A direct sum theorem in communication complexity via message compression

Rahul Jain* Jaikumar Radhakrishnan† Pranab Sen‡

Abstract

We prove lower bounds for the *direct sum* problem for two-party bounded error randomised multiple-round communication protocols. Our proofs use the notion of *information cost* of a protocol, as defined by Chakrabarti et al. [CSWY01] and refined further by Bar-Yossef et al. [BJKS02]. Our main technical result is a 'compression' theorem saying that, for any probability distribution μ over the inputs, a k-round private coin bounded error protocol for a function f with information cost c can be converted into a k-round deterministic protocol for f with bounded distributional error and communication cost $O(kc)$. We prove this result using a *substate* theorem about *relative entropy* and a *rejection sampling* argument. Our direct sum result follows from this 'compression' result via elementary information theoretic arguments.

We also consider the direct sum problem in quantum communication. Using a probabilistic argument, we show that messages cannot be compressed in this manner even if they carry small information. Hence, new techniques may be necessary to tackle the direct sum problem in quantum communication.

1 Introduction

We consider the two-party *communication complexity* of computing a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$. There are two players Alice and Bob. Alice is given an input $x \in \mathcal{X}$ and Bob is given an input $y \in \mathcal{Y}$. They then exchange messages in order to determine $f(x, y)$. The goal is to devise a protocol that minimises the amount of communication. In the *randomised* communication complexity model, Alice and Bob are allowed to toss coins and base their actions on the outcome of these coin tosses, and are required to determine the correct value with high probability for every input. There are two models for randomised protocols: in the *private coin* model the coin tosses are private to each player; in the *public coin* model the two players share a string that is generated randomly (independently of the input). A protocol where k messages are exchanged between the two players is called a k-round protocol. One also considers protocols where the two parties send a message each to a referee who determines the answer: this is the *simultaneous* message model.

The starting point of our work is a recent result of Chakrabarti, Shi, Wirth and Yao [CSWY01] concerning the *direct sum* problem in communication complexity. For a function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$, the m-fold *direct sum* is the function $f^m : \mathcal{X}^m \times \mathcal{Y}^m \rightarrow \mathcal{Z}^m$, defined by $f^m(\langle x_1, \ldots, x_m \rangle, \langle y_1, \ldots, y_m \rangle) \triangleq \langle f(x_1, y_1), \ldots, f(x_m, y_m) \rangle$. One then studies the communication complexity of f^m as the parameter m increases. Chakrabarti et al. [CSWY01] considered the direct sum problem in the bounded error simultaneous

*School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India. Email: rahulj@tcs.tifr.res.in.

†School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai 400005, India. Email: jaikumar@tcs.tifr.res.in. Part of this work was done while visiting MSRI, Berkeley.

‡Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario N2L 3C1, Canada. Email: p2sen@cacr.math.uwaterloo.ca. This work was done while visiting TIFR, Mumbai and MSRI, Berkeley.

message private coin model and showed that for the equality function $\mathrm{EQ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, the communication complexity of EQ_n^m is $\Omega(m)$ times the communication complexity of EQ_n. In fact, their result is more general. Let $R^{\mathrm{sim}}(f)$ be the bounded error simultaneous message private coin communication complexity of $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, and let $\tilde{R}^{\mathrm{sim}}(f) \stackrel{\Delta}{=} \min_S R^{\mathrm{sim}}(f|_{S\times S})$, where S ranges over all subsets of $\{0,1\}^n$ of size at least $(\frac{2}{3})2^n$.

Theorem ([CSWY01]) $R^{\mathrm{sim}}(f^m) = \Omega(m(\tilde{R}^{\mathrm{sim}}(f) - O(\log n)))$. A similar result holds for two-party bounded error one-round protocols too.

The proof of this result in [CSWY01] had two parts. The first part used the notion of information cost of randomised protocols, which is the mutual information between the inputs (which were chosen with uniform distribution in [CSWY01]) and the transcript of the communication between the two parties. Clearly, the information cost is bounded by the length of the transcript. So, showing lower bounds on the information cost gives a lower bound on the communication complexity. Chakrabarti et al. showed that the information cost is super-additive, that is, the information cost of f^m is at least m times the information cost of f. The second part of their argument showed an interesting message compression result for communication protocols. This result can be stated informally as follows: if the message contains at most a bits of information about a player's input, then one can modify the (one-round or simultaneous message) protocol so that the length of the message is $O(a + \log n)$. Thus, one obtains a lower bound on the information cost of f if one has a suitable lower bound on the communication complexity f. By combining this with the first part, we see that the communication complexity of f^m is at least m times this lower bound on the communication complexity of f.

In this paper, we examine if this approach can be employed for protocols with more than one-round of communication. Let $R^k_\delta(f)$ denote the k-round private coin communication complexity of f where the protocol is allowed to err with probability at most δ on any input. Let μ be a probability distribution on the inputs of f. Let $C^k_{\mu,\delta}(f)$ denote the deterministic k-round communication complexity of f, where the protocol errs for at most δ fraction, according to the distribution μ, of the inputs. Let $C^k_{[],\delta}(f)$ denote the maximum, over all product distributions μ, of $C^k_{\mu,\delta}(f)$. We prove the following.

Theorem: Let m, k be positive integers, and $\epsilon, \delta > 0$. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Then, $R^k_\delta(f^m) \geq m \cdot (\frac{\epsilon^2}{2k} \cdot C^k_{[],\delta+2\epsilon}(f) - 2)$.

The proof this result, like the proof in [CSWY01], has two parts, where the first part uses a notion of information cost for k-round protocols, and the second shows how messages can be compressed in protocols with low information cost. We now informally describe the ideas behind these results. To keep our presentation simple, we will assume that Alice's and Bob's inputs are chosen uniformly at random from their input sets.

The first part of our argument uses the extension of the notion of information cost to k-round protocols. The information cost of a k-round randomised protocol is the mutual information between the inputs and the transcript. This natural extension, and its refinement to *conditional information cost* by [BJKS02] has proved fruitful in several other contexts [BJKS02, JRS03]. It is easy to see that it is bounded above by the length of the transcript, and a lower bound on the information cost of protocols gives a lower bound on the randomised communication complexity. The first part of the argument in [CSWY01] is still applicable: the information cost is super-additive; in particular, the k-round information cost of f^m is at least m times the k-round information cost of f.

The main contribution of this work is in the second part of the argument. This part of Chakrabarti et al. [CSWY01] used a technical argument to compress messages by exploiting the fact that they carry low information. Our proof is based on the connection between mutual information of random variables and the relative entropy of probability distributions (see Section 2 for definition). Intuitively, it is reasonable to expect that if the message sent by Alice contains little information about her input X, then for various values x of X, the conditional distribution on the message, denoted by P_x, are similar. In fact, if we use relative entropy to compare distributions, then one can show that the mutual information is the average taken over x of the relative entropy $S(P_x \| Q)$ of P_x and Q, where $Q = \mathrm{E}_X[P_X]$. Thus, if the information between Alice's input and her message is bounded by a, then typically $S(P_x \| Q)$ is about a. To exploit this fact, we use the Substate theorem of [JRS02] which states (roughly) that if $S(P_x \| Q) \leq a$, then $P_x \leq 2^{-a} Q$. Using a standard *rejection sampling* idea we then show that Alice can restrict herself to a set of just $2^{O(a)} n$ messages; consequently, her messages can be encoded in $O(a + \log n)$ bits. In fact, such a compact set of messages can be obtained by sampling $2^{O(a)} n$ times from distribution Q.

We believe this connection between relative entropy and sampling is an important contribution of this work. Besides giving a more direct proof of the second part of Chakrabarti et al.'s [CSWY01] argument, our approach quickly generalises to two party bounded error private coin multiple round protocols, and allows us to prove a message compression result and a direct sum lower bound for such protocols. Direct sum lower bounds for such protocols were not known earlier. In addition, our message compression result and direct sum lower bound for multiple round protocols hold for protocols computing relations too.

The second part of our argument raises an interesting question in the setting of quantum communication. Can we always make the length of quantum messages comparable to the amount of information they carry about the inputs without significantly changing the error probability of the protocol? That is, for $x \in \{0,1\}^n$, instead of distributions P_x we have density matrices ρ_x so that the expected quantum relative entropy $\mathrm{E}_X[S(\rho_x \| \rho)] \leq a$, where $\rho \stackrel{\Delta}{=} \mathrm{E}_X[\rho_x]$. Also, we are given measurements (POVM elements) M_y^x, $x, y \in \{0,1\}^n$. Then, we wish to replace ρ_x by ρ'_x so that there is a subspace of dimension $n \cdot 2^{O(a/\epsilon)}$ that contains the support of each ρ'_x; also, there is a set $A \subseteq \{0,1\}^n$, $|A| \geq \frac{2}{3} \cdot 2^n$ such that for each $(x, y) \in A \times \{0,1\}^n$, $|\mathrm{Tr}\, M_y^x \rho_x - \mathrm{Tr}\, M_y^x \rho'_x| \leq \epsilon$. Fortunately, the quantum analogue of the Substate theorem has already been proved by Jain, Radhakrishnan and Sen [JRS02]. Unfortunately, it is the rejection sampling argument that does not generalise to the quantum setting. Indeed, we can prove the following strong negative result about compressibility of quantum information: For sufficiently large constant a, there exist ρ_x, M_y^x, $x, y \in \{0,1\}^n$ as above such that any subspace containing the supports of ρ'_x as above has dimension at least $2^{n/6}$. This strong negative result seems to suggest that new techniques may be required to tackle the direct sum problem for quantum communication.

1.1 Previous results

The direct sum problem for communication complexity has been extensively studied in the past (see Kushilevitz and Nisan [KN97]). Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Let $C(f)$ ($R(f)$) denote the deterministic (bounded error private coin randomised) two-party communication complexity of f. Ceder, Kushilevitz, Naor and Nisan [FKNN95] showed that there exists a partial function f with $C(f) = \Theta(\log n)$, whereas solving m copies takes only $C(f^m) = O(m + \log m \cdot \log n)$. They also showed a lower bound $C(f^m) \geq m(\sqrt{C(f)/2} - \log n - O(1))$ for total functions f. For the one-round deterministic model, they showed that $C(f^m) \geq m(C(f) - \log n - O(1))$ even for partial functions. For the two-round deterministic model, Karchmer, Kushilevitz and Nisan [KKN92] showed that $C(f^m) \geq m(C(f) - O(\log n))$ for any relation f. Feder et al. [FKNN95] also showed that for the equality problem $R(EQ_n^m) = O(m + \log n)$.

1.2 Our results

We now state the new results in this paper.

Result 1 (Compression result, multiple-rounds) *Suppose that Π is a k-round private coin randomised protocol for $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Let the average error of Π under a probability distribution μ on the inputs $\mathcal{X} \times \mathcal{Y}$ be δ. Let X, Y denote the random variables corresponding to Alice's and Bob's inputs respectively. Let T denote the complete transcript of messages sent by Alice and Bob. Suppose $I(XY : T) \leq a$. Let $\epsilon > 0$. Then, there is another deterministic protocol Π' with the following properties:*

(a) The communication cost of Π' is at most $\frac{2k(a+1)}{\epsilon^2} + \frac{2k}{\epsilon}$ bits;

(b) The distributional error of Π' under μ is at most $\delta + 2\epsilon$.

Result 2 (Direct sum, multiple-rounds) *Let m, k be positive integers, and $\epsilon, \delta > 0$. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Then, $R^k_\delta(f^m) \geq m \cdot \left(\frac{\epsilon^2}{2k} \cdot C^k_{[],\delta+2\epsilon}(f) - 2\right)$.*

Result 3 (Quantum incompressibility) *Let m, n, d be positive integers and $k \geq 7$. Let $d \geq 160^2$, $1600 \cdot d^4 \cdot k2^k \ln(20d^2) < m$ and $3200 \cdot d^5 \cdot 2^{2k} \ln d < n$. Let the underlying Hilbert space be $\mathbb{C}^m$. There exist n states ρ_l and n orthogonal projections M_l, $1 \leq l \leq n$, such that*

(a) $\forall l \; \mathrm{Tr}\, M_l \rho_l = 1$.

(b) $\rho \stackrel{\Delta}{=} \frac{1}{n} \cdot \sum_l \rho_l = \frac{1}{m} \cdot I$*, where I is the identity operator on $\mathbb{C}^m$.*

(c) $\forall l \; S(\rho_l \| \rho) = k$.

(d) For all d-dimensional subspaces W of $\mathbb{C}^m$, for all ordered sets of density matrices $\{\sigma_l\}_{l \in [n]}$ with support in W, $|\{l : \mathrm{Tr}\, M_l \sigma_l \leq 1/10\}| \geq n/4$.

Remark: The above result intuitively says that the states ρ_l on $\log m$ qubits cannot be compressed to less than $\log d$ qubits with respect to the measurements M_l.

1.3 Organisation of the rest of the paper

Section 2 defines several basic concepts which will be required for the proofs of the main results. In Section 3, we prove a version of the message compression result for bounded error private coin simultaneous message protocols and state the direct sum result for such protocols. Our version is slightly stronger than the one in [CSWY01]. The main ideas of this work (i.e. the use of the Substate theorem and rejection sampling) are already encountered in this section. In Section 4, we prove the compression result for k-round bounded error private coin protocols, and state the direct sum result for such protocols. We prove the impossibility of quantum compression in Section 5. Finally, we conclude by mentioning some open problems in Section 6.

2 Preliminaries

2.1 Information theoretic background

In this paper, $\ln$ denotes the natural logarithm and $\log$ denotes logarithm to base 2. All random variables will have finite range. Let $[k] \stackrel{\Delta}{=} \{1, \ldots, k\}$. Let $P, Q : [k] \to \mathbb{R}$. The *total variation distance* (also known as

ℓ_1-*distance*) between P, Q is defined as $\|P - Q\|_1 \triangleq \sum_{i \in [k]} |P(i) - Q(i)|$. We say $P \leq Q$ iff $P(i) \leq Q(i)$ for all $i \in [k]$. Suppose X, Y, Z are random variables with some joint distribution. The *Shannon entropy* of X is defined as $H(X) \triangleq -\sum_x \Pr[X = x] \log \Pr[X = x]$. The mutual information of X and Y is defined as $I(X : Y) \triangleq H(X) + H(Y) - H(XY)$. For $z \in \mathsf{range}(Z)$, $I((X : Y) \mid Z = z)$ denotes the mutual information of X and Y conditioned on the event $Z = z$ i.e. the mutual information arising from the joint distribution of X, Y conditioned on $Z = z$. Define $I((X : Y) \mid Z) \triangleq \mathrm{E}_Z\, I((X : Y) \mid Z = z)$. It is readily seen that $I((X : Y) \mid Z) = H(XZ) + H(YZ) - H(XYZ) - H(Z)$. For a good introduction to information theory, see e.g. [CT91].

We now recall the definition of an important information theoretic quantity called *relative entropy*, also known as *Kullback-Leibler divergence*.

Definition 1 (Relative entropy) *Let P and Q be probability distributions on a set $[k]$. The relative entropy of P and Q is given by* $S(P\|Q) \triangleq \sum_{i \in [k]} P(i) \log \frac{P(i)}{Q(i)}$.

The following facts follow easily from the definitions.

Fact 1 *Let X, Y, Z, W be random variables with some joint distribution. Then,*

(a) $I(X : YZ) = I(X : Y) + I((X : Z) \mid Y)$;

(b) $I(XY : Z \mid W) \geq I(XY : Z) - H(W)$.

Fact 2 *Let (X, M) be a pair of random variables with some joint distribution. Let P be the (marginal) probability distribution of M, and for each $x \in \mathsf{range}(X)$, let P_x be the conditional distribution of M given $X = x$. Then $I(X : M) = \mathrm{E}_X[S(P_x\|P)]$, where the expectation is taken according to the marginal distribution of X.*

Thus, if $I(X : M)$ is small, then we can conclude that $S(P_x\|P)$ is small on the average.

Using Jensen's inequality, one can derive the following property of relative entropy.

Fact 3 (Monotonicity) *Let P and Q be probability distributions on the set $[k]$ and $\mathcal{E} \subseteq [k]$. Let $D_P = (P(\mathcal{E}), 1 - P(\mathcal{E}))$ and $D_Q = (Q(\mathcal{E}), 1 - Q(\mathcal{E}))$ be the two-point distributions determined by $\mathcal{E}$. Then, $S(D_P\|D_Q) \leq S(P\|Q)$.*

Our main information theoretic tool in this paper is the following theorem (see [JRS02]).

Fact 4 (Substate theorem) *Suppose P and Q are probability distributions on $[k]$ such that $S(P\|Q) = a$. Let $r \geq 1$. Then,*

(a) the set $\mathsf{Good} \triangleq \{i \in [k] : \frac{P(i)}{2^{r(a+1)}} \leq Q(i)\}$ *has probability at least* $1 - \frac{1}{r}$ *in P;*

(b) There is a distribution $\widetilde{P}$ on $[k]$ such that $\left\|P - \widetilde{P}\right\|_1 \leq \frac{2}{r}$ *and* $\alpha \widetilde{P} \leq Q$*, where* $\alpha \triangleq \left(\frac{r-1}{r}\right) 2^{-r(a+1)}$.

Proof: Let $\mathsf{Bad} \triangleq [k] - \mathsf{Good}$. Consider the two-point distributions $D_P = (P(\mathsf{Good}), 1 - P(\mathsf{Good}))$ and $D_Q = (Q(\mathsf{Good}), 1 - Q(\mathsf{Good}))$. By Fact 3, $S(D_P\|D_Q) \leq a$, that is,

$$P(\mathsf{Good}) \log \frac{P(\mathsf{Good})}{Q(\mathsf{Good})} + P(\mathsf{Bad}) \log \frac{P(\mathsf{Bad})}{Q(\mathsf{Bad})} \leq a.$$

From our definition, $P(\mathsf{Bad})/Q(\mathsf{Bad}) > 2^{(a+1)}$. Now, $P(\mathsf{Good}) \log \frac{P(\mathsf{Good})}{Q(Good)} \geq P(\mathsf{Good}) \log P(\mathsf{Good}) > -1$ (because $x \log x \geq (-\log e)/e > -1$ for $0 \leq x \leq 1$). It follows that $P(\mathsf{Bad}) \leq \frac{1}{r}$, thus proving part (a). Let $\widetilde{P}(i) \stackrel{\Delta}{=} P(i)/P(\mathsf{Good})$ for $i \in \mathsf{Good}$ and $\widetilde{P}(i) = 0$ otherwise. Then, $\widetilde{P}$ satisfies the requirements for part (b). ■

2.2 Chernoff-Hoeffding bounds

We will need the following standard Chernoff-Hoeffding bounds on tails of probability distributions of sequences of bounded, independent, identically distributed random variables. Below, the notation $B(t, q)$ stands for the binomial distribution got by t independent coin tosses of a binary coin with success probability q for each toss. A *randomised predicate* S on $[k]$ is a function $S : [k] \to [0, 1]$. For proofs of the following bounds, see e.g. [AS00, Corollary A.7, Theorem A.13].

Fact 5

(a) Let P be a probability distribution on $[k]$ and S a randomised predicate on $[k]$. Let $p \stackrel{\Delta}{=} \mathop{\mathrm{E}}_{x \in_P [k]}[S(x)]$.

Let $\mathbf{Y} \stackrel{\Delta}{=} \langle Y_1, \ldots, Y_r \rangle$ be a sequence of r independent random variables, each with distribution P. Then,

$$\Pr_{\mathbf{Y}}[|\mathop{\mathrm{E}}_{i \in_U [r]}[S(Y_i)] - p| > \epsilon] < 2\exp(-2\epsilon^2 r).$$

(b) Let R be a random variable with binomial distribution $B(t, q)$. Then,

$$\Pr[R < \frac{1}{2}tq] < \exp\left(-\frac{1}{8}tq\right).$$

2.3 Communication complexity background

In the two-party private coin randomised communication complexity model [Yao79], two players Alice and Bob are required to collaborate to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Alice is given $x \in \mathcal{X}$ and Bob is given $y \in \mathcal{Y}$. Let $\Pi(x, y)$ be the random variable denoting the entire transcript of the messages exchanged by Alice and Bob by following the protocol Π on input x and y. We say Π is a δ-error protocol if for all x and y, the answer determined by the players is correct with probability (taken over the coin tosses of Alice and Bob) at least $1 - \delta$. The communication cost of Π is the maximum length of $\Pi(x, y)$ over all x and y, and over all random choices of Alice and Bob. The k-round δ-error private coin randomised communication complexity of f, denoted $R^k_\delta(f)$, is the communication cost of the best private coin k-round δ-error protocol for f. When δ is omitted, we mean that $\delta = \frac{1}{3}$.

We also consider private coin randomised simultaneous protocols in this paper. $R^{\mathrm{sim}}_\delta(f)$ denotes the δ-error private coin randomised simultaneous communication complexity of f. When δ is omitted, we mean that $\delta = \frac{1}{3}$.

Let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. A deterministic protocol Π has distributional error δ if the probability of correctness of Π, averaged with respect to μ, is least $1 - \delta$. The k-round δ-error distributional communication complexity of f, denoted $C^k_{\mu,\delta}(f)$, is the communication cost of the best k-round deterministic protocol for f with distributional error δ. μ is said to be a product distribution if there exist probability distributions $\mu_{\mathcal{X}}$ on $\mathcal{X}$ and $\mu_{\mathcal{Y}}$ on $\mathcal{Y}$ such that $\mu(x, y) = \mu_{\mathcal{X}}(x) \cdot \mu_{\mathcal{Y}}(y)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. The k-round δ-error product distributional communication complexity of f is defined as

$C^k_{[],\delta}(f) = \sup_\mu C^k_{\mu,\delta}(f)$, where the supremum is taken over all product distributions μ on $\mathcal{X} \times \mathcal{Y}$. When δ is omitted, we mean that $\delta = \frac{1}{3}$.

We now recall the definition of the important notion of *information cost* of a communication protocol from Bar-Yossef et al. [BJKS02].

Definition 2 (Information cost) *Let Π be a private coin randomised protocol for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Let $\Pi(x,y)$ be the entire message transcript of the protocol on input (x,y). Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, and let the input random variable (X,Y) have distribution μ. The* information cost of Π under μ *is defined to be $I(XY : \Pi(X,Y))$. The k-round δ-error information complexity of f under the distribution μ, denoted by $\mathrm{IC}^k_{\mu,\delta}(f)$, is the infimum information cost under μ of a k-round δ-error protocol for f. $\mathrm{IC}^{\mathrm{sim}}_\delta(f)$ denotes the infimum information cost under the uniform probability distribution on the inputs of a private coin simultaneous δ-error protocol for f.*

Remark: In Chakrabarti et al. [CSWY01], the information cost of a private coin δ-error simultaneous message protocol Π is defined as follows: Let X (Y) denote the random variable corresponding to Alice's (Bob's) input, and let M (N) denote the random variable corresponding to Alice's (Bob's) message to the referee. The information cost of Π is defined as I(X:M) + I(Y:N). We note that our definition of information cost coincides with Chakrabarti et al.'s definition for simultaneous message protocols.

Let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. The probability distribution μ^m on $\mathcal{X}^m \times \mathcal{Y}^m$ is defined as $\mu^m(\langle x_1, \ldots, x_m\rangle, \langle y_1, \ldots, y_m\rangle) \stackrel{\Delta}{=} \mu(x_1,y_1) \cdot \mu(x_2,y_2) \cdots \mu(x_m,y_m)$. Suppose μ is a product probability distribution on $\mathcal{X} \times \mathcal{Y}$. It can be easily seen (see e.g. [BJKS02]) that for any positive integers m, k, and real $\delta > 0$, $IC^k_{\mu^m,\delta}(f^m) \geq m \cdot IC^k_{\mu,\delta}(f)$. The reason for requiring μ to be a product distribution is as follows. We define the notion of information cost for private coin protocols only. This is because the proof of our message compression theorem (Theorem 3), which makes use of information cost, works for private coin protocols only. If μ is not a product distribution, the protocol for f which arises out of the protocol for f^m in the proof of the above inequality fails to be a private coin protocol, even if the protocol for f^m was private coin to start with. To get over this restriction on μ, Bar-Yossef et al. [BJKS02] introduced the notion of *conditional information cost* of a protocol. Suppose the distribution μ is expressed as a convex combination $\mu = \sum_{d \in K} \kappa_d \mu_d$ of product distributions μ_d, where K is some finite index set. Let κ denote the probability distribution on K defined by the numbers κ_d. Define the random variable D to be distributed according to κ. Conditioned on D, μ is a product distribution on $\mathcal{X} \times \mathcal{Y}$. We will call μ a mixture of product distributions $\{\mu_d\}_{d \in K}$ and say that κ *partitions* μ. The probability distribution κ^m on K^m is defined as $\kappa^m(d_1, \ldots, d_m) \stackrel{\Delta}{=} \kappa(d_1) \cdot \kappa(d_2) \cdots \kappa(d_m)$. Then κ^m partitions μ^m in a natural way. The random variable D^m has distribution κ^m. Conditioned on D^m, μ^m is a product distribution on $\mathcal{X}^m \times \mathcal{Y}^m$.

Definition 3 (Conditional information cost) *Let Π be a private coin randomised protocol for a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Let $\Pi(x,y)$ be the entire message transcript of the protocol on input (x,y). Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, and let the input random variable (X,Y) have distribution μ. Let μ be a mixture of product distributions partitioned by κ. Let the random variable D be distributed according to κ. The* conditional information cost of Π under (μ, κ) *is defined to be $I((XY : \Pi(X,Y)) \mid D)$. The k-round δ-error conditional information complexity of f under (μ,κ), denoted by $\mathrm{IC}^k_{\mu,\delta}(f \mid \kappa)$, is the infimum conditional information cost under (μ,κ) of a k-round δ-error protocol for f.*

The following facts follow easily from the results in Bar-Yossef et al. [BJKS02] and Fact 1.

Fact 6 *Let μ be a probability distribution on $\mathcal{X} \times \mathcal{Y}$. Let κ partition μ. For any $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, positive integers m, k, real $\delta > 0$, $IC^k_{\mu^m,\delta}(f^m \mid \kappa^m) \geq m \cdot IC^k_{\mu,\delta}(f \mid \kappa) \geq m \cdot (IC^k_{\mu,\delta}(f) - H(\kappa))$.*

Fact 7 *With the notation and assumptions of Fact 6, $R^k_\delta(f) \geq IC^k_{\mu,\delta}(f \mid \kappa)$.*

2.4 Sampling uniformly random orthonormal sets of vectors

To prove our result about the incompressibility of quantum information, we need to define the notion of a uniformly random set of size d of orthonormal vectors from $\mathbb{C}^m$. Let $\mathbf{U}(m)$ denote the group (under matrix multiplication) of $m \times m$ complex unitary matrices. Being a compact topological group, it has a unique Haar probability measure on its Borel sets which is both left and right invariant under multiplication by unitary matrices (see e.g. [Chapter 14, Corollary 20][Roy88]). Let $\mathbf{U}_{m,d}$, ($1 \leq d \leq m$) denote the topological space of $m \times d$ complex matrices with orthonormal columns. $\mathbf{U}_{m,d}$ is compact, and the group $\mathbf{U}(m)$ acts on $\mathbf{U}_{m,d}$ via multiplication from the left. Let $f_{m,d} : \mathbf{U}(m) \to \mathbf{U}_{m,d}$ be the map got by discarding the last $m - d$ columns of a unitary matrix. $f_{m,d}$ induces a probability measure $\mu_{m,d}$ on the Borel sets of $\mathbf{U}_{m,d}$ from the Haar probability measure on $\mathbf{U}(m)$. $\mu_{m,d}$ is invariant under the action of $\mathbf{U}(m)$, and is in fact the unique $\mathbf{U}(m)$-invariant probability measure on the Borel sets of $\mathbf{U}_{m,d}$ (see e.g. [Chapter 14, Theorem 25][Roy88]). By a uniformly random ordered set $(v_1, \ldots, v_d)$, $1 \leq d \leq m$ of orthonormal vectors from $\mathbb{C}^m$, we mean an element of $\mathbf{U}_{m,d}$ chosen according to $\mu_{m,d}$. By a uniformly random d dimensional subspace V of $\mathbb{C}^m$, we mean a subspace $V \stackrel{\Delta}{=} \mathrm{Span}(v_1, \ldots, v_d)$, where $(v_1, \ldots, v_d)$ is a uniformly random ordered set of orthonormal vectors from $\mathbb{C}^m$.

Let $\mathbf{O}(m)$ denote the group (under matrix multiplication) of $m \times m$ real orthogonal matrices. Identify $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ by treating a complex number as a pair of real numbers. A uniformly random unit vector in $\mathbb{C}^m$ (i.e. a vector distributed according to $\mu_{m,1}$) is the same as a uniformly random unit vector in $\mathbb{R}^{2m}$, since $\mathbf{U}(m)$ is contained in $\mathbf{O}(2m)$. From now on, while considering metric and measure theoretic properties of $\mathbf{U}_{m,1}$, it may help to keep the above identification of $\mathbb{C}^n$ and $\mathbb{R}^{2m}$ in mind.

One way of generating a uniformly random unit vector in $\mathbb{R}^m$ is as follows: First choose $\langle y_1, \ldots, y_m \rangle$ independently, each y_i being chosen according to the one dimensional Gaussian distribution with mean 0 and variance 1 (i.e. a real valued random variable with probability density function $\frac{\exp(-y^2)}{\sqrt{2\pi}}$). Normalise to get the unit vector $\langle x_1, \ldots, x_m \rangle$, where $x_i \stackrel{\Delta}{=} \frac{y_i}{\sqrt{y_1^2 + \cdots + y_m^2}}$ (note that any $y_i = 0$ with zero probability). It is easily seen that the resulting distribution on unit vectors is $\mathbf{O}(m)$-invariant, and hence, the above process generates a uniformly random unit vector in $\mathbb{R}^m$.

From the above discussion, one can prove the following fact.

Fact 8

(a) Let $1 \leq d \leq m$. Let $(v_1, \ldots, v_d)$ be distributed according to $\mu_{m,d}$. Then for each i, v_i is distributed according to $\mu_{m,1}$, and for each i, j, $i \neq j$, (v_i, v_j) is distributed according to $\mu_{m,2}$,

(b) Suppose x, y are independent unit vectors, each distributed according to $\mu_{m,1}$. Let $w'' \stackrel{\Delta}{=} y - \langle x|y \rangle x$, and set $w \stackrel{\Delta}{=} x$ and $w' \stackrel{\Delta}{=} \frac{w''}{\|w''\|}$ (note that $w'' = 0$ with probability zero). Then the pair (w, w') is distributed according to $\mu_{m,2}$.

(c) Suppose x, y are independent unit vectors, each distributed according to $\mu_{m,1}$. Let V be a subspace of $\mathbb{C}^m$ and define $\widehat{x} \stackrel{\Delta}{=} \frac{Px}{\|Px\|}$, $\widehat{y} \stackrel{\Delta}{=} \frac{Py}{\|Py\|}$, where P is the orthogonal projection operator onto V (note that $Px = 0$, $Py = 0$ are each zero probability events). Then $\widehat{x}, \widehat{y}$ are uniformly random independent unit vectors in V.

We will need to 'discretise' the set of d-dimensional subspaces of $\mathbb{C}^m$. The discretisation is done by using a δ-*dense* subset of $\mathbf{U}_{m,1}$. A subset $\mathcal{N}$ of $\mathbf{U}_{m,1}$ is said to be δ-dense if each vector $v \in \mathbf{U}_{m,1}$ has some vector in $\mathcal{N}$ at distance no larger than δ from it. We require the following fact about δ-dense subsets of $\mathbf{U}_{m,1}$.

Fact 9 ([Mat02, Lemma 13.1.1, Chapter 13]) *For each $0 < \delta \leq 1$, there is a δ-dense subset $\mathcal{N}$ of $\mathbf{U}_{m,1}$ satisfying $|\mathcal{N}| \leq (4/\delta)^{2m}$.*

A mapping f between two metric spaces is said to be 1-*Lipschitz* if the distance between $f(x)$ and $f(y)$ is never larger than the distance between x and y. The following fact says that a 1-Lipschitz function $f : \mathbf{U}_{m,1} \to \mathbb{R}$ greatly exceeds its expectation with very low probability. It follows by combining Theorem 14.3.2 and Proposition 14.3.3 of [Mat02, Chapter 14].

Fact 10 *Let $f : \mathbf{U}_{m,1} \to \mathbb{R}$ be 1-Lipschitz. Then for all $0 \leq t \leq 1$, $\Pr[f > \mathrm{E}[f] + t + 12/\sqrt{2m}] \leq 2\exp(-t^2 m)$.*

2.5 Quantum information theoretic background

We consider a quantum system with Hilbert space $\mathbb{C}^m$. For A, B Hermitian operators on $\mathbb{C}^m$, $A \leq B$ is a shorthand for the statement "$B - A$ is positive semidefinite". A *POVM element* M over $\mathbb{C}^m$ is a Hermitian operator satisfying the property $0 \leq M \leq I$, where $0, I$ are the zero and identity operators respectively on $\mathbb{C}^m$. For a POVM element M over $\mathbb{C}^m$ and a subspace W of $\mathbb{C}^m$, define $M(W) \stackrel{\Delta}{=} \max_{w \in W : \|w\|=1} \langle w|M|w\rangle$. For subspaces W, W' of $\mathbb{C}^m$, define $\Delta(W, W') \stackrel{\Delta}{=} \max_M |M(W) - M(W')|$, where the maximum is taken over all POVM elements M over $\mathbb{C}^m$. $\Delta(W, W')$ is a measure of how well one can distinguish between subspaces W, W' via a measurement. For a good introduction to quantum information theory, see [NC00].

The following fact can be proved from the results in [AKN98].

Fact 11 *Let M be a POVM element over $\mathbb{C}^m$ and let $w, \widehat{w} \in \mathbb{C}^m$ be unit vectors. Then, $|\langle w|M|w\rangle - \langle \widehat{w}|M|\widehat{w}\rangle| \leq \|w - \widehat{w}\|$.*

A density matrix ρ over $\mathbb{C}^m$ is a Hermitian, positive semidefinite operator on $\mathbb{C}^m$ with unit trace. If A is a quantum system with Hilbert space $\mathbb{C}^m$ having density matrix ρ, then $S(A) \stackrel{\Delta}{=} S(\rho) \stackrel{\Delta}{=} -\mathrm{Tr}\, \rho \log \rho$ is the *von Neumann entropy* of A. If A, B are two disjoint quantum systems, the *mutual information* of A and B is defined as $I(A : B) \stackrel{\Delta}{=} S(A) + S(B) - S(AB)$. For density matrices ρ, σ over $\mathbb{C}^m$, their *relative entropy* is defined as $S(\rho\|\sigma) \stackrel{\Delta}{=} \mathrm{Tr}\, \rho(\log \rho - \log \sigma)$. Let X be a classical random variable with finite range and M be a m-dimensional quantum encoding of X i.e. for every $x \in \mathsf{range}(X)$ there is a density matrix σ_x over $\mathbb{C}^m$ (σ_x represents a 'quantum encoding' of x). Let $\sigma \stackrel{\Delta}{=} \mathrm{E}_X\, \sigma_x$, where the expectation is taken over the (marginal) probability distribution of X. Then, $I(X : M) = \mathrm{E}_X\, S(\sigma_x \|\sigma)$.

3 Simultaneous message protocols

In this section, we prove a result of [CSWY01], which states that if the mutual information between the message and the input is at most k, then the protocol can be modified so that the players send messages of length at most $O(k + \log n)$ bits. Our proof will make use of the Substate Theorem and a rejection sampling argument. In the next section, we will show how to extend this argument to multiple-round protocols.

Before we formally state the result and its proof, let us outline the main idea. Fix a simultaneous message protocol for computing the function $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \mathcal{Z}$. Let $X \in_U \{0,1\}^n$. Suppose $I(X : M) \leq a$, where M be the message sent by Alice to the referee when her input is X. Let $s_{xy}(m)$ be conditional probability that the referee computes $f(x,y)$ correctly when Alice's message is m, her input is x and Bob's input is y.

We want to show that we can choose a small subset $\mathcal{M}$ of possible messages, so that for most x, Alice can generate a message M'_x from this subset (according to some distribution that depends on x), and still ensure that $\mathrm{E}[s_{xy}(M'_x)]$ is close to 1, for all y. Let P_x be the distribution of M conditioned on the event $X = x$. For a fixed x, it is possible to argue that we can confine Alice's messages to a certain small subset $\mathcal{M}_x \subseteq [k]$. Let $\mathcal{M}_x$ consist of $O(n)$ messages picked according to the distribution P_x. Then, instead of sending messages according to the distribution P_x, Alice can send a random message chosen from $\mathcal{M}_x$. Using Chernoff-Hoeffding bounds one can easily verify that $\mathcal{M}_x$ will serve our purposes with exponentially high probability.

However, what we really require is a set of samples $\{\mathcal{M}_x\}$ whose union is small, so that she and the referee can settle on a common succinct encoding for the messages. Why should such samples exist? Since $I(X : M)$ is small, we have by Fact 2 that for most x, the relative entropy $S(P_x \| Q)$ is bounded (here Q is the distribution of the message M, i.e., $Q = \mathrm{E}_X[P_X]$). By combining this fact, the Substate Theorem (Fact 4) and a *rejection sampling* argument (see e.g. [Ros97, Chapter 4, Section 4.4]), one can show that if we choose a sample of messages according to the distribution Q, then, for most x, roughly one in every $2^{O(a)}$ messages 'can serve' as a message sampled according to the distribution P_x. Thus, if we pick a sample of size $n \cdot 2^{O(a)}$ according to Q, then for most x we can get a the required sub-sample $\mathcal{M}_x$. of $O(n)$ elements. The formal arguments are presented below.

The following easy lemma is the basis of the rejection sampling argument.

Lemma 1 (Rejection sampling) *Let P and Q be probability distributions on $[k]$ such that $2^{-a}P \leq Q$. Then, there exist correlated random variables X and χ taking values in $[k] \times \{0,1\}$, such that: (a) X has distribution Q, (b) $\Pr[\chi = 1] = 2^{-a}$ and (c) $\Pr[X = i \mid \chi = 1] = P(i)$.*

Proof: Since the distribution of X is required to be Q, we will just describe the conditional distribution of χ for each potential value i for X: let $\Pr[\chi = 1 \mid X = i] = P(i)/(2^a Q(i))$. Then,

$$\Pr[\chi = 1] = \sum_{i \in [k]} P[X = i] \cdot \Pr[\chi = 1 \mid X = i] = 2^{-a}$$

and

$$\Pr[X = i \mid \chi = 1] = \frac{\Pr[X = i \wedge \chi = 1]}{\Pr[\chi = 1]} = \frac{Q(i) \cdot P(i)/(2^a Q(i))}{2^{-a}} = P(i).$$

■

In order to combine this argument with the Substate Theorem to generate simultaneously a sample $\mathcal{M}$ of messages according to the distribution Q and several subsamples $\mathcal{M}_x$, we will need a slight extension of the above lemma.

Lemma 2 *Let P and Q be probability distributions on $[k]$ such that $2^{-a}P \leq Q$. Then, for each integer $t \geq 1$, there exist correlated random variables $\mathbf{X} = \langle X_1, X_2, \ldots, X_t \rangle$ and $\mathbf{Y} = \langle Y_1, Y_2, \ldots, Y_R \rangle$ such that*

(a) The random variables $(X_i : i \in [t])$ are independent and each X_i has distribution Q;

(b) R is a random variable with binomial distribution $B(t, 2^{-a})$;

(c) *Conditioned on the event $R = r$, the random variables $(Y_i : i \in [r])$ are independent and each Y_i has distribution P.*

(d) *$\mathbf{Y}$ is a subsequence of $\mathbf{X}$ (with probability 1).*

Proof: We generate t independent copies of the random variables (X, χ) promised by Lemma 1; this gives us $\mathbf{X} = \langle X_1, X_2, \ldots, X_t \rangle$ and $\chi = \langle \chi_1, \chi_2, \ldots, \chi_t \rangle$. Let $\mathbf{Y} \stackrel{\Delta}{=} \langle X_i : \chi_i = 1 \rangle$. It is easy to verify that $\mathbf{X}$ and $\mathbf{Y}$ satisfy conditions (a)–(d). ■

Our next lemma uses Lemma 2 to pick a sample of messages according to the average distributions Q and find sub-samples inside it for several distributions P_x. This lemma will be crucial to show the compression result for simultaneous message protocols (Theorem 1).

Lemma 3 *Let Q and $P_1, P_2, \ldots, P_N$ be probability distributions on $[k]$. Define $a_i \stackrel{\Delta}{=} S(P_i \| Q)$. Suppose $a_i < \infty$ for all $i \in [N]$. Let $s_{ij}, s_{ij}, \ldots, s_{ij}$ be functions from $[k]$ to $[0, 1]$. (In our application, they will correspond to conditional probability that the referee gives the correct answer when Alice sends a certain message from [k]). Let $p_{ij} \stackrel{\Delta}{=} E_{y \in_{P_i} [k]}[s_{ij}(y)]$. Fix $\epsilon \in (0, 1]$. Then, there exists a sequence $\mathbf{x} \stackrel{\Delta}{=} \langle x_1, \ldots, x_t \rangle$ of elements of $[k]$ and subsequences $\mathbf{y}^1, \ldots, \mathbf{y}^N$ of $\mathbf{x}$ such that*

(a) *$\mathbf{y}^i$ is a subsequence of $\langle x_1, \ldots, x_{t_i} \rangle$ where, $t_i \stackrel{\Delta}{=} \left\lceil \frac{8 \cdot 2^{(a_i+1)/\epsilon} \cdot \log(2N)}{(1-\epsilon)\epsilon^2} \right\rceil$.*

(b) *For $i, j = 1, 2, \ldots, N$, $\left| \mathop{\mathrm{E}}_{\ell \in_U [r_i]} [s_{ij}(\mathbf{y}^i[\ell])] - p_{ij} \right| \leq 2\epsilon$, where r_i is the length of $\mathbf{y}^i$.*

(c) *$t \stackrel{\Delta}{=} \max_i t_i$.*

Proof: Using part (b) of Fact 4, we obtain distributions $\widetilde{P}_i$ such that

$$\forall i \in [k], \; \left\| P_i - \widetilde{P}_i \right\|_1 \leq 2\epsilon \quad \text{and} \quad (1-\epsilon) 2^{-(a_i+1)/\epsilon} \widetilde{P}_i \leq Q.$$

Using Lemma 2, we can construct correlated random variables $(\mathbf{X}, \mathbf{Y}^1, \mathbf{Y}^2, \ldots, \mathbf{Y}^N)$ such that $\mathbf{X}$ is a sequence of $t \stackrel{\Delta}{=} \max_i t_i$ independent random variables, each distributed according to Q, and $(\mathbf{X}[1, t_i], \mathbf{Y}^i)$ satisfying conditions (a)–(d) (with $P = P_i$, $a = (a_i + 1)/\epsilon - \log(1 - \epsilon)$ and $t = t_i$). We will show that with non-zero probability these random variables satisfy conditions (a) and (b) of the present lemma. This implies that there is a choice $(\mathbf{x}, \mathbf{y}^1, \ldots, \mathbf{y}^N)$ for $(\mathbf{X}, \mathbf{Y}^1, \ldots, \mathbf{Y}^N)$ satisfying parts (a) and (b) of the present lemma.

Let R_i denote the length of $\mathbf{Y}^i$. Using part (b) of Fact 5, $\Pr[\exists i, R_i < (4/\epsilon^2)\log(2N)] < N \cdot \frac{1}{2N} = \frac{1}{2}$. Now, condition on the event $R_i \geq \left(\frac{4}{\epsilon^2}\right)\log(2N)$, for all $1 \leq i \leq N$. Define $\widetilde{p}_j \stackrel{\Delta}{=} \mathop{\Pr}_{y \in_{\widetilde{P}_i} [k]}[s_{ij}(y)]$. We use part (a) of Fact 5 to conclude that

$$\mathop{\Pr}_{\mathbf{Y}^i} \left[\left| \mathop{\mathrm{E}}_{\ell \in_U [r_i]} [s_{ij}(\mathbf{Y}^i[\ell])] - \widetilde{p}_{ij} \right| > \epsilon \right] < \frac{2}{(2N)^8}, \quad \forall i, j = 1, \ldots, N, \tag{1}$$

implying that

$$\mathop{\Pr}_{\mathbf{Y}^1, \ldots, \mathbf{Y}^N} \left[\exists i, j, \left| \mathop{\mathrm{E}}_{\ell \in_U [r_i]} [s_{ij}(\mathbf{Y}^i[l])] - \widetilde{p}_{ij} \right| > \epsilon \right] \leq N^2 \times \frac{2}{(2N)^8} < \frac{1}{2}. \tag{2}$$

From (1), (2) and the fact that $\forall i, j \ \ |p_{ij} - \widetilde{p}_{ij}| \leq \epsilon$ (since $\left\| P_i - \widetilde{P}_i \right\|_1 \leq 2\epsilon$), it follows that part (b) of our lemma holds with non-zero probability. Part (a) is never violated. Part (c) is true by definition of t. ∎

Theorem 1 (Compression result, simultaneous messages) *Suppose that Π is a δ-error private coin simultaneous message protocol for $f : \{0,1\}^n \times \{0,1\}^n \to \mathcal{Z}$. Let the inputs to f be chosen according to the uniform distribution. Let X, Y denote the random variables corresponding to Alice's and Bob's inputs respectively, and M_A, M_B denote the random variables corresponding to Alice's and Bob's messages respectively. Suppose $I(X : M_A) \leq a$ and $I(Y : M_B) \leq b$. Then, there exist sets $\mathsf{Good}_A, \mathsf{Good}_B \subseteq \{0,1\}^n$ such that $|\mathsf{Good}_A| \geq \frac{2}{3} \cdot 2^n$ and $|\mathsf{Good}_B| \geq \frac{2}{3} \cdot 2^n$, and a private coin simultaneous message protocol Π' with the following properties:*

(a) In Π', Alice sends messages of length at most $\frac{3a+1}{\epsilon} + \log(n+1) + \log\frac{1}{\epsilon^2(1-\epsilon)} + 4$ bits and Bob sends messages of length at most $\frac{3b+1}{\epsilon} + \log(n+1) + \log\frac{1}{\epsilon^2(1-\epsilon)} + 4$ bits.

(b) For each input $(x, y) \in \mathsf{Good}_A \times \mathsf{Good}_B$, the error probability of Π' is at most $\delta + 4\epsilon$.

Proof: Let P be the distribution of M_A, and let P_x be its distribution under the condition $X = x$. Note that by Fact 2, we have $\mathrm{E}_X[S(P_x\|P)] \leq a$, where the expectation is got by choosing x uniformly from $\{0,1\}^n$. Therefore there exists a set Good_A, $|\mathsf{Good}_A| \geq \frac{2}{3} \cdot 2^n$, such that for all $x \in \mathsf{Good}_A$, $S(P_x\|P) \leq 3a$.

Define $t_a \stackrel{\Delta}{=} \frac{8(n+1)2^{(3a+1)/\epsilon}}{\epsilon^2(1-\epsilon)}$. From Lemma 3, we know that there is a sequence of messages $\sigma = \langle m_1, \ldots, m_{t_a} \rangle$ and subsequences σ_x of σ such that on input $x \in \mathsf{Good}_A$, if Alice sends a uniformly chosen random message of σ_x instead of sending messages according to distribution P_x, the probability of error for any $y \in \{0,1\}^n$ changes by at most 2ϵ. We now define an intermediate protocol Π'' as follows. The messages in σ are encoded using at most $\log t_a + 1$ bits. In protocol Π'' for $x \in \mathsf{Good}_A$, Alice sends a uniformly chosen random message from σ_x; for $x \notin \mathsf{Good}_A$, Alice sends a fixed arbitrary message from σ. Bob's strategy in Π'' is the same as in Π. In Π'', the error probability of an input $(x, y) \in \mathsf{Good}_A \times \{0,1\}^n$ is at most $\delta + 2\epsilon$, and $I(Y : M_B) \leq b$. Now arguing similarly, the protocol Π'' can be converted to a protocol Π' by compressing Bob's message to at most $\log t_b + 1$ bits, where $t_b \stackrel{\Delta}{=} \frac{8(n+1)2^{(3b+1)/\epsilon}}{\epsilon^2(1-\epsilon)}$. In Π', the error for an input $(x, y) \in \mathsf{Good}_A \times \mathsf{Good}_B$ is at most $\delta + 4\epsilon$. ∎

Corollary 1 *Let $\delta, \epsilon > 0$. Let $f : \{0,1\}^n \times \{0,1\}^n \to \mathcal{Z}$ be a function. Let the inputs to f be chosen according to the uniform distribution. Then there exist sets $\mathsf{Good}_A, \mathsf{Good}_B \subseteq \{0,1\}^n$ such that $|\mathsf{Good}_A| \geq \frac{2}{3} \cdot 2^n$, $|\mathsf{Good}_B| \geq \frac{2}{3} \cdot 2^n$, and $IC^{\mathrm{sim}}_\delta(f) \geq \frac{\epsilon}{3}(R^{\mathrm{sim}}_{\delta+4\epsilon}(f') - 2\log(n+1) - 2\log\frac{1}{\epsilon^2(1-\epsilon)} - \frac{2}{\epsilon} - 8)$, where f' is the restriction of f to $\mathsf{Good}_A \times \mathsf{Good}_B$.*

We can now prove the key theorem of Chakrabarti et al. [CSWY01].

Theorem 2 (Direct sum, simultaneous messages) *Let $\delta, \epsilon > 0$. Let $f : \{0,1\}^n \times \{0,1\}^n \to \mathcal{Z}$ be a function. Define $\tilde{R}^{\mathrm{sim}}_\delta(f) \stackrel{\Delta}{=} \min_{f'} R^{\mathrm{sim}}_\delta(f')$, where the minimum is taken over all functions f' which are the restrictions of f to sets of the form $A \times B$, $A, B \subseteq \{0,1\}^n$, $|A| \geq \frac{2}{3} \cdot 2^n$, $|B| \geq \frac{2}{3} \cdot 2^n$. Then, $R^{\mathrm{sim}}_\delta(f^m) \geq \frac{m\epsilon}{3}(\tilde{R}^{\mathrm{sim}}_{\delta+4\epsilon}(f) - 2\log(n+1) - 2\log\frac{1}{\epsilon^2(1-\epsilon)} - \frac{2}{\epsilon} - 8)$.*

Proof: Immediate from Fact 7, Fact 6 and Corollary 1. ∎

Remarks:
1. The above theorem implies lower bounds for the simultaneous direct sum complexity of equality, as well as lower bounds for some related problems as in Chakrabarti et al. [CSWY01]. The dependence of the bounds on ϵ is better in our version.
2. A very similar direct sum theorem can be proved about two-party one-round private coin protocols.
3. All the results in this section, including the above remark, hold even when f is a relation.

4 Two-party multiple-round protocols

We first prove Lemma 4, which intuitively shows that if P, Q are probability distributions on $[k]$ such that $P \leq 2^a Q$, then about it is enough to sample Q independently $2^{O(a)}$ times to produce one sample element Y according to P. In the statement of the lemma, the random variable $\mathbf{X}$ represents an infinite sequence of independent sample elements chosen according to Q, the random variable R indicates how many of these elements have to be considered till 'stopping'. $R = \infty$ indicates that we do not 'stop'. If we do 'stop', then either we succeed in producing a sample according P (in this case, the sample $Y = X_R$), or we give up (in this case, we set $Y = 0$). In the proof of the lemma, $\star$ indicates that we do not 'stop' at the current iteration and hence the rejection sampling process must go further.

Lemma 4 *Let P and Q be probability distributions on $[k]$, such that* $\mathsf{Good} \stackrel{\Delta}{=} \{i \in [k] : \frac{P(i)}{2^a} \leq Q(i)\}$ *has probability exactly $1 - \epsilon$ in P. Then, there exist correlated random variables $\mathbf{X} \stackrel{\Delta}{=} \langle X_i \rangle_{i \in \mathbb{N}_+}$, R and Y such that*

(a) the random variables $(X_i : i \in \mathbb{N}_+)$ are independent and each has distribution Q;

(b) R takes values in $\mathbb{N}_+ \cup \{\infty\}$ and $\mathrm{E}[R] = 2^a$;

(c) if $R \neq \infty$, then $Y = X_R$ or $Y = 0$;

(d) Y takes values in $\{0\} \cup [k]$, such that: $\Pr[Y = i] = \begin{cases} P(i) & \text{if } i \in \mathsf{Good} \\ 0 & \text{if } i \in [k] - \mathsf{Good} \\ \epsilon & \text{if } i = 0. \end{cases}$

Proof: First, we define a pair of correlated random variables (X, Z), where X takes values in $[k]$ and Z in $[k] \cup \{0, \star\}$. Let $P' : [k] \to [0, 1]$ be defined by $P'(i) = P(i)$ for $i \in \mathsf{Good}$, and $P'(i) = 0$ for $i \in [k] - \mathsf{Good}$. Let $\beta \stackrel{\Delta}{=} \epsilon 2^{-a}/(1 - (1 - \epsilon)2^{-a})$ and $\gamma_i \stackrel{\Delta}{=} P'(i)2^{-a}/Q(i)$. The joint probability distribution of X and Z is given by

$$\forall i \in [k], \ \Pr[X = i] = Q(i) \quad \text{and} \quad \Pr[Z = j \mid X = i] = \begin{cases} \gamma_i & \text{if } j = i \\ \beta(1 - \gamma_i) & \text{if } j = 0 \\ 1 - \gamma_i - \beta(1 - \gamma_i) & \text{if } j = \star \\ 0 & \text{otherwise.} \end{cases}$$

Note that this implies that

$$\Pr[Z \neq \star] = \sum_{i \in [k]} Q(i) \cdot [\gamma_i + \beta(1 - \gamma_i)] = \beta + (1 - \beta) \sum_{i \in [k]} P'(i) 2^{-a} = \beta + (1 - \beta)(1 - \epsilon)2^{-a} = 2^{-a}.$$

Now, consider the sequence of random variables $\mathbf{X} \triangleq \langle X_i \rangle_{i \in \mathbb{N}_+}$ and $\mathbf{Z} \triangleq \langle Z_i \rangle_{i \in \mathbb{N}_+}$, where each (X_i, Z_i) has the same distribution as (X, Z) defined above and (X_i, Z_i) is independent of all $(X_j, Z_j), j \neq i$. Let $R \triangleq \min\{i : Z_i \neq \star\}$; $R \triangleq \infty$ if $\{i : Z_i \neq \star\}$ is the empty set. R is a geometric random variable with success probability 2^{-a}, and so satisfies part (b) of the present lemma. Let $Y \triangleq Z_R$ if $R \neq \infty$ and $Y \triangleq 0$ if $R = \infty$. Parts (a) and (c) are satisfied by construction.

We now verify that part (d) is satisfied. Since $\Pr[R = \infty] = 0$, we see that

$$\begin{aligned}
\Pr[Y = i] &= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \Pr[Z_r = i \mid R = r] \\
&= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \Pr[Z_r = i \mid Z_r \neq \star] \\
&= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \frac{\Pr[Z_r = i]}{\Pr[Z_r \neq \star]},
\end{aligned}$$

where the second equality follows from the independence of (X_r, Z_r) from all $(X_j, Z_j), j \neq r$. If $i \in [k]$, we see that

$$\begin{aligned}
\Pr[Y = i] &= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \frac{\Pr[Z_r = i]}{\Pr[Z_r \neq \star]} \\
&= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \frac{\Pr[X_r = i] \cdot \Pr[Z_r = i \mid X_r = i]}{\Pr[Z_r \neq \star]} \\
&= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \frac{Q(i)\gamma_i}{2^{-a}} \\
&= \sum_{r \in \mathbb{N}_+} \Pr[R = r] P'(i) = P'(i).
\end{aligned}$$

Thus, for $i \in \mathsf{Good}$, $\Pr[Y = i] = P(i)$, and for $i \in [k] - \mathsf{Good}$, $\Pr[Y = i] = 0$. Finally,

$$\begin{aligned}
\Pr[Y = 0] &= \sum_{r \in \mathbb{N}_+} \Pr[R = r] \cdot \frac{\Pr[Z_r = 0]}{\Pr[Z_r \neq \star]} \\
&= \sum_{r \in \mathbb{N}_+} \frac{\Pr[R = r]}{2^{-a}} \sum_{j \in [k]} \Pr[X_r = j] \cdot \Pr[Z_r = 0 \mid X_r = j] \\
&= \sum_{r \in \mathbb{N}_+} \frac{\Pr[R = r]}{2^{-a}} \sum_{j \in [k]} Q(j) \cdot \beta(1 - \gamma_j) \\
&= \sum_{r \in \mathbb{N}_+} \Pr[R = r]\epsilon = \epsilon.
\end{aligned}$$

■

Lemma 5 follows from Lemma 4, and will be used to prove the message compression result for two-party multiple-round protocols (Theorem 3).

Lemma 5 *Let Q and $P_1, \ldots, P_N$ be probability distributions on $[k]$. Define $S(P_i \| Q) = a_i$. Suppose $a_i < \infty$ for all $i \in [N]$. Fix $\epsilon \in (0, 1]$. Then, there exist random variables $\mathbf{X} = \langle X_i \rangle_{i \in \mathbb{N}_+}$, $R_1, \ldots, R_N$ and $Y_1, \ldots, Y_N$ such that*

(a) *$(X_i : i \in \mathbb{N}_+)$ are independent random variables, each having distribution Q;*

(b) *R_i takes values in $\mathbb{N}_+ \cup \{\infty\}$ and $\mathrm{E}[R_i] = 2^{(a_i+1)/\epsilon}$;*

(c) *Y_j takes values in $[k] \cup \{0\}$, and there is a set $\mathsf{Good}_j \subseteq [k]$ with $P_j(\mathsf{Good}_j) \geq 1-\epsilon$ such that for all $\ell \in \mathsf{Good}_j$, $\Pr[Y_j = \ell] = P_j(\ell)$, for all $\ell \in [k] - \mathsf{Good}_j$, $\Pr[Y_j = \ell] = 0$ and $\Pr[Y_j = 0] = 1 - P_j(\mathsf{Good}_j) \leq \epsilon$;*

(d) *if $R_j < \infty$, then $Y_j = X_{R_j}$ or $Y = 0$.*

Proof: Using part (a) of Fact 4, we obtain for $j = 1, \ldots, N$, a set $\mathsf{Good}_j \subseteq [k]$ such that $P_j(\mathsf{Good}_j) \geq 1-\epsilon$ and $P_j(i)2^{-(a_j+1)/\epsilon} \leq Q(i)$ for all $i \in \mathsf{Good}_j$. Now from Lemma 4, we can construct correlated random variables $\mathbf{X}$, $Y_1, \ldots, Y_N$, and $R_1, \ldots, R_N$ satisfying the requirements of the present lemma. ∎

Theorem 3 (Compression result, multiple rounds) *Suppose Π is a k-round private coin randomised protocol for $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Let the average error of Π under a probability distribution μ on the inputs $\mathcal{X} \times \mathcal{Y}$ be δ. Let X, Y denote the random variables corresponding to Alice's and Bob's inputs respectively. Let T denote the complete transcript of messages sent by Alice and Bob. Suppose $I(XY : T) \leq a$. Let $\epsilon > 0$. Then, there is another deterministic protocol Π' with the following properties:*

(a) *The communication cost of Π' is at most $\frac{2k(a+1)}{\epsilon^2} + \frac{2k}{\epsilon}$ bits;*

(b) *The distributional error of Π' under μ is at most $\delta + 2\epsilon$.*

Proof: The proof proceeds by defining a series of intermediate k-round protocols $\Pi'_k, \Pi'_{k-1}, \ldots, \Pi'_1$. Π'_i is obtained from Π'_{i+1} by compressing the message of the ith round. Thus, we first compress the kth message, then the $(k-1)$th message, and so on. Each message compression step introduces an additional additive error of at most ϵ/k for every input (x, y). Protocol Π'_i uses private coins for the first $i-1$ rounds, and public coins for rounds i to k. In fact, Π'_i behaves the same as Π for the first $i-1$ rounds. Let Π'_{k+1} denote the original protocol Π.

We now describe the construction of Π'_i from Π'_{i+1}. Suppose the ith message in Π'_{i+1} is sent by Alice. Let M denote the random variable corresponding to the first i messages in Π'_{i+1}. M can be expressed as (M_1, M_2), where M_2 represents the random variable corresponding to the ith message and M_1 represents the random variable corresponding to the initial $i-1$ messages. From Fact 1 (note that the distributions below are as in protocol Π'_{i+1} with the input distributed according to μ),

$$I(XY : M) = I(XY : M_1) + \mathop{\mathrm{E}}_{M_1}[I((XY : M_2) \mid M_1 = m_1)] = I(XY : M_1) + \mathop{\mathrm{E}}_{M_1 XY}[S(M_2^{xym_1} \| M_2^{m_1})]$$

where $M_2^{xym_1}$ denotes the distribution of M_2 when $(X, Y) = (x, y)$ and $M_1 = m_1$, and $M_2^{m_1}$ denotes the distribution of M_2 when $M_1 = m_1$. Note that the distribution of $M_2^{xym_1}$ is independent of y, as Π'_{i+1} is private coin up to the ith round. Define $a_i \stackrel{\Delta}{=} \mathrm{E}_{M_1 XY}[S(M_2^{xym_1} \| M_2^{m_1})]$.

Protocol Π'_i behaves the same as Π'_{i+1} for the first $i-1$ rounds; hence Π'_i behaves the same as Π for the first $i-1$ rounds. In particular, it is private coin for the first $i-1$ rounds. Alice generates the ith message of Π'_i using a fresh public coin C_i as follows: For each distribution $M_2^{m_1}$, m_1 ranging over all possible initial $i-1$ messages, C_i stores an infinite sequence $\mathbf{\Gamma}^{m_1} \stackrel{\Delta}{=} \langle \gamma_j^{m_1} \rangle_{j \in \mathbb{N}_+}$, where $(\gamma_j^{m_1} : j \in \mathbb{N}_+)$ are chosen independently from distribution $M_2^{m_1}$. Note that the distribution $M_2^{m_1}$ is known to both Alice and Bob as

m_1 is known to both of them; so both Alice and Bob know which part of C_i to 'look' at in order to read from the infinite sequence Γ^{m_1}. Using Lemma 5, Alice generates the ith message of Π'_i which is either $x_j^{m_1}$ for some j, or the dummy message 0. The probability of generating 0 is less than or equal to $\frac{\epsilon}{k}$. If Alice does not generate 0, her message lies in a set Good_{xm_1} which has probability at least $1 - \frac{\epsilon}{k}$ in the distribution $M_2^{xym_1}$. The probability of a message $m_2 \in \mathsf{Good}_{xm_1}$ being generated is exactly the same as the probability of m_2 in $M_2^{xym_1}$. The expected value of j is $2^{k(S(M_2^{xym_1} \| M_2^{m_1})+1)/\epsilon}$. Actually, Alice just sends the value of j or the dummy message 0 to Bob, using a prefix free encoding, as the ith message of Π'_i. After Alice sends off the ith message, Π'_i behaves the same as Π'_{i+1} for rounds $i+1$ to k. In particular, the coin C_i is not 'used' for rounds $i+1$ to k; instead, the public coins of Π'_{i+1} are 'used' henceforth.

By the concavity of the logarithm function, the expected length of the ith message of Π'_i is at most $2k\epsilon^{-1}(S(M_2^{xym_1} \| M_2^{m_1}) + 1) + 2$ bits for each (x, y, m_1) (The multiplicative and additive factors of 2 are there to take care of the prefix-free encoding). Also in Π'_i, for each (x, y, m_1), the expected length (averaged over the public coins of Π'_i, which in particular include C_i and the public coins of Π'_{i+1}) of the $(i+1)$th to kth messages does not increase as compared to the expected length (averaged over the public coins of Π'_{i+1}) of the $(i+1)$th to kth messages in Π'_{i+1}. This is because in the ith round of Π'_i, the probability of any non-dummy message does not increase as compared to that in Π'_{i+1}, and if the dummy message 0 is sent in the ith round Π'_i aborts immediately. For the same reason, the increase in the error from Π'_{i+1} to Π'_i is at most an additive term of $\frac{\epsilon}{k}$ for each (x, y, m_1). Thus the expected length, averaged over the inputs and public and private coin tosses, of the ith message in Π'_i is at most $2k\epsilon^{-1}(a_i + 1) + 2$ bits. Also, the average error of Π'_i under input distribution μ increases by at most an additive term of $\frac{\epsilon}{k}$.

By Fact 1, $\sum_{i=i}^{k} a_i = I(XY : T) \leq a$, where $I(XY : T)$ is the mutual information in the original protocol Π. This is because the quantity $\mathrm{E}_{M_1 XY}[S(M_2^{xym_1} \| M_2^{m_1})]$ is the same irrespective of whether it is calculated for protocol Π or protocol Π'_{i+1}, as Π'_{i+1} behaves the same as Π for the first i rounds. Doing the above 'compression' procedure k times gives us a public coin protocol Π'_1 such that the expected communication cost (averaged over the inputs as well as all the public coins of Π'_1) of Π'_1 is at most $2k\epsilon^{-1}(a+1)+2k$, and the average error of Π'_1 under input distribution μ is at most $\delta + \epsilon$. By restricting the maximum communication to $2k\epsilon^{-2}(a+1) + 2k\epsilon^{-1}$ bits and applying Markov's inequality, we get a public coin protocol Π'' from Π'_1 which has average error under input distribution μ at most $\delta + 2\epsilon$. By setting the public coin tosses to a suitable value, we get a deterministic protocol Π' from Π'' where the maximum communication is at most $2k\epsilon^{-2}(a+1) + 2k\epsilon^{-1}$ bits, and the distributional error under μ is at most $\delta + 2\epsilon$. ■

Corollary 2 *Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Let μ be a product distribution on the inputs $\mathcal{X} \times \mathcal{Y}$. Let $\delta, \epsilon > 0$. Then, $IC^k_{\mu,\delta}(f) \geq \frac{\epsilon^2}{2k} \cdot C^k_{\mu,\delta+2\epsilon}(f) - 2$.*

Theorem 4 (Direct sum, k-round) *Let m, k be positive integers, and $\epsilon, \delta > 0$. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Then, $R^k_\delta(f^m) \geq m \cdot \sup_{\mu,\kappa} \left(\frac{\epsilon^2}{2k} \cdot C^k_{\mu,\delta+2\epsilon}(f) - 2 - H(\kappa) \right)$, where the supremum is over all probability distributions μ on $\mathcal{X} \times \mathcal{Y}$ and partitions κ of μ.*

Proof: Immediate from Fact 7, Fact 6 and Corollary 2. ■

Corollary 3 *Let m, k be positive integers, and $\epsilon, \delta > 0$. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Then, $R^k_\delta(f^m) \geq m \cdot \left(\frac{\epsilon^2}{2k} \cdot C^k_{[],\delta+2\epsilon}(f) - 2 \right)$.*

Remarks:

1. Note that all the results in this section hold even when f is a relation.

2. The above corollary implies that the direct sum property holds for constant round protocols for the pointer jumping problem with the 'wrong' player starting (the bit version, the full pointer version and the tree version), since the product distributional complexity (in fact, for the uniform distribution) of pointer jumping is the same as its randomised complexity [NW93, PRV01].

5 Impossibility of quantum compression

In this section, we show that the information cost based message compression approach does not work in the quantum setting. We first need some preliminary definitions and lemmas.

Lemma 6 *Fix positive integers d, m and real $\epsilon > 0$. Then there is a set $\mathcal{S}$ of at most d-dimensional subspaces of $\mathbb{C}^m$ such that*

(a) $|\mathcal{S}| \leq \left(\frac{8\sqrt{d}}{\epsilon}\right)^{2md}$.

(b) For all d-dimensional subspaces W of $\mathbb{C}^m$, there is an at most d-dimensional subspace $\widehat{W} \in \mathcal{S}$ such that $\Delta(W, \widehat{W}) \leq \epsilon$.

Proof: Let $\mathcal{N}$ be a δ-dense subset of $\mathbf{U}_{m,1}$ satisfying Fact 9. For a unit vector $v \in \mathbb{C}^m$, let $\widetilde{v}$ denote the vector in $\mathcal{N}$ closest to it (ties are broken arbitrarily). Let W be a subspace of $\mathbb{C}^m$ of dimension d. Let $w = \sum_{i=1}^{d} \alpha_i w_i$ be a unit vector in W, where $\{w_1, \ldots, w_d\}$ is an orthonormal basis for W and $\sum_{i=1}^{d} |\alpha_i|^2 = 1$. Define $w' \stackrel{\Delta}{=} \sum_{i=1}^{d} \alpha_i \widetilde{w}_i$ and $\widehat{w} \stackrel{\Delta}{=} \frac{w'}{\|w'\|}$ if $w' \neq 0$, $\widehat{w} \stackrel{\Delta}{=} 0$ if $w' = 0$. It is now easy to verify the following.

(a) $\|w - w'\| \leq \delta\sqrt{d}$.

(b) $\|w'\| \geq 1 - \delta\sqrt{d}$.

(c) $\|w - \widehat{w}\| \leq 2\delta\sqrt{d}$.

Choose $\delta \stackrel{\Delta}{=} \frac{\epsilon}{2\sqrt{d}}$. Define $\widehat{W}$ to be the subspace spanned by the set $\{\widetilde{w_1}, \ldots, \widetilde{w_d}\}$. $\dim(\widehat{W}) \leq d$. By Fact 11 and (c) above, $\Delta(W, \widehat{W}) \leq \epsilon$. Define $\mathcal{S} \stackrel{\Delta}{=} \{\widehat{W} : W \text{ subspace of } \mathbb{C}^m \text{ of dimension } d\}$. $\mathcal{S}$ satisfies part (b) of the present lemma. Also $|\mathcal{S}| \leq (4/\delta)^{2md} = (8\sqrt{d}/\epsilon)^{2md}$, thus proving part (a) of the present lemma. ■

We next prove the following two propositions using Fact 10.

Proposition 1 *Let m, d, l be positive integers such that $d < \sqrt{\frac{m}{l}}$ and $l < \frac{m}{20}$. Let V be a fixed subspace of $\mathbb{C}^m$ of dimension m/l. Let P be the orthogonal projection operator on V. Let (w, w') be an independently chosen random pair of unit vectors from $\mathbb{C}^m$. Then,*

(a) $\Pr\left[|\langle w|w'\rangle| \geq \frac{1}{5d^2}\right] \leq 2\exp\left(-\frac{m}{100d^4}\right)$,

(b) $\Pr\left[\|Px\| \geq \frac{2}{\sqrt{l}}\right] \leq 2\exp\left(-\frac{m}{4l}\right)$, $x = w, w'$,

(c) $\Pr\left[|\langle w|P|w'\rangle| \geq \frac{4}{5d^2 l}\right] \leq 6\exp\left(-\frac{m}{100d^4 l}\right)$.

Proof: To prove the first inequality, we can assume by the $\mathbf{U}(m)$-invariance of $\mu_{m,1}$ that $w' = e_1$. The map $w \mapsto |\langle w|e_1\rangle|$ is 1-Lipschitz, with expectation at most $\frac{1}{\sqrt{m}}$ by $\mathbf{U}(m)$-symmetry and using convexity of the square function. By Fact 10,

$$\Pr\left[|\langle w|w'\rangle| \geq \frac{1}{5d^2}\right] \leq \Pr\left[|\langle w|w'\rangle| > 1/\sqrt{m} + 12/\sqrt{2m} + \frac{1}{10d^2}\right] \leq 2\exp\left(-\frac{m}{100d^4}\right),$$

proving part (a) of the present proposition.

The argument for the second inequality is similar. By $\mathbf{U}(m)$-symmetry and using convexity of the square function, $\mathrm{E}[\|Pw\|] = \mathrm{E}[\|Pw'\|] \leq \frac{1}{\sqrt{l}}$. Since the map $w \mapsto \|Pw\|$ is 1-Lipschitz, by Fact 10 we get that

$$\Pr\left[\|Px\| \geq \frac{2}{\sqrt{l}}\right] \leq \Pr\left[\|Px\| > \frac{1}{\sqrt{l}} + \frac{12}{\sqrt{2m}} + \frac{1}{2\sqrt{l}}\right] \leq 2\exp\left(-\frac{m}{4l}\right), x = w, w',$$

proving part (b) of the present proposition.

We now prove part (c) of the present proposition. Let $\widehat{w} \triangleq \frac{Pw}{\|Pw\|}$ and $\widehat{w'} \triangleq \frac{Pw'}{\|Pw'\|}$ (note that $\|Pw\| = 0$ and $\|Pw'\| = 0$ are each zero probability events). By Fact 8, $\widehat{w}, \widehat{w'}$ are random independently chosen unit vectors in V. By the argument used in the proof of part (a) of the present proposition, we get that

$$\Pr\left[|\langle \widehat{w}|\widehat{w'}\rangle| \geq \frac{1}{5d^2}\right] \leq 2\exp\left(-\frac{m}{100ld^4}\right).$$

Now,

$$\Pr\left[|\langle Pw|Pw'\rangle| \geq \frac{4}{5d^2l}\right] \leq 2\exp\left(-\frac{m}{100d^4l}\right) + 4\exp\left(-\frac{m}{4l}\right) \leq 6\exp\left(-\frac{m}{100d^4l}\right),$$

proving part (c) of the present proposition. ■

Proposition 2 *Let m, d, l be positive integers such that $d < \sqrt{\frac{m}{l}}$ and $l < \frac{m}{20}$. Let V be a fixed subspace of $\mathbb{C}^m$ of dimension m/l. Let P be the orthogonal projection operator on V. Let (w, w') be a random pair of orthonormal vectors from $\mathbb{C}^m$. Then,*

$$\Pr\left[|\langle w|P|w'\rangle| \geq \frac{2}{d^2l}\right] \leq 10\exp\left(-\frac{m}{100d^4l}\right).$$

Proof: By Fact 8, to generate a random pair of orthonormal vectors (w, w') from $\mathbb{C}^m$ we can do as follows: First generate unit vectors $x, y \in \mathbb{C}^m$ randomly and independently, let $w'' \triangleq y - \langle x|y\rangle x$, and set $w \triangleq x$ and $w' \triangleq \frac{w''}{\|w''\|}$. Now (note that $\Pr[w'' = 0] = 0$),

$$|\langle w|P|w'\rangle = \frac{|\langle w|P|w''\rangle|}{\|w''\|} \leq \frac{|\langle x|P|y\rangle + |\langle x|y\rangle| \langle x|P|x\rangle}{1 - |\langle x|y\rangle|}.$$

By Proposition 1 we see that,

$$\begin{aligned}
\Pr\left[|\langle w|P|w'\rangle| \geq \frac{2}{d^2l}\right] &\leq \Pr\left[|\langle w|P|w'\rangle| \geq \frac{4/(5d^2l) + (1/(5d^2)) \cdot (4/l)}{1 - (1/(5d^2))}\right] \\
&\leq 6\exp\left(-\frac{m}{100d^4l}\right) + 2\exp\left(-\frac{m}{100d^4}\right) + 2\exp\left(-\frac{m}{4l}\right) \\
&\leq 10\exp\left(-\frac{m}{100d^4l}\right),
\end{aligned}$$

proving the present proposition. ■

Lemma 7 *Let m, d, l be positive integers such that $200d^4 l \ln(20d^2) < m$. Let V be a fixed subspace of $\mathbb{C}^m$ of dimension m/l. Let P be the orthogonal projection operator on V. Let W be a random subspace of $\mathbb{C}^m$ of dimension d. Then,*

$$\Pr[\exists w \in W, \|w\| = 1 \text{ and } |\langle w|P|w\rangle| \geq 6/l] \leq \exp\left(-\frac{m}{200d^4 l}\right).$$

Proof: Let $(w_1, \ldots, w_d)$ be a randomly chosen ordered orthonormal set of size d in $\mathbb{C}^m$, and let $W \triangleq \mathrm{Span}(w_1, \ldots, w_d)$. By Fact 8, each w_i is a random unit vector of $\mathbb{C}^m$ and each (w_i, w_j), $i \neq j$ is a random pair of orthonormal vectors of $\mathbb{C}^m$. By Propositions 1 and 2, we have with probability at least $1 - 2d \exp\left(-\frac{m}{4l}\right) - 10d^2 \exp\left(-\frac{m}{100d^4 l}\right)$,

$$\forall i, \langle w_i|P|w_i\rangle < \frac{4}{l} \text{ and } \forall i, j, i \neq j, |\langle w_i|P|w_j\rangle| < \frac{2}{d^2 l}.$$

We show that whenever this happens $|\langle w|P|w\rangle| \leq 6/l$ for all $w \in W$, $\|w\| = 1$. Let $w \triangleq \sum_{i=1}^d \alpha_i w_i$, where $\sum_{i=1}^d |\alpha_i|^2 = 1$. Then,

$$\begin{aligned}
|\langle w|P|w\rangle| &= \left|\sum_{i,j} \alpha_i^* \alpha_j \langle w_i|P|w_j\rangle\right| \\
&\leq \sum_i |\alpha_i|^2 |\langle w_i|P|w_i\rangle| + \sum_{i,j:i\neq j} |\alpha_i^* \alpha_j||\langle w_i|P|w_j\rangle| \\
&< \frac{4}{l} + d^2 \cdot \frac{2}{d^2 l} \\
&= \frac{6}{l}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\Pr[\exists w \in W, \|w\| = 1 \text{ and } |\langle w|P|w\rangle| \geq 6/l] &\leq 2d \exp\left(-\frac{m}{4l}\right) + 10d^2 \exp\left(-\frac{m}{100d^4 l}\right) \\
&\leq \exp\left(-\frac{m}{200d^4 l}\right),
\end{aligned}$$

completing the proof of the present lemma. ∎

We can now prove the following 'incompressibility' theorem about (mixed) state compression in the quantum setting.

Theorem 5 (Quantum incompressibility) *Let m, d, n be positive integers and k a positive real number such that $k > 7$, $d > 160^2$, $1600d^4 k 2^k \ln(20d^2) < m$ and $32002^{2k} d^5 \ln d < n$. Let the underlying Hilbert space be $\mathbb{C}^m$. There exist n states ρ_l and n orthogonal projections M_l, $1 \leq l \leq n$ such that*

(a) $\forall l\, \mathrm{Tr}\, M_l \rho_l = 1$.

(b) $\rho \triangleq \frac{1}{n} \cdot \sum_l \rho_l = \frac{1}{m} \cdot I$, *where I is the identity operator on $\mathbb{C}^m$.*

(c) $\forall l\, S(\rho_l \| \rho) = k$.

(d) For all subspaces W of dimension d, $|\{M_l : M_l(W) \leq 1/10\}| \geq n/4$.

Proof: In the proof, we will index the n states ρ_l, $1 \leq l \leq n$ as ρ_{ij}, $1 \leq i \leq \frac{n}{2^k}$, $1 \leq j \leq 2^k$. We will also index the n orthogonal projections M_l as M_{ij}. For $1 \leq i \leq \frac{n}{2^k}$, choose $\mathcal{B}^i = (|b^i_1\rangle, \ldots, |b^i_m\rangle)$ to be a random ordered orthonormal basis of $\mathbb{C}^m$. $\mathcal{B}^i$ is chosen independently of $\mathcal{B}^{i'}$, $i' \neq i$. Partition the sequence $\mathcal{B}^i$ into 2^k equal parts; call these parts $\mathcal{B}^{ij}$, $1 \leq j \leq 2^k$. Define $\rho_j \triangleq \frac{2^k}{m} \cdot \sum_{v \in \mathcal{B}^{ij}} |v\rangle\langle v|$. Define $M_{ij} \triangleq \sum_{v \in \mathcal{B}^{ij}} |v\rangle\langle v|$. Define $V_{ij} \triangleq \mathrm{Span}(v : v \in \mathcal{B}^{ij})$. V_{ij} is the support of ρ_{ij}. It is easy to see that ρ_{ij}, M_{ij} satisfy parts (a), (b) and (c) of the present theorem.

To prove part (d), we reason as follows. Let W be a fixed subspace of $\mathbb{C}^m$ of dimension d. Let P_{ij} denote the orthogonal projection operator onto V_{ij}. By the $\mathbf{U}(m)$-invariance of the distribution $\mu_{m,d}$ and from Lemma 7, for each i, j,

$$\Pr\left[\exists w \in W,\ \|w\| = 1 \ \text{ and } \ |\langle w|P_{ij}|w\rangle| \geq \frac{6}{2^k}\right] \leq \exp\left(-\frac{m}{200 \cdot 2^k d^4}\right),$$

where the probability is over the random choice of the bases $\mathcal{B}^i$, $1 \leq i \leq \frac{n}{2^k}$. Define the set

$$\mathsf{Bad} \triangleq \{i \in [n/2^k] : \exists j \in [2^k], M_{ij}(W) \geq \frac{6}{2^k}\}.$$

Hence for a fixed $i \in \left[\frac{n}{2^k}\right]$,

$$\Pr[i \in \mathsf{Bad}] \leq 2^k \exp\left(-\frac{m}{200 \cdot 2^k d^4}\right) \leq \exp\left(-\frac{m}{400 \cdot 2^k d^4}\right).$$

Since the events $i \in Bad$ are independent,

$$\Pr\left[|\mathsf{Bad}| \geq \frac{3}{4} \cdot \frac{n}{2^k}\right] \leq \binom{\frac{n}{2^k}}{\frac{3n}{4 \cdot 2^k}} \exp\left(-\frac{3mn}{1600 \cdot 2^{2k} d^4}\right). \leq \left(\frac{4e}{3}\right)^{\frac{3n}{2^{k+2}}} \exp\left(-\frac{3mn}{1600 \cdot 2^{2k} d^4}\right).$$

So,

$$\Pr\left[\left|\left\{M_{ij} : M_{ij}(W) \geq \frac{6}{2^k}\right\}\right| \geq \frac{3n}{4}\right] \leq \left(\frac{4e}{3}\right)^{\frac{3n}{2^{k+2}}} \exp\left(-\frac{3mn}{1600 \cdot 2^{2k} d^4}\right).$$

By setting $\epsilon = 1/20$ in Lemma 6, we get

$$\begin{aligned}
&\Pr\left[\exists \widehat{W} \in \mathcal{S}, \left|\left\{M_{ij} : M_{ij}(W) \geq \frac{1}{20}\right\}\right| \geq \frac{3n}{4}\right] \\
\leq & \quad \left(\frac{4e}{3}\right)^{\frac{3n}{2^{k+2}}} (8\sqrt{d}/\epsilon)^{2md} \exp\left(-\frac{3mn}{1600 \cdot 2^{2k} d^4}\right) \\
< & \quad 1,
\end{aligned}$$

for the given constraints on the parameters. Again by Lemma 6, we get

$$\begin{aligned}
&\Pr\left[\exists W \text{ subspace of } \mathbb{C}^m, \dim(W) = d, \left|\left\{M_{ij} : M_{ij}(W) \geq \frac{1}{10}\right\}\right| \geq \frac{3n}{4}\right] \\
= & \quad \Pr\left[\exists \widehat{W} \in \mathcal{S}, \left|\left\{M_{ij} : M_{ij}(W) \geq \frac{1}{20}\right\}\right| \geq \frac{3n}{4}\right] \\
< & \quad 1.
\end{aligned}$$

This completes the proof of part (d) of the present theorem. ■

6 Conclusion and open problems

In this paper, we have shown a compression theorem and a direct sum theorem for two party multiple round private coin protocols. Our proofs use the notion of information cost of a protocol. The main technical ingredient in our compression proof is a connection between relative entropy and sampling. It is an interesting open problem to strengthen this connection, so as to obtain better lower bounds for the direct sum problem for multiple round protocols. In particular, can one improve the dependence on the number of rounds in the compression result (by information cost based methods or otherwise)?

We have also shown a strong negative result about the compressibility of quantum information. Our result seems to suggest that to tackle the direct sum problem in quantum communication, techniques other than information cost based message compression may be necessary. Buhrman et al. [BCWdW01] have shown that the bounded error simultaneous quantum complexity of EQ_n is $\theta(\log n)$, as opposed to $\theta(\sqrt{n})$ in the classical setting [NS96, BK97]. An interesting open problem is whether the direct sum property holds for simultaneous quantum protocols for equality.

Acknowledgements

We thank Ravi Kannan and Sandeep Juneja for helpful discussions, and Siddhartha Bhattacharya for enlightening us about unitarily invariant measures on homogeneous spaces. We also thank the anonymous referees for their comments on the conference version of this paper, which helped us to improve the presentation of the paper.

References

[AKN98] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 20–30, 1998. Also quant-ph/9806029.

[AS00] N. Alon and J. Spencer. *The probabilistic method*. John Wiley and Sons, 2000.

[BCWdW01] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16), 2001.

[BJKS02] Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 209–218, 2002.

[BK97] L. Babai and P. Kimmel. Randomized simultaneous messages. In *Proceedings of the 12th IEEE Conference on Computational Complexity*, pages 239–246, 1997.

[CSWY01] A. Chakrabarti, Y. Shi, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 33st Annual ACM Symposium on Theory of Computing*, pages 270–278, 2001.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley and Sons, 1991.

[FKNN95] T. Feder, E. Kushilevitz, M. Naor, and N. Nisan. Amortized communication complexity. In *SIAM Journal of Computing*, pages 239–248, 1995.

[JRS02] R. Jain, J. Radhakrishnan, and P. Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 429–438, 2002.

[JRS03] R. Jain, J. Radhakrishnan, and P Sen. A lower bound for bounded round quantum communication complexity of set disjointness function. Manuscript at quant-ph/0303138, 2003.

[KKN92] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. In *Structure in Complexity Theory*, pages 262–274, 1992.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[Mat02] J. Matoušek. *Lectures on Discrete Geometry*. Graduate Texts in Mathematics. Springer-Verlag, 2002.

[NC00] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[NS96] I. Newman and M. Szegedy. Public vs. private coin flips in one round communication games. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pages 561–570, 1996.

[NW93] N. Nisan and A. Wigderson. Rounds in communication complexity revisited. *SIAM Journal of Computing*, 22:211–219, 1993.

[PRV01] S. Ponzio, J. Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62(2):323–355, 2001.

[Ros97] S. Ross. *Simulation*. Academic Press, 1997.

[Roy88] H. Royden. *Real Analysis*. Prentice-Hall of India Pvt. Ltd., 1988.

[Yao79] A. C-C. Yao. Some complexity questions related to distributed computing. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing*, pages 209–213, 1979.