq \Omega(GDM_{1/5}(f) - O(1))$, which combined with equation (29) completes the proof. $\square$

6 From AND to Disj

In this section, we show that a protocol with low quantum information cost for AND implies a protocol with low quantum information cost for Disjointness

Lemma 6.1.

$$\max_{\nu} QIC(DISJ_n, \nu, 2/n) \leq n \cdot QIC_0^r(AND, 1/n^2) + O(r \cdot \log^5(n)) + o(\sqrt{n}) \tag{30}$$

Proof. Let $QIC_0^r(AND, 1/n^2) = I$. Suppose π is a protocol for AND which has error $\leq 1/n^2$ for all inputs and s.t. $\max_{\mu \text{ s.t. } \mu(1,1) = 0} QIC(\pi, \mu) \leq I + \delta$, for arbitrary small δ. Using π, we will construct a protocol for $DISJ_n$. The protocol will have low information cost w.r.t. any distribution ν. Suppose τ_k is a quantum protocol for $DISJ_k$ that has worst case error $\leq 1/k^{10}$ and communication cost $O(\sqrt{k} \log(k))$. For example, use the protocol from [AA03] and amplify the error to $1/k^{10}$. We'll drop the subscript k when it is clear from the context. Consider the protocol π_n described as Protocol 4.

Inputs: $(x, y) \in \{0, 1\}^n \times \{0, 1\}^n$, $(x, y) \sim \nu$
Goal: check if $DISJ_n(x, y) = 1$ or not.

1. Alice and Bob share a maximally entangled state $\phi_S^{S_A S_B}$ that will serve as shared randomness in order to sample uniformly at random $n/\log^3(n)$ coordinates from $[n]$ (with replacement). Alice has the register S_A and Bob has S_B.

2. On the random coordinates, run τ. Suppose O_A is the output register for Alice and O_B is the output register for Bob. Note that all this can be implemented using unitaries. Also note either $O_A = O_B = 1$ or $O_A = O_B = 0$.

3. If $O_A = O_B = 1$, then run π on each coordinate. If π outputs 1 on any coordinate, then output 0, otherwise output 1. If $O_A = O_B = 0$, Alice and Bob will keep running a dummy protocol (for example keep exchanging a freshly prepared register $|0\rangle$ of dimension same as to be sent in π^n in the corresponding step). In the end they output 0.

Protocol 4: Subsampling Protocol π_n

We'll denote the protocol in which π is run independently on each coordinate by π^n. First lets analyze the error of the protocol π_n. Suppose (x, y) were disjoint. Then probability that we output

0 because of τ is at most $\log^{30}(n)/n^{10} \leq 1/n$. And the probability that we output 0 because of π^n is at most $n/n^2 = 1/n$ because of union bound. So error in this case $\leq 2/n$. If the sets were intersecting, even if we don't output 0 because of τ, we will output 0 because of π^n w.p. at least $1 - 1/n^2$ (because on the intersecting coordinate, $1/n^2$ is the probability of failure). So in both cases, probability of error $\leq 2/n$.

Now lets figure out the information cost of π_n. For running τ, we just bound the information cost by communication cost, which is at most $\sqrt{n}/\sqrt{\log(n)} = o(\sqrt{n})$. The interesting part is what happens after τ. Lets look at the state of Alice and Bob after τ is over. Alice holds the registers A_τ, O_A, S_A, where A_τ is what is left behind with Alice after τ, O_A is Alice's output register for τ and S_A is the entanglement register which acts as shared randomness. Similarly Bob holds B_τ, O_B, S_B. After running i steps of π^n (just before the $(i+1)^{\text{th}}$ message is transmitted), Alice and Bob hold registers A_{i+1} and B_{i+1} respectively, with C_{i+1} (the register to be sent next) with Alice if i even and with Bob if i odd. Note that the number of rounds of π is r. Then the information cost of step 3 is:

$$\frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; R|B_{i+1}, B_\tau, O_B, S_B) + \frac{1}{2} \cdot \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; R|A_{i+1}, A_\tau, O_A, S_A)$$

$$\leq \frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; R, B_\tau, O_B, S_B|B_{i+1}) + \frac{1}{2} \cdot \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; R, A_\tau, O_A, S_A|A_{i+1})$$

$$\leq \frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; R, B_\tau, O_B, S_B, A_\tau, O_A, S_A|B_{i+1})+$$

$$\frac{1}{2} \cdot \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; R, B_\tau, O_B, S_B, A_\tau, O_A, S_A|A_{i+1})$$

$$= \frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; O_A|B_{i+1}) + \frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A|B_{i+1}, O_A)+$$

$$\frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; O_B|B_{i+1}, R, B_\tau, S_B, A_\tau, S_A, O_A)+$$

$$\frac{1}{2} \cdot \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; O_A|A_{i+1}) + \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A|A_{i+1}, O_A)+ \qquad (31)$$

$$\frac{1}{2} \cdot \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; O_B|A_{i+1}, R, B_\tau, S_B, A_\tau, S_A, O_A)$$

$$\leq \frac{1}{2} \cdot \sum_{i=0,i \ even}^{r-1} I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A|B_{i+1}, O_A)+$$

$$\frac{1}{2} \cdot \sum_{i=0,i \ odd}^{r-1} I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A|A_{i+1}, O_A) + O(r)$$

$$= \frac{1}{2} \cdot \sum_{i=0,i\ even}^{r-1} \Pr[O_A = 1] \cdot I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A | B_{i+1}, O_A = 1) +$$

$$\frac{1}{2} \cdot \sum_{i=0,i\ odd}^{r-1} \Pr[O_A = 1] \cdot I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A | A_{i+1}, O_A = 1) + O(r)$$

The first two inequalities are by properties of mutual information. The first equality is just chain rule. Third inequality follows from the fact that O_A, O_B are one dimensional systems. The last equality is true because O_B is just a copy of O_A, so tracing out O_B, O_A becomes a classical system and also conditioned on $O_A = 0$, the mutual information expressions are 0 since in that case the C_{i+1} registers are independent of everything else. Now lets analyze the term:

$$\frac{1}{2} \cdot \sum_{i=0,i\ even}^{r-1} I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A | B_{i+1}, O_A = 1) + \frac{1}{2} \cdot \sum_{i=0,i\ odd}^{r-1} I(C_{i+1}; R, B_\tau, S_B, A_\tau, S_A | A_{i+1}, O_A = 1)$$

We claim that this is equal to $QIC(\pi^n, \nu')$, where ν' is the distribution $\nu | O_A = 1$. This follows from the following observations:

- Since O_B is just a copy of O_A, for all i, the state of systems $A_{i+1}, B_{i+1}, C_{i+1}, R, B_\tau, S_B, A_\tau, S_A$ conditioned on $O_A = 1$ (the post-measurement state if O_A is measured and the result is 1) is pure.

- For all i, the marginal state of systems $A_{i+1}, B_{i+1}, C_{i+1}$ conditioned on $O_A = 1$ is the same as it would have been if π^n was run starting from the distribution ν'. This is because π^n never touches the registers B_τ, S_B, A_τ, S_A.

- If $|\phi\rangle^{R',A,B,C}$ and $|\psi\rangle^{R,A,B,C}$ are two pure states such that $\mathrm{Tr}_{R'} |\phi\rangle^{R',A,B,C} = \mathrm{Tr}_R |\psi\rangle^{R,A,B,C}$. Then $I(C; R'|B)_\phi = I(C; R|B)_\psi$.

Remark 6.2. The reader might have noticed that the trick of merging stuff with the purification register and then applying the last observation is used at a lot of places in this paper. This seems to be a very useful trick and seems to replace the classical Proposition 2.9 from [Bra12].

Putting it all together, we have the following upper bound on information cost of step 3:

$$\Pr[O_A = 1] \cdot QIC(\pi^n, \nu') + O(r)$$

$$\leq \Pr[O_A = 1] \cdot \left(\sum_{i=1}^{n} QIC(\pi, \nu'_i) \right) + O(r)$$

$$\leq \Pr[O_A = 1] \cdot n \cdot QIC\left(\pi, \sum_{i=1}^{n} \nu'_i/n \right) + O(r)$$

$$\leq \Pr[O_A = 1] \cdot n \cdot (I + \delta) + O(\Pr[O_A = 1] \cdot n \cdot rH(w)) + O(r) \qquad (32)$$

Here ν'_i is the marginal distribution on the ith coordinate and $w = \sum_{i=1}^{n} \nu'_i(1,1)/n$. First inequality is by lemma 4.14. Second inequality is just concavity of information cost, lemma 3.13. The last

inequality follows from corollary 4.9. Now we can assume that $\Pr[O_A = 1] \geq 1/n$, otherwise (32) is trivially bounded by $O(r)$. Now let us bound w. Suppose (X, Y) are random variables s.t. $(X, Y) \sim \nu$. Also let $N(x, y)$ be the number of intersections in x and y i.e. number of i such that $x_i = y_i = 1$. Then

$$
\begin{aligned}
\Pr[N(X, Y) = d | O_A = 1] &= \frac{\Pr[N(X, Y) = d] \cdot \Pr[O_A = 1 | N(X, Y) = d]}{\Pr[O_A = 1]} \\
&\leq \Pr[N(X, Y) = d] \cdot \Pr[O_A = 1 | N(X, Y) = d] \cdot n \\
&\leq \Pr[N(X, Y) = d] \cdot \left(\left(1 - \frac{d}{n}\right)^{n/\log^3(n)} + \frac{\log^{30}(n)}{n^{10}} \right) \cdot n \\
&\leq e^{-d/\log^3(n)} \cdot n + \frac{\log^{30}(n)}{n^9}
\end{aligned}
$$

The second inequality follows because if there are d intersections, then getting no intersection in $n/\log^3(n)$ uniformly random coordinates is at most the first term. The second term is due to the error of the amplified protocol for disjointness. So for $d \geq 9\ln(2)\log^4(n)$, $\Pr[N(X, Y) = d | O_A = 1] \leq 1/n^8$. Thus

$$
w = \sum_{i=1}^{n} \nu_i'(1, 1)/n = \mathbb{E}_{(X,Y) \sim \nu'} N(X, Y)/n \leq O(\log^4(n)/n)
$$

Thus we can bound (32) as follows:

$$
\begin{aligned}
&\Pr[O_A = 1] \cdot n \cdot (I + \delta) + O(\Pr[O_A = 1] \cdot n \cdot rH(w)) + O(r) \\
&\leq n \cdot (I + \delta) + O(n \cdot rH(w)) + O(r) \\
&\leq n \cdot (I + \delta) + O(r \log^5(n))
\end{aligned}
$$

Since δ was arbitrary small, this completes the proof. $\qquad\square$

7 Proof of the main result

We now put everything together to get a lower bound on $QIC_0^r(AND, 1/3)$.

Lemma 7.1. *For all r, it holds that*

$$
QIC_0^r(AND, 1/3) \geq \Omega\left(\frac{1}{r \cdot \log^8 r}\right).
$$

Proof. We know by theorem 3.17 that $GDM_{1/5}(DISJ_n) \geq \Omega(\sqrt{n})$. Hence, by Theorem 5.7, we must have that $\max_\mu QIC(DISJ_n, \mu, 2/n) \geq \Omega(\sqrt{n})$. Putting this together with Lemma 6.1 and Corollary 4.17, and let $r = \Theta\left(\frac{\sqrt{n}}{\log^6 n}\right)$, we have,

$$
QIC_0^r(AND, 1/3) = \Omega\left(\frac{1}{\sqrt{n} \cdot \log^2 n}\right) = \Omega\left(\frac{1}{r \cdot \log^8 r}\right).
$$

$\qquad\square$

Corollary 7.2. *Let μ^* be the distribution such that $\mu^*(0,0) = 1/3, \mu^*(0,1) = 1/3, \mu^*(1,0) = 1/3$. Then*

$$\inf_{\Pi \in \mathcal{T}^r(AND, 1/3)} QIC(\Pi, \mu^*) = \Omega \left(\frac{1}{r \cdot \log^8 r} \right).$$

Proof. For any distribution μ_0 such that $\mu_0(1,1) = 0$, it is easy to see that μ^* can be written as $\mu^* = \frac{1}{3}\mu_0 + \frac{2}{3}\mu'$ where μ' is some other valid distribution. By Lemma 3.13, we have

$$QIC(\Pi, \mu^*) \geq \frac{1}{3}QIC(\Pi, \mu_0) + \frac{2}{3}QIC(\Pi, \mu') \geq \frac{1}{3}QIC(\Pi, \mu_0).$$

Then we have

$$QIC(\Pi, \mu^*) \geq \frac{1}{3} \max_{\mu_0, \mu_0(1,1)=0} QIC(\Pi, \mu_0).$$

Therefore by Lemma 7.1, we have

$$\inf_{\Pi \in \mathcal{T}^r(AND, 1/3)} QIC(\Pi, \mu^*) \geq \frac{1}{3}QIC_0^r(AND, 1/3) = \Omega \left(\frac{1}{r \cdot \log^8 r} \right).$$

$\square$

Theorem 7.3. *For all $r, n \in \mathbb{N}$, $QCC^r(DISJ_n, 1/3) = \Omega\big(\frac{n}{r \cdot \log^8 r}\big)$.*

Proof. Combining Lemma 4.20 and Lemma 7.1, we get this theorem. $\square$

8 Low information protocol for AND

In this section, we exhibit a $\tilde{O}(1/r)$ information $4r$-round protocol for AND (w.r.t. the prior $1/3, 1/3, 1/3, 0$) which computes correctly on all inputs with probability 1. The protocol is due to Jain, Radhakrishnan and Sen. Consider the protocol described in Protocol 5.

First let us see why it computes AND. Let $|\psi_i^{x,y}\rangle = \cos(\phi_i^{x,y})|0\rangle + \sin(\phi_i^{x,y})|1\rangle$ be the state of qubit C after i rounds when the input is (x, y). If the input is $0, 0$, $\phi_i^{0,0}$ is always 0. Also when the input is $0, 1$, $\phi_i^{0,1}$ is always 0. So $\left|\psi_i^{0,0}\right\rangle = \left|\psi_i^{0,1}\right\rangle = |0\rangle$ always. When the input is $1, 0$, $\phi_i^{1,0}$ follows the trajectory $2\theta \to 2\theta \to 0 \to 0 \to 2\theta \to \cdots$. So $\left|\psi_{4r-1}^{1,0}\right\rangle = |0\rangle$ as well. When the input is $1, 1$, $\phi_i^{1,1}$ follows the trajectory $2\theta \to -2\theta \to 4\theta \to -4\theta \to \cdots \to -\pi/2$. So $\left|\psi_{4r-1}^{1,1}\right\rangle = -|1\rangle$. Thus the players compute AND correctly.

Now let us analyze the information cost of this protocol. Note that after i rounds the full state can be written as follows:

$$|\psi_i\rangle^{XYCR} = \sum_{x, y \text{ s.t. } x \wedge y = 0} \frac{1}{\sqrt{3}} |x\rangle^X |y\rangle^Y |\psi_i^{x,y}\rangle^C |x, y\rangle^R$$

Then information cost is given by:

$$\frac{1}{2} \cdot \sum_{i=1, odd}^{4r-1} I(C; R|Y)_{\psi_i} + \frac{1}{2} \cdot \sum_{i=1, even}^{4r-1} I(C; R|X)_{\psi_i}$$

Protocol 5: Protocol for AND

Let us look at a particular term:

$$
\begin{aligned}
I(C; R|Y)_{\psi_i} &= H(C, Y)_{\psi_i} + H(R, Y)_{\psi_i} - H(C, R, Y)_{\psi_i} - H(Y)_{\psi_i} \\
&= H(C, Y)_{\psi_i} + H(C, X)_{\psi_i} - H(X)_{\psi_i} - H(Y)_{\psi_i} \\
&= H(C|Y)_{\psi_i} + H(C|X)_{\psi_i} \\
&= \frac{2}{3}H(C|Y = 0)_{\psi_i} + \frac{1}{3}H(C|Y = 1)_{\psi_i} + \frac{2}{3}H(C|X = 0)_{\psi_i} + \frac{1}{3}H(C|X = 1)_{\psi_i} \\
&= \frac{2}{3}H(C|Y = 0)_{\psi_i}
\end{aligned}
$$

First equality is by definition. For second equality, we are using the fact that for a pure state on some systems A, B, $H(A) = H(B)$. Third equality is again by definition. For fourth equality, we use the fact that if we trace out R, X, Y become classical. For the fifth equality, we use the fact that conditioned on $Y = 1$, system C is in a pure state, namely $\left|\psi_i^{0,1}\right\rangle$. Similarly conditioned on $X = 1$, it is in state $\left|\psi_i^{1,0}\right\rangle$. Conditioned on $X = 0$, C is in the state $|0\rangle$. Now conditioned on $Y = 0$, C is in the state:

$$
\frac{1}{2}\left|\psi_i^{0,0}\right\rangle\left\langle\psi_i^{0,0}\right| + \frac{1}{2}\left|\psi_i^{1,0}\right\rangle\left\langle\psi_i^{1,0}\right|
$$

This is $|0\rangle$ if $i \equiv 3 \pmod 4$ and if $i \equiv 1 \pmod 4$, the density matrix is given by:

$$
\rho = \begin{bmatrix} \frac{1}{2} + \frac{1}{2}\cos^2(2\theta) & \frac{1}{2}\cos(2\theta)\sin(2\theta) \\ \frac{1}{2}\cos(2\theta)\sin(2\theta) & \frac{1}{2}\sin^2(2\theta) \end{bmatrix}
$$

Eigenvalue computation shows that $H(\rho) = H(\sin^2(\theta)) = O(\theta^2 \log(1/\theta)) = O(\log(r)/r^2)$. So some of Alice's terms are 0 and some are $O(\log(r)/r^2)$. Similarly some of Bob's terms are 0 and some

36

are $O(\log(r)/r^2)$. So in total we get that the information cost is $O(\log(r)/r)$. Note that from the protocol it might seem that since the roles of Alice and Bob are asymmetric, only Alice is sending information and Bob is not. However this definition of quantum information cost also accounts for sending back information in some sense. For example, in some of the rounds, Alice is sending Bob some information but Bob is sending it back, so that is accounted for. This results in Bob's part of the cost to be non-zero and in fact equal to that of Alice.

Now let us see what happens if we place a small mass w on $(1,1)$ entry. Then the full state can be described as follows:

$$|\psi_i\rangle^{XYCR} = \sum_{x,y \text{ s.t. } x \wedge y = 0} \sqrt{\frac{1-w}{3}} |x\rangle^X |y\rangle^Y |\psi_i^{x,y}\rangle^C |x,y\rangle^R + \sqrt{w} |1\rangle^X |1\rangle^Y \left|\psi_i^{1,1}\right\rangle^C |1,1\rangle^R$$

The ith term of the information cost as before is given by:

$$\frac{2(1-w)}{3} H(C|Y=0)_{\psi_i} + \frac{1+2w}{3} H(C|Y=1)_{\psi_i} + \frac{2(1-w)}{3} H(C|X=0)_{\psi_i} + \frac{1+2w}{3} H(C|X=1)_{\psi_i}$$

$$= \frac{2(1-w)}{3} H(C|Y=0)_{\psi_i} + \frac{1+2w}{3} H(C|Y=1)_{\psi_i} + \frac{1+2w}{3} H(C|X=1)_{\psi_i}$$

As before $H(C|X=0)_{\psi_i} = 0$. But the other three terms are non-zero. $H(C|Y=0)_{\psi_i}$ is the same as before. Let us focus on $H(C|Y=1)_{\psi_i}$. State of C conditioned on $Y=1$ is given by:

$$\frac{1-w}{1+2w} \left|\psi_i^{0,1}\right\rangle \left\langle\psi_i^{0,1}\right| + \frac{3w}{1+2w} \left|\psi_i^{1,1}\right\rangle \left\langle\psi_i^{1,1}\right|$$

For i odd, the density matrix is given by:

$$\rho = \begin{bmatrix} \frac{1-w}{1+2w} + \frac{3w}{1+2w}\cos^2((i+1)\theta) & \frac{3w}{1+2w}\cos((i+1)\theta)\sin((i+1)\theta) \\ \frac{3w}{1+2w}\cos((i+1)\theta)\sin((i+1)\theta) & \frac{3w}{1+2w}\sin^2((i+1)\theta) \end{bmatrix}$$

Eigenvalue computation shows that

$$H(\rho) = H\left(\frac{1 - \sqrt{1 - \frac{12w(1-w)\sin^2((i+1)\theta)}{(1+2w)^2}}}{2}\right)$$

Now assuming $w \leq 1/6$ and considering i such that $\sin^2((i+1)\theta) \geq 4/5$, we get that

$$\frac{1 - \sqrt{1 - \frac{12w(1-w)\sin^2((i+1)\theta)}{(1+2w)^2}}}{2} \geq \frac{1 - \sqrt{1 - \frac{8w}{(1+2w)^2}}}{2}$$

$$= \frac{1 - \frac{1-2w}{1+2w}}{2}$$

$$= \frac{2w}{1+2w}$$

Since other terms involving w either have positive contribution or are of lower order, we get that for a constant fraction of the rounds, the information cost term increases by an additive $\Omega(H(w))$. And hence overall the increase in information cost is at least $\Omega(rH(w))$.

References

[AA03]   Scott Aaronson and Andris Ambainis. Quantum search of spatial regions. In *Foundations of Computer Science, 2003. Proceedings. 44th Annual IEEE Symposium on*, pages 200–209. IEEE, 2003.

[BBCR10]   Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

[BBHT98]   Michel Boyer, Gilles Brassard, Peter Hoyer, and Alain Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46:493–506, 1998.

[BCK14]   Joshua Brody, Amit Chakrabarti, and Ranganath Kondapally. Certifying equality with limited interaction. *APPROX-RANDOM*, pages 545–581, 2014.

[BCW98]   Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *STOC*, 1998.

[BGPW13a] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. *STOC*, 2013.

[BGPW13b] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. Information lower bounds via self-reducibility. In *Computer Science–Theory and Applications*, pages 183–194. Springer, 2013.

[BHOS14]   Fernando G.S.L. Brandao, Aram W. Harrow, Jonathan Oppenheim, and Sergii Strelchuk. Quantum conditional mutual information, reconstructed states, and state redistribution. *http://arxiv.org/abs/1411.4921*, 2014.

[BM13]   Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. *STOC*, 2013.

[BP13]   Gabor Braun and Sebastian Pokutta. Common information and unique disjointness. *FOCS*, 2013.

[Bra12]   Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012.

[BRWY13]   Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. *FOCS*, 2013.

[BT15]   Mario Berta and Marco Tomamichel. The fidelity of recovery is multiplicative. *http://arxiv.org/abs/1502.07973*, 2015.

[BW92]   Charles H. Bennett and Stephen Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Physical Review Letters*, 69, 1992.

[BYJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[CA08]   Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(002), 2008.

[CvDNT98] Richard Cleve, Wim van Dam, Michael Nielsen, and Alain Tapp. Quantum entanglement and the communication complexity of the inner product function. *Lecture Notes in Computer Science*, 1509:61–74, 1998.

[DP]   Devdatt P. Dubhashi and Alessandro Panconesi. Concentration of measure for the analysis of randomised algorithms.

[EV93]   Avshalom C Elitzur and Lev Vaidman. Quantum mechanical interaction-free measurements. *Foundations of Physics*, 23(7):987–997, 1993.

[FR14]   Omar Fawzi and Renato Renner. Quantum conditional mutual information and approximate markov chains. *http://arxiv.org/abs/1410.0664*, 2014.

[GKR14]   Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 176–185. IEEE, 2014.

[Gro96]   Lov Kumar Grover. A fast quantum mechanical algorithm for database search. *STOC*, 1996.

[Hol73]   Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973.

[JRS03]   Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for bounded round quantum communication complexity of set disjointness. *FOCS*, pages 220 – 229, 2003.

[JSWZ13]   Rahul Jain, Yaoyun Shi, Zhaohui Wei, and Shengyu Zhang. Efficient protocols for generating bipartite classical distributions and quantum states. *IEEE Transactions of Information Theory*, 59(8):5171–5178, 2013.

[Kla98]   Hartmut Klauck. Lower bounds for computation with limited nondeterminism. *Computational Complexity*, pages 141–152, 1998.

[KLL$^+$12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 500–509. IEEE, 2012.

[KN97]   Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997.

[KNTSZ01] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman. Interaction in quantum communication and the complexity of set disjointness. *STOC*, 2001.

[KR11]   Bo'az Klartag and Oded Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *STOC*, pages 31–40, 2011.

[KS92]   Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, November 1992.

[KSDW04]   H. Klauck, R. Spalek, and R. De Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Foundations of Computer Science, 2004. Proceedings. 45th Annual IEEE Symposium on*, pages 12–21. IEEE, 2004.

[KWHZ95]   Paul Kwiat, Harald Weinfurter, Thomas Herzog, and Anton Zeilinger. Interaction-free measurement. *Physical Review Letters*, 74(24), 1995.

[LRS15]   James R. Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. *STOC*, 2015.

[NW93]   Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM Journal on Computing*, 22(211-219), 1993.

[PRV01]   Stephen J. Ponzio, Jaikumar Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62:323–355, 2001.

[Raz92]   Alexander Razborov. On the distributed complexity of disjointness. *TCS: Theoretical Computer Science*, 106, 1992.

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

[Raz02]   Alexander A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Science, Mathematics*, 67, 2002.

[She07]   Alexander A. Sherstov. The pattern matrix method for lower bounds on quantum communication. *STOC*, 2007.

[She12]   Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM Journal on Computing*, 41(5):1122–1165, 2012.

[She14]   Alexander A. Sherstov. Communication lower bounds using directional derivatives. *Journal of the ACM*, 61(6):1–71, 2014.

[ST13]   Mert Saglam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. *FOCS*, 2013.

[Ter72]   Frode Terkelsen. Some minimax theorems. *Mathematica Scandinavica*, 31:405–413, 1972.

[Tou15]   Dave Touchette. A new, fully quantum notion of information complexity, and an application to direct sum for bounded round quantum communication complexity. *STOC*, 2015.

[Wat13]   John Watrous. Theory of quantum information. *Lecture notes*, 2013.

[Wil13]   Mark Wilde. Quantum information theory. *Cambridge University Press*, June 10, 2013.

[Yao79]   Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *STOC*, pages 209–213, 1979.

[Yao93]   Andrew Chi-Chih Yao. Quantum circuit complexity. *FOCS*, 1993.

[Zal99]   Christof Zalka. Grover's quantum searching algorithm is optimal. *Physical Review A*, 60(4), 1999.