Information theoretic problems in computational complexity theory

A thesis submitted to the University of Mumbai
for the degree of
Doctor of Philosophy in Computer Science

by

Rahul Jain

School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai 400005, India

2004

STATEMENT BY THE CANDIDATE

As required by the University Ordinance 770, I wish to state that the work embodied in this thesis titled "**Information theoretic problems in computational complexity theory**" forms my own contribution to the research work carried out under the guidance of **Prof. R. K. Shyamasundar** at the Tata Institute of Fundamental Research. This work has not been submitted for any other degree of this or any other University. Whenever references have been made to previous works of others, it has been clearly indicated as such and included in the bibliography.

Certified by

Signature of Guide

Prof. R. K. Shyamasundar
Name of Guide

Signature of Candidate

Rahul Jain
Name of Candidate

To my family.

Acknowledgements

I am deeply thankful to my adviser, Jaikumar Radhakrishnan, for giving me an opportunity to work with him and for his support and guidance. His insights and clarity of thought have helped me a lot in proceeding with my work. Developing such insights and clarity looks like a (life time) hard problem for me!

I thank R. K. Shyamasundar for serving as my official guide, and giving me freedom to pursue my research interests in TIFR.

I am greatful to Funda da (a.k.a. Pranab Sen) for being very true to the name that we have given him. A very large part of all that I know in computer science goes to his credit.

I would like to thank Ashwin Nayak, Hartmut Klauck, and Andris Ambainis for several useful discussions and comments, which have directly influenced part of this work. I am extremely thankful to Andrew Yao for being very generous in his appreciation of a part of this work.

I would like to thank all the members of the School of Technology and Computer Science, for their encouragement and help that they extended to me at various stages of my stay here. I wish to thank Vivek Borkar, Subir Ghosh, Paritosh Pandya, Jaikumar Radhkrishnan, N. Raja and R. K. Shyamasundar for the courses that they have given, and all that I have learnt from them. I thank John Barretto and the other office staff for their excellent administrative support.

I would like to thank all my friends in the computer science department namely Aghav, Anoop, Benny, Kavitha, Krishnan, Narayanan, Siva Kiran, and Venkatesh for useful (and non-useful) academic and non-academic discussions.

I have been fortunate to have had great company of many friends in TIFR outside the department. In maths — IG, Holla, Amitava, Siddhartha, Arvind, Keshari and in physics — Yeshpal (specially for the fruitful fruits), Yogesh, Neel, Pratik, Ashok, Arvinder, Lakshmi, Poonam, Rajeev, and Surjeet.

And finally, I express deep thanks to my family for being very supportive in all my endeavours. Whatever I am is due to their incessant love and encouragement. I dedicate this thesis to them. I also thank GOD for everything.

Synopsis

Introduction

One of the main aims of *computational complexity theory* is determining upper and lower bounds on the amount of resources required to solve a computational problem. The resources considered most commonly are time and space. However, when several agents are required to arrive at an answer to a problem whose input is distributed among them, the amount of communication required often determines whether the solution is efficient. Thus, one is lead to study various forms of communication costs associated with computational problems whose input is distributed between several agents. This has resulted in a rich area called *communication complexity theory*, with connections and applications to various other branches of complexity theory.

The advent of the quantum model of computation, lead to the re-examination of communication complexity of various problems when the agents exchange qubits instead of classical bits. For several problems, the quantum model is known to be much more powerful than the classical model. That is, if the agents are allowed quantum operations, then they can solve certain computational problem by exchanging far fewer qubits than they would require if they were constrained to communicate classical bits and perform classical operations. The study of such solutions and their limitations is the aim of *quantum communication complexity theory*. The notion of information plays an important role in communication complexity both in the classical and the quantum settings. In this thesis, we develop and apply information theoretic tools to show lower bounds on the communication complexity of several problems.

In the following sections, we formally define the communication complexity models, define the problems considered in this thesis, and present the results we obtain.

Computational models and problems studied

Two-party communication model

We first describe the classical model. In the two-party private coin randomised communication complexity model [Yao79], two players Alice and Bob are required to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$. Alice is given $x \in \mathcal{X}$ and Bob is given $y \in \mathcal{Y}$ by following a protocol Π. 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 $R_\delta^k(f)$, is the communication cost of the best private coin k-round δ-error protocol for f.

We also consider private coin randomised simultaneous message protocols. In such protocols, in addition to the two players there is a referee. The inputs are still with Alice and Bob, and are not known to the referee. Each player sends a message to the referee who then computes the function. $R_\delta^{\mathrm{sim}}(f)$ denotes the δ-error private coin randomised simultaneous message communication complexity of f.

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, when the inputs are drawn according to the distribution μ, is least $1-\delta$. The k-round δ-error distributional communication complexity of f, denoted $C_{\mu,\delta}^k(f)$, is the communication cost of the best k-round deterministic protocol for f with distributional error δ. We say that μ is 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 distributional communication complexity of f under product distributions is defined as $C_{[],\delta}^k(f) = \sup_\mu C_{\mu,\delta}^k(f)$, where the supremum is taken over all product distributions μ on $\mathcal{X} \times \mathcal{Y}$.

Whenever δ is omitted, we mean that $\delta = \frac{1}{3}$.

Problems studied

The first two results in this thesis are related the direct sum problem in classical communication complexity. For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, let $f^m : \mathcal{X}^m \times \mathcal{Y}^m \to \mathcal{Z}^m$ be defined by $f^m(\langle x_1, \ldots, x_m \rangle, \langle y_1, \ldots, y_m \rangle) \stackrel{\Delta}{=} \langle f(x_1, y_1), \ldots, f(x_m, y_m) \rangle$. In the direct sum problem, one studies the communication complexity of f^m as the parameter m increases.

The direct sum problem has received a lot of attention because of its connections with showing lower bounds in *circuit complexity*. Recently, interest in this problem was revived through a result of Chakrabarti et al. [CSWY01], who showed a lower bound for the communication complexity of f^m in the simultaneous message model. We extend their arguments to get a similar result when there are more rounds of communication.

Result (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,

$$R_\delta^k(f^m) \geq m \cdot \left(\frac{\epsilon^2}{2k} \cdot C_{[],\delta+2\epsilon}^k(f) - 2 \right).$$

The key ingredient of our proof is a result showing that in a communication protocol, messages can be compressed roughly to the amount of information they carry about the inputs. A similar result was proved using ad hoc arguments in [CSWY01]. Our proof makes direct use of the relative entropy of distributions, and gives us the following stronger result.

Result (Message compression, multiple-rounds) Suppose that Π is a k-round private coin randomised protocol for $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$. Let the average error of Π under a probability distribution μ on the inputs 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 $I(XY : T) \leq a$, where I(XY:T) is the mutual information between the random variables XY and T. 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 $\delta + 2\epsilon$.

We also consider the corresponding problem in the context of quantum communication. We show that the compression of messages which is possible in classical communication is impossible in quantum communication. This amounts to showing the existence of certain quantum states and projective measurements.

Result(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) $\forall l \operatorname{Tr} M_l \rho_l = 1$.

(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, $|\{l : \operatorname{Tr} M_l \sigma_l \leq 1/10\}| \geq n/4$.

The two-party quantum communication model

This model was defined by Yao [Yao93] to study communication as a resource in quantum computation. Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be arbitrary finite sets and $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ be a function. There are two players Alice and Bob, who hold qubits. When the communication game starts, Alice holds $|x\rangle$ where $x \in \mathcal{X}$ together with some ancilla qubits in the state $|0\rangle$, and Bob holds $|y\rangle$ where $y \in \mathcal{Y}$ together with some ancilla qubits in the state $|0\rangle$. Thus the qubits of Alice and Bob are initially in computational basis states, and the initial superposition is simply $|x\rangle_A |0\rangle_A |y\rangle_B |0\rangle_B$. Here the subscripts denote the ownership of the qubits by Alice and Bob. The players take turns to communicate to compute $f(x, y)$. Suppose it is Alice's turn. Alice can make an arbitrary unitary transformation on her qubits and then send one or more qubits to Bob. Sending qubits does not change the overall superposition, but rather changes the ownership of the qubits, allowing Bob to apply his next unitary transformation on his original qubits plus the newly received qubits. At the end of the protocol, the last recipient of qubits performs a measurement on the qubits in his/her possession to output an answer. We say a quantum protocol computes f with ϵ-error in the worst case, if for any input $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the probability that the protocol outputs the correct result $f(x, y)$ is greater than $1 - \epsilon$. The term 'bounded error quantum protocol' means that $\epsilon = 1/3$.

We require that Alice and Bob make a secure copy of their inputs before beginning the protocol. This is possible since the inputs to Alice and Bob are in computational basis states. Thus, without loss of generality, the input qubits of Alice and Bob are never sent as messages, their state remains unchanged throughout the protocol, and they are never measured i.e. some work qubits are measured to determine the result of the protocol. We call such protocols *secure*. We will assume henceforth that all our protocols are secure.

We also consider *protocols with prior entanglement*, where Alice and Bob already possess parts of a common state which is unentangled with the inputs. Our results apply only to protocols without prior entanglement unless we explicitly state that prior entanglement is allowed.

Problems studied

Substate theorem

An important contribution of this work is a theorem, called *Substate Theorem*, about relative entropy; it states, roughly, that if the relative entropy, $S(\rho \| \sigma) \stackrel{\Delta}{=} \mathrm{Tr}\ \rho(\log \rho - \log \sigma)$, of two quantum states ρ and σ is at most c, then $\frac{\rho}{2^{O(c)}}$ *sits inside* σ. We shall present below two natural problems in whose solution this result plays a crucial part.

Result (Substate theorem) Suppose ρ and σ are quantum states in the same finite dimensional Hilbert space, and $S(\rho \| \sigma) \leq c$. Then, for all $r > 1$, there are states ρ' and ρ'' such that $\|\rho - \rho'\|_t \leq \frac{2}{\sqrt{r}}$ and $\sigma = \alpha\rho' + (1 - \alpha)\rho''$, where $\alpha = 2^{-O(rc)}$.

A consequence of the above result is that if a POVM element F has probability p in ρ, then it has probability at least $\frac{p}{2^{O(c/p^2)}}$ in σ. Another consequence is that $\|\rho - \sigma\|_t \leq 2 - 2^{-O(c)}$. Fuchs and van de Graaf's connection [FC95] between fidelity [Joz94] and trace distance now implies that the fidelity of ρ and σ is lower bounded by $2^{-O(c)}$.

Pointer chasing problem: The full version

Our first application of the Substate Theorem concerns the *pointer chasing* problem in two-party communication complexity.

> Let V_A and V_B be disjoint subsets of size n. Player A is given a function $F_A : V_A \rightarrow V_B$ and player B is given a function $F_B : V_B \rightarrow V_A$. Let $F \triangleq F_A \cup F_B$. There is a fixed vertex s in V_B. A and B need to communicate to determine $t = F^{(k+1)}(s)$, where k and s are known to both parties in advance.

If B starts the communication, then there is a straightforward classical deterministic protocol where one of the players can determine t after k messages of $\log n$ bits have been exchanged. It appears much harder, however, to solve the problem efficiently with k messages, when A is required to send the first message. We refer to this as the pointer chasing problem P_k.

The pointer chasing problem was studied recently in the quantum communication complexity model by Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01a], who, using interesting information-theoretic techniques, showed a lower bound of $\Omega(\frac{n}{2^{2^{O(k)}}})$ for the *bit version* (defined later in this section) of this problem. They *did not consider* the full pointer version of the problem. We prove the following for the full version.

Result For any constant k, the bounded error quantum communication complexity of the pointer jumping problem P_k (full pointer version) is $\Omega(n \log^{(k)} n)$.

Our proof uses a round elimination argument (using substate theorem) and correlated input generation to arrive at this result. This matches an upper bound due to Damm, Jukna and Sgall [DJS98] of $O(n \log^{(k)} n)$, for constant k. Ponzio, Radhakrishnan and Venkatesh [PRV01a] have shown the same lower bound in the classical communication model.

Privacy and communication complexity

Our second application of the substate theorem concerns the index function problem [MNSW98a, Nay99].

> There are two players A and B. A is given an input $x \in \{0,1\}^n$ and B is given an index $i \in [n]$. They must exchange messages so that in the end B knows x_i.

In the classical setting, the index function problem (under the name *set membership problem*) was considered by Miltersen, Nisan, Safra and Wigderson [MNSW98a] in the classical setting. They showed that if B sends a total of at most b bits, then A must send $n/2^{O(b)}$ bits. Note that this is optimal as there is a trivial protocol where B sends the first b bits of his index to A, and A replies by sending the corresponding part of her bit string.

In the quantum setting, Nayak [Nay99] (see also Cleve et al. [CvDNT98]) showed that if B sends no messages at all, then A must send at least $\Omega(n)$ bits. This bound holds even if the players share EPR pairs in advance, or if A and B interact but B's messages do not depend on his input i. However, the case where B is allowed to send a few qubits based on his input in order to reduce the communication from A, does not seem to have been considered before.

In this thesis, we generalise Nayak's result to a statement of the following form: if B 'leaks' only a small number of bits of information about his input, then A must 'leak' a large number of bits of information about her input. Before we present our result, let us explain what we mean when we say that B 'leaks' only a small number of bits of information about his input. Fix a protocol for the index function problem. Assume that B's input J is a random index $i \in [n]$. Suppose B operates faithfully according to the protocol, but A deviates from it and manages to get her registers R entangled with J: we say that B leaks only b bits of information about his input if the mutual information between J and R, $I(J : R)$, is at most b. This upper bound of b on the information loss must hold for all strategies adopted by A, which have the property that the reduced density matrix of Bob's qubits is, at all times the same as in the original protocol. In other words, A wants to cheat and gather a lot of information about B's input, but B should not be able to figure out that A is cheating. Note that we do not assume that B's messages contain only b qubits, they can be arbitrarily long. In the quantum setting, A has a big bag of tricks she can use in order to extract information from B; for example, she can place a superposition of states in her input register and extract information about B's input (see [CvDNT98, Kla02] for details). Our definition of privacy loss is inspired by the above example. Let Π be a protocol for solving the index function problem INDEX_n. Let X and Y be the input registers of Alice and Bob respectively. Let A and B be other workspace registers in the possession of Alice and Bob respectively. Let $\mu_{\mathcal{X}}$ and $\mu_{\mathcal{X}}$ be distributions on $\mathcal{X}$ and $\mathcal{Y}$ respectively. We consider a 'cheating' run of Π when mixture $\mu_{\mathcal{X}}$ is fed to register X and superposition $|\mu_{\mathcal{Y}}\rangle \stackrel{\Delta}{=} \sum_y \sqrt{\mu_{\mathcal{Y}}(y)}\,|y\rangle$ fed to register Y. Let $\mu = \mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$. Let $I(X : BY)$ denote the mutual information X with Bob's registers BY at the end of this run of Π. We make the following definition.

Definition (Privacy loss) The *privacy loss* of Π for function f on the product distribution μ from Alice to Bob is defined as $L^{\Pi}(f, \mu, A, B) \stackrel{\Delta}{=} I(X : BY)$. The privacy loss from Bob to Alice, $L^{\Pi}(f, \mu, B, A)$, is defined similarly. The privacy loss of Π for f under distribution μ, $L^{\Pi}(f, \mu)$, is the larger of $L^{\Pi}(f, \mu, A, B)$ and $L^{\Pi}(f, \mu, B, A)$.

We show the following.

Result (INDEX_n privacy loss) Consider a quantum protocol Π for index function problem (INDEX_n) with worst case error at most $1/3$. Let μ denote the uniform probability distribution on Alice's and Bob's inputs. Suppose $L^{\Pi}(\mathrm{INDEX}_n, \mu, B, A) \leq k$. Then, $L^{\Pi}(\mathrm{INDEX}_n, \mu, A, B) \geq n/2^{O(k)}$.

The following corollaries are immediate.

Corollary 1 Let Π be a quantum protocol for INDEX_n with worst case error at most $1/3$. Suppose Bob sends at most k qubits to Alice. Let m be the number of qubits communicated by Alice to Bob. Then, $m = n/2^{O(k)}$.

Corollary 2 For the index function problem, one of the players must leak $\Omega(\log n)$ bits of information about his input, i.e. for any protocol Π, $L^{\Pi}(\mathrm{INDEX}_n, \mu) = \Omega(\log n)$, where μ is the uniform distribution on the inputs.

The index function problem is just one of several problems where results like above can be proved using our technique. In fact, it follows easily that if the communication matrix of the function has VC-dimension at least k, then one of the players must leak at least $\Omega(\log k)$ bits of information about his input, when the inputs are chosen according to the uniform distribution. In particular, this implies an $\Omega(\log n)$ loss in privacy for the set-disjointness and inner product modulo 2 problems.

Pointer chasing problem: The bit version

We also consider the bit version of the pointer chasing problem. It was originally studied by Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01b] in the two-party quantum communication model.

> Let V_A and V_B be disjoint sets of size n. Alice is given a function $F_A : V_A \rightarrow V_B$ and player Bob is given a function $F_B : V_B \rightarrow V_A$. Let $F \overset{\Delta}{=} F_A \cup F_B$. There is a fixed vertex s in V_B. The players need to exchange messages and determine the most significant bit of $F^{(k+1)}(s)$, where k and s are known to both parties in advance.

We refer to this as the pointer chasing problem P_k^{bit}. We prove the following.

Result In any quantum protocol for P_k^{bit}, the two players must exchange $\Omega(\frac{n}{k^2})$ qubits.

This improves the previous best bound of $\Omega(\frac{n}{2^{2^{O(k)}}})$ in [KNTZ01b] (although their bound holds also for protocols which start with some prior entanglement), and comes significantly closer to the best upper bounds known $O(n + k \log n)$ (classical deterministic [PRV01a]) and $O(k \log n + \frac{n}{k}(\log^{\lceil k/2 \rceil}(n) + \log k))$ (classical randomised [KNTZ01b]). Our proof uses similar round elimination argument with correlated input generation (as in the problem P_k,) making better use of the information theoretic tools than in previous papers.

Lower bounds for multi-party quantum communication complexity

We show lower bounds in the multi-party quantum communication complexity model. In this model, there are t parties where the ith party has input $X_i \subseteq [n]$. These parties communicate with each other by transmitting qubits to determine with high probability the value of some function F of their combined input $(X_1, X_2, \ldots, X_t)$. We consider the class of functions whose value depends only on the intersection of $X_1, X_2, \ldots, X_t$; that is, for each F in this class there is an $f_F : 2^{[n]} \rightarrow \{0, 1\}$, such that

$$F(X_1, X_2, \ldots, X_t) = f_F(X_1 \cap X_2 \cap \ldots \cap X_t).$$

The special case of this problem when there are two parties, and the function $F(X_1, X_2)$ is 1 if and only if X_1 and X_2 are disjoint, is the set-disjointness problem. This problem has a long history. In the bounded error classical setting Babai, Frankl and Simon [BFS86] showed a lower bound of $\Omega(\sqrt{n})$. This was improved to an $\Omega(n)$ lower bound by Kalyanasundaram and Schnitger [KS92]; their proof was simplified by Razborov [Raz92]. There is a straightforward protocol with $n+1$ bits of communication where Alice sends her entire input to Bob, who computes the answer and returns it to Alice. Interest in the communication complexity of several problems related to the set-disjointness function has been revived recently because of their connection to showing lower bounds in the classical data-stream model. One of these problem is the $\mathcal{L}_\infty$ promise problem: Alice and Bob are given inputs $X_A, X_B \in \{0, 1, \ldots, m\}^n$, with the promise that either for all $i \in [n]$, $|X_A[i] - X_B[i]| \leq 1$ or there exists an $i \in [n]$, such that $|X[i] - Y[i]| = m$; they must communicate in order to distinguish between these two types of inputs. For this problem, Saks and Sun [SS02] showed a lower bound of $\Omega(n/m^2)$ in a restricted model; their lower bound was strengthened by Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02], who obtained the same lower bound without any restrictions.

In the quantum setting, the set-disjointness function was first addressed by Buhrman, Cleve and Wigderson [BCW98], who showed that there is a protocol for this problem with $O(\sqrt{n}\log n)$ bits of communication. This bound was improved to $O(\sqrt{n}c^{\log^* n})$, where c is a small constant, by Hoyer and de Wolf [HdW02], and recently to $O(\sqrt{n})$ by Aaronson and Ambainis [AA03]. By a result of Razborov [Raz02] this last bound is optimal.

We show the following.

Result The t-party k-round quantum communication complexity of F is $\Omega(s_m(f_F)/(k^2))$, where $s_m(f_F)$ stands for the 'monotone sensitivity of f_F' and is defined by

$$s_m(f_F) \stackrel{\Delta}{=} \max_{S \subseteq [n]} |\{i : f_F(S \cup \{i\}) \neq f_F(S)\}|.$$

This result also holds for protocols with prior entanglement. For two-party quantum communication protocols for the set- disjointness problem, this implies that the two parties must exchange $\Omega(n/k^2)$ qubits. An upper bound of $O(n/k)$ can be derived from the $O(\sqrt{n})$ upper bound due to Aaronson and Ambainis [AA03](see also [BCW98] and [HdW02]). For $k = 1$, our lower bound matches the $\Omega(n)$ lower bound observed by Buhrman and de Wolf [BdW01] (based on a result of Nayak [Nay99]), and for $2 \leq k \ll n^{1/4}$, improves the lower bound of $\Omega(\sqrt{n})$ shown by Razborov [Raz02]. For protocols with no restrictions on the number of rounds, we can conclude that the two parties must exchange $\Omega(n^{1/3})$ qubits. This, however, falls short of the optimal $\Omega(\sqrt{n})$ lower bound shown by Razborov [Raz02].

Our result is obtained by adapting to the quantum setting the elegant *information-theoretic* arguments of Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02]. Using this method, in a related work (not included in this thesis) we can show similar lower bounds for the $\mathcal{L}_\infty$ function considered in [BJKS02].

Publications

1 "Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states." In proceedings of 43rd IEEE Symposium on Foundations of Computer Science (FOCS), 2002, pp. 429–438. (With Jaikumar Radhakrishnan and Pranab Sen.)

2 "The quantum communication complexity of the pointer chasing problem: the bit version." In proceedings of 22nd conference on the Foundations of Software Technology and Theoretical Computer Science (FSTTCS), 2002, pp. 218–229. (With Jaikumar Radhakrishnan and Pranab Sen.)

3 "A direct sum theorem in communication complexity via message compression." In proceedings of 30th International Colloquium on Automata, Languages and Programming (ICALP), 2003. (With Jaikumar Radhakrishnan and Pranab Sen.)

4 "A lower bound for bounded round quantum communication complexity of set disjointness." Submitted, available at quant-ph/0303138, 2003. (With Jaikumar Radhakrishnan and Pranab Sen.)

Contents

List of Figures

Chapter 1

Introduction

In *computational complexity* theory, one studies the amount of resources required to perform various computational tasks. In any such study one has to fix the model of computation and identify the resources that are to be used to measure the efficiency of algorithms that perform those tasks. For example, in classical complexity theory, one uses the Turing machine model for computation and studies the space and time used by algorithms. Similarly in the quantum circuits model (see Nielsen and Chuang [NC00]), one computes using circuits made out of basic quantum gates with the goal of minimizing the number of operations.

In this thesis, we study computational tasks where the input is distributed among several agents. We focus on the communication complexity of such tasks, that is, the minimum number of bits that the agents must exchange in order to complete the task. Communication complexity has a well-developed theory, with surprising and deep connections to other areas of computational complexity, e.g., VLSI circuits, data structures, pseudorandomness and boolean circuits. The book of Kushilevitz and Nisan [KN97] contains a comprehensive discussion of the techniques and applications of communication complexity.

We study communication complexity problems using information theoretic tools, some standard and some developed in this thesis. In this chapter, we define our models, both in the classical and quantum settings, introduce the problems we consider, and present our results. In our discussions of the quantum model, we assume that the reader is familiar with the basics of quantum computation.

1.1 Classical communication complexity

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 work. In this, both Alice and Bob send a message each to a referee who then decides on the answer $f(x,y)$. $R^{\mathrm{sim}}_\delta(f)$ 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

$$C^k_{\times,\delta}(f) = \sup_{\mu} C^k_{\mu,\delta}(f),$$

where the supremum is taken over all product distributions μ on $\mathcal{X} \times \mathcal{Y}$. When δ is omitted, we mean that $\delta = \frac{1}{3}$.

1.1.1 The direct sum problem

For a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$, the m-fold *direct sum* is the function $f^m : \mathcal{X}^m \times \mathcal{Y}^m \to \mathcal{Z}^m$, defined by

$$f^m(\langle x_1, \ldots, x_m\rangle, \langle y_1, \ldots, y_m\rangle) \stackrel{\Delta}{=} \langle f(x_1,y_1), \ldots, f(x_m,y_m)\rangle.$$

One then studies the communication complexity of f^m as the parameter m increases. This is referred to as the *direct sum problem* for communication complexity.

Background: The direct sum problem for communication complexity has been extensively studied in the past (see Kushilevitz and Nisan [KN97]). One of its important applications is in showing lower bounds in circuit complexity. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. In the deterministic two-party model, Feder, 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 the deterministic model for non partial 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. For the private coin randomised model, [FKNN95] showed that for the equality function,

$$\mathrm{EQ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}, R(EQ_n^m) = O(m + \log n).$$

Recently Chakrabarti et al. [CSWY01] considered the direct sum problem in the bounded error simultaneous message private coin model and showed that 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

$$\tilde{R}^{\mathrm{sim}}(f) \triangleq \min_S R(f|_{S\times S}),$$

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.

In this work, we prove lower bounds for the direct sum problem for protocols with more than one round of communication. We prove the following theorem.

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,

$$R_\delta^k(f^m) \geq m \cdot \left(\frac{\epsilon^2}{2k} \cdot C_{\times,\delta+2\epsilon}^k(f) - 2\right).$$

Above theorem follows from the following compression result via standard information theoretic arguments.

Theorem Let X and M be random variables (with some joint distribution), where X is uniformly distributed over $\{0,1\}^n$ and their mutual information $I(X : M) \leq a$. Let $[m]$ be the range of M. Let $S_y^x, x, y \in \{0,1\}^n$ be randomised predicates from $[m]$ to $[0,1]$. Then, there exists a random variable M' (correlated with X) such that

(a) M' takes values in a set of size $n \cdot 2^{O(a/\epsilon)}$;

(b) There exists $A \subseteq \{0,1\}^n$ of size at least $\frac{2}{3} \cdot 2^n$ such that for all $x \in A$ and $y \in \{0,1\}^n$,

$$|\Pr[S_y^x(M') \mid X = x] - \Pr[S_y^x(M) \mid X = x]| \leq \epsilon.$$

In our context, the above theorem states that if Alice's message contains only a bits of information about her input, she can compress it to $O(a + \log n)$ bits without changing the error probability of the protocol significantly. A similar message compression argument holds for Bob too. This gives us an alternative proof of the main result of Chakrabarti et al. [CSWY01], with better dependence on the parameters.

In order to prove the above result we establish a connection between *relative entropy* (defined later) and sampling which we believe is an important contribution of this work. Besides giving a simpler and more transparent proof of Chakrabarti et al.'s [CSWY01] main result, 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.

A quantum analogue? One might ask if a similar compression of messages is possible in the quantum setting (see section 1.2 for definition of quantum communication complexity model). 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 [JRS02a]. 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 density matrices ρ'_x, $x \in \{0,1\}^n$ such that there is no subspace of dimension less than $n^{1/5}$ that contains the supports of most of the ρ'_x. This strong negative result seems to suggest that new techniques (not based on information cost) may be required to tackle the direct sum problem for quantum communication.

1.2 Quantum communication complexity

We consider two-party quantum communication protocols as defined by Yao [Yao93]. Let $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{Z}$ be a function. There are two players Alice and Bob, who hold qubits. When the communication protocol $\mathcal{P}$ starts, Alice holds $|x\rangle$ where $x \in \mathcal{X}$ together with some ancilla qubits in the state $|0\rangle$, and Bob holds $|y\rangle$ where $y \in \mathcal{Y}$ together with some ancilla qubits in the state $|0\rangle$. Alice and Bob may also share an input independent prior entanglement. Thus, the initial superposition is simply $|x\rangle_A|0\rangle_A|\psi\rangle|y\rangle_B|0\rangle_B$, where $|\psi\rangle$ is a pure state providing the input independent prior entanglement. Here the subscripts denote the ownership of the qubits by Alice and Bob. Some of the qubits of $|\psi\rangle$ belong to Alice, the rest belong to Bob. The players take turns to communicate to compute $f(x, y)$. Suppose it is Alice's turn. Alice can make an arbitrary unitary transformation on her qubits and then send one or more qubits to Bob. Sending qubits does not change the overall superposition, but rather changes the ownership of the qubits, allowing Bob to apply his next unitary transformation on his original qubits plus the newly received qubits. At the end of the protocol, the last recipient of qubits performs a measurement in the computational basis of some qubits in her possession (the 'answer qubits') to output an answer $\mathcal{P}(x, y)$. We say that protocol $\mathcal{P}$ computes f with ϵ-error in the worst case, if $\max_{x,y} \Pr[\mathcal{P}(x, y) = f(x, y)] \geq 1 - \epsilon$. We say that $\mathcal{P}$ computes f with ϵ-error with respect to a probability distribution μ on $\mathcal{X} \times \mathcal{Y}$, if

$$\Pr_{\mu}[\mathcal{P}(x, y) = f(x, y)] \geq 1 - \epsilon.$$

We require that Alice and Bob make a 'safe' copy of their inputs (using, for example, CNOT gates) before beginning protocol $\mathcal{P}$. This is possible since the inputs x and y are in computational basis states. Thus, the input qubits of Alice and Bob are never sent as messages, their state remains unchanged throughout the execution of $\mathcal{P}$, and they are never measured i.e. some work qubits are measured to determine the result $\mathcal{P}(x, y)$. We call such protocols *safe*, and henceforth, we will assume without loss of generality that all our protocols are safe. Note that in a safe protocol one can assume that, in each round, the player whose turn it is to send the next message, has a *set of unitary transformation, one for each input,* that she applies in the rest of the qubits in her possession to generate her message.

Given a probability distribution μ on $\mathcal{X} \times \mathcal{Y}$, we define

$$|\mu\rangle \triangleq \sum_{x,y} \sqrt{\mu(x, y)}\, |x\rangle|y\rangle.$$

We define the success probability of $\mathcal{P}$ when superposition $|\mu\rangle$ is fed to Alice's and Bob's inputs, to be the probability that measuring the inputs and the answer qubits in the computational basis at the end of $\mathcal{P}$ produces consistent results. Since $\mathcal{P}$ is safe, the success probability of $\mathcal{P}$ on superposition $|\mu\rangle$ is equal to

$$E_{\mu}[\mathcal{P}(x, y) = f(x, y)].$$

We call a protocol *clean* if the final state of the work qubits (i.e all the qubits other than the input and the answer qubits) of Alice and Bob, is the state $|\mathbf{0}\rangle$.

1.2.1 The pointer chasing problem

In the two-party quantum communication complexity model we consider two versions of the *pointer chasing problem*, namely the *full pointer version* and the *bit version*, and give lower bounds on the amount of communication required.

The full pointer version

The full version of the pointer chasing problem is defined as follows.

> Let V_A and V_B be disjoint subsets of size n. Player A is given a function $F_A : V_A \to V_B$ and player B is given a function $F_B : V_B \to V_A$. Let $F = F_A \cup F_B$. There is a fixed vertex s in V_B. A and B need to communicate to determine $t = F^{(k+1)}(s)$ with probability of correctness being at least $3/4$; k and s are known to both parties in advance.

If B starts the communication, then there is a straightforward classical deterministic protocol where one of the players can determine t after k messages of $\log n$ bits have been exchanged. It appears much harder, however, to solve the problem efficiently with k messages, when A is required to send the first message. We refer to this as the pointer chasing problem P_k.

Background: The pointer chasing problem has been well-studied in the past to show rounds versus communication tradeoffs in classical communication complexity. Following some earlier results of Papadimitriou and Sipser [PS84], and Duris, Galil and Schnitger [DGS87], Nisan and Wigderson [NW93] showed that A and B must exchange $\Omega(n/k - k \log n)$ bits to solve P_k; their bound was improved by Klauck [Kla00] to $\Omega(\frac{n}{k} + k)$. These lower bounds hold even when A and B can toss coins and err with some small probability. This bound was further improved by [PRV01b] to $\Omega(n \log^{(k)} n)$ and thus matching the upper bound of $O(n \log^{(k)} n)$ due to [DJS98].

The pointer chasing problem has been studied recently in the quantum communication complexity model by Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01a], who, using interesting information-theoretic techniques, showed a lower bound of $\Omega(n/2^{2^{O(k)}})$ for the *bit version* of this problem where only the least significant bit of t is required. They *did not consider* the full version of the problem. (Note that the classical application of the lower bound for P_k to monotone circuit depth in the paper of Nisan and Wigderson [NW93, Theorem 2.7] is valid for the full version of the problem, not just for the bit version.) We show the following for the full version.

Result 2: For any constant k, the bounded error quantum communication complexity of the pointer jumping problem P_k (full pointer version) is $\Omega(n \log^{(k)} n)$.

In order prove this result, an important information theoretic tool that we developed and used is the quantum analogue of the *substate theorem.* We describe the substate theorem and its quantum analogue in the next section (Section 1.3).

The bit version

The bit version of the pointer chasing problem P_k^{bit} is formally defined as follows:.

> Let V_A and V_B be disjoint sets of size n. Alice is given a function $F_A : V_A \to V_B$ and player Bob is given a function $F_B : V_B \to V_A$. Let $F \triangleq F_A \cup F_B$. There is a fixed vertex s in V_B. The players need to exchange messages and determine the least significant bit of $F^{(k+1)}(s)$, where k and s are known to both parties in advance.

As is the case with the full version, if Bob starts the communication, there is a straightforward classical deterministic protocol where one of the players can determine the answer after k messages of $\log n$ bits have been exchanged and it appears much harder, to solve the problem efficiently with k messages, when Alice is required to send the first message.

Background: The results of Nisan and Wigderson [NW93] of $\Omega(n/k - k \log n)$ and of Klauck [Kla00] of $\Omega(\frac{n}{k} + k)$ hold for the P_k^{bit} as well. As mentioned earlier these lower bounds hold even if randomisation is allowed. A deterministic protocol with $O(n + k \log n)$ bits of communication was given by Ponzio, Radhakrishnan and Venkatesh [PRV01b], and a classical randomised protocol with $O(k \log n + \frac{n}{k}(\log^{\lceil k/2 \rceil}(n) + \log k))$ bits by Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01b]. Thus, the lower and upper bounds are quite close in the the classical setting.

As mentioned earlier, this problem has been studied recently by Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01b] in the quantum communication complexity, who, using interesting information-theoretic techniques, showed a lower bound of $\Omega(\frac{n}{2^{2^{O(k)}}})$. This bound deteriorates rapidly with k, and becomes trivial for $k \geq \log \log n$. We improve this lower bound.

Result: In any bounded error quantum protocol for the pointer chasing problem P_k^{bit}, Alice and Bob must exchange $\Omega(\frac{n}{k^2})$ qubits.

1.3 The substate theorem

Let P and $\widetilde{P}$ be two classical distributions. Let $\left\|P - \widetilde{P}\right\|_t$ denote the *total variation distance* between probability distributions P and $\widetilde{P}$. The following is easy to show:

Proposition (Substate theorem, classical version) Suppose P and Q are probability distributions on $[k]$ such that $S(P\|Q) = a$. Let $r \geq 1$. Then there is a distribution $\widetilde{P}$ on $[k]$ such that

$$\left\|P - \widetilde{P}\right\|_t \leq \frac{2}{r}, \qquad \alpha\widetilde{P} \leq Q,$$

where

$$\alpha \triangleq \left(\frac{r-1}{r}\right) 2^{-r(a+1)}.$$

An important contribution of this work is a quantum analogue of Proposition 1.3.

Result 3 (Substate theorem, quantum version) Suppose ρ and σ are quantum states with $S(\rho\|\sigma) \leq c$. Then, for all $r > 1$, there are states ρ' and ρ'' such that

$$\|\rho - \rho'\|_t \leq 2/\sqrt{r}$$

and

$$\sigma = \alpha\rho' + (1-\alpha)\rho'',$$

where $\alpha = 2^{O(rc)}$.

(This has been stated here in a form that brings out the analogy with the classical statement above. In a later chapter we give a more nuanced statement (Theorem 4.1) which is better suited for our applications.)

1.4 Privacy model

Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a boolean valued function. Let $\mu_{\mathcal{X}}, \mu_{\mathcal{Y}}$ be probability distributions on $\mathcal{X}, \mathcal{Y}$, and let $\mu \triangleq \mu_{\mathcal{X}} \times \mu_{\mathcal{Y}}$ denote the product distribution on $\mathcal{X} \times \mathcal{Y}$. Let registers A, X, B, Y denote Alice's work qubits, Alice's input qubits, Bob's work qubits and Bob's input qubits respectively, at a particular point in time. Let $\rho_{XA}^{\mathcal{P}}$ denote the density matrix of Alice's qubits in protocol $\mathcal{P}$ at this point in time, when $\mathcal{P}$ is started off with distribution μ on (x, y). Now let us suppose that Bob turns malicious, and he wants to know as much as he can about Alice's input, without letting Alice realize this. Thus, Alice and Bob are now actually running a 'cheating' protocol $\widetilde{\mathcal{P}}$. Let $\rho_{XA}^{\widetilde{\mathcal{P}}}$ denote the density matrix of Alice's qubits in protocol $\mathcal{P}'$ at the same point in time. Alice does not realize the difference between $\mathcal{P}$ and $\widetilde{\mathcal{P}}$ iff $\rho_{XA}^{\mathcal{P}} = \rho_{XA}^{\widetilde{\mathcal{P}}}$. The privacy loss from Alice to Bob is captured by the mutual information $\widetilde{I}(X : BY)$ between Alice's input register X and Bob's qubits BY in $\widetilde{\mathcal{P}}$. We want to study how large $\sup \widetilde{I}(X : BY)$ can be for a given boolean valued function f, product distribution μ, and protocol $\mathcal{P}$, where the supremum is taken over all 'cheating' protocols $\widetilde{\mathcal{P}}$ that 'mimic' $\mathcal{P}$ with respect to Alice.

One of the ways that Bob can cheat without Alice being wiser is by running $\mathcal{P}$ with the superposition

$$|\mu_{\mathcal{Y}}\rangle \triangleq \sum_y \sqrt{\mu_{\mathcal{Y}}(y)}\,|y\rangle$$

fed to register Y. This method of cheating gives Bob at least as much information about Alice's input as in the 'honest' run of $\mathcal{P}$ when the mixture $\mu_{\mathcal{Y}}$ is fed to Y. Sometimes it can give much more. Consider the *index function problem*, where now Bob has a bit string y and Alice has an index i and Alice is supposed to determine the value of the ith bit y_i of y. Consider a safe protocol $\mathcal{P}$ for the index function problem in which Alice sends the index i to Bob who just sends back to Alice both i and a copy of y_i. For simplicity, assume that it is error less (an error of $1/4$ will only change the privacy losses by a multiplicative constant). Bob can cheat by feeding a uniform superposition over bit strings into his input register Y, and then running $\mathcal{P}$. Alice is honest, and she has a random index $i \in [n]$. At the end of this 'cheating' run of $\mathcal{P}$, Bob applies a Hadamard transformation on each of the registers $Y_j, 1 \leq j \leq n$. Suppose he were to measure them now in the computational basis. For all $j \neq i$, he would measure $|0\rangle$ with probability 1. For $j = i$, he would measures 1 with probability $1/2$. Thus, Bob has extracted about $\log n/2$ bits of information about Alice's index i. An 'honest' run of $\mathcal{P}$ would have yielded Bob only 1 bit of information about i. Klauck [Kla02] (based on Cleve et al. [CvDNT98]) has made a similar observation about $\Omega(n)$ privacy loss for clean protocols computing the inner product mod 2 function. The significance of our lower bounds on privacy loss is that they make *no assumptions* about the protocol $\mathcal{P}$.

Our definition of privacy loss is inspired by the above example.

Definition 1.1 (Privacy loss) *We consider a 'cheating' run of $\mathcal{P}$ when mixture $\mu_{\mathcal{X}}$ is fed to register X and superposition $|\mu_{\mathcal{Y}}\rangle$ to register Y. Let $I'(X : BY)$ denote the mutual information of* Alice*'s input register X with* Bob*'s registers BY at the end of this run of $\mathcal{P}$. The* privacy loss *of $\mathcal{P}$ for function f on the product distribution μ from* Alice *to* Bob *is defined as*

$$L^{\mathcal{P}}(f, \mu, A, B) \triangleq I'(X : BY).$$

The privacy loss from Bob *to* Alice*, $L^{\mathcal{P}}(f, \mu, B, A)$, is defined similarly. The privacy loss of $\mathcal{P}$ for f under distribution μ, $L^{\mathcal{P}}(f, \mu)$, is the larger of $L^{\mathcal{P}}(f, \mu, A, B)$ and $L^{\mathcal{P}}(f, \mu, B, A)$. The privacy loss of $\mathcal{P}$ for f, $L^{\mathcal{P}}(f)$, is the maximum over all product distributions μ, of $L^{\mathcal{P}}(f, \mu)$.*

Remarks:

1. Since we are only considering a particular class of 'cheating' protocols that 'mimic' $\mathcal{P}$ with respect to Alice, lower bounds for privacy loss proved in this model also hold for the model of general 'cheating' protocols that mimic $\mathcal{P}$ with respect to Alice.
2. Our notion of 'superpositional' privacy loss can be viewed as a quantum analogue of the "combinatorial-informational" bounded error measure of privacy loss, $I^*_{\mathrm{c-i}}$, in Bar-Yehuda et. al [BCKO93].
3. In [Kla02], Klauck defines a similar notion of privacy loss. In his definition, a mixture according to distribution μ (not necessarily a product distribution) is fed to both Alice's and Bob's input registers. He does not consider the case of superpositions being fed to input registers. For product distributions, our notion of privacy is more stringent than Klauck's, and in fact, the $L^{\mathcal{P}}(f, \mu, A, B)$ defined above is an upper bound (to within an additive factor of 1) on Klauck's privacy loss function.
4. We restrict ourselves to product distributions because we allow Bob to cheat by putting a superposition in his input register Y. He should be able to do this without any *a priori* knowledge of x, which implies that the distribution μ should be a product distribution.

We now define two more communication problems for which we will show logarithmic privacy lower bounds later on.

Definition 1.2 (Set disjointness) *In the* set disjointness *problem* DISJ_n*,* Alice *has a bit string* $x \in \{0,1\}^n$*,* Bob *has a bit string* $y \in \{0,1\}^n$*, and they want to communicate and determine the value of* $\bigvee_{i=1}^{n}(x_i \wedge y_i)$*.*

Definition 1.3 (Inner product mod 2) *In the* inner product mod 2 *problem* IP_n*,* Alice *has a bit string* $x \in \{0,1\}^n$*,* Bob *has a bit string* $y \in \{0,1\}^n$*, and they want to communicate and determine the value of* $\bigoplus_{i=1}^{n}(x_i \wedge y_i)$*.*

Klauck [Kla00] has given a lower bound for the communication complexity of bounded error one-way quantum protocols for f in terms of the VC-dimension (see Definition 2.2). We prove a lower bound for the privacy loss of bounded error safe quantum protocols for f in terms of the VC-dimension. Note that INDEX_n (defined below), DISJ_n and IP_n each have VC-dimension n for $\mathcal{X}$.

1.4.1 The index function problem

Let us recall the definition of the index function problem INDEX_n [MNSW98b, ANTV99, Nay99].

> There are two players A and B. A is given an input $x \in \{0,1\}^n$ and B is given an index $i \in [n]$. They must exchange messages so that in the end B knows x_i.

Background: Miltersen, Nisan, Safra and Wigderson [MNSW98b] considered this problem (they called it the *set membership problem*) in the classical setting, and showed that if B sends a total of at most b bits, then A must send $n/2^{O(b)}$ bits. Note that this is optimal as there is a trivial protocol where B sends the first b bits of his index to A, and A replies by sending the corresponding part of her bit string. This was one of the problems where they applied their *richness technique.* However, there is a natural round-elimination argument that gives this lower bound. Fix a protocol where B sends a total of at most b bits, perhaps spread over several rounds. Modify this protocol as follows. Let A guess all of B's messages. A sends her guesses as well as her responses to those guesses to B. Now, if B finds that A guessed all the messages correctly on his behalf, he accepts the answer given by the original protocol; otherwise, he tosses a fair coin. Thus, if the original protocol was correct with probability $1/2+\epsilon$, the new one-round protocol is correct with probability at least $\frac{1}{2}+\frac{\epsilon}{2^{O(b)}}$. A standard information theoretic argument now shows that A must send $n\epsilon^2/2^{O(b)}$ bits.

In the quantum setting, Nayak [Nay99] (see also Cleve et al. [CvDNT98]), showed that if B sends no messages at all, then A must send at least $\Omega(n)$ bits. This bound holds even if the players share EPR pairs in advance, or if A and B interact but B's messages do not depend on his input i. However, the case where B is allowed to send a few qubits based on his input in order to reduce the communication from A, does not seem to have been considered before. Nayak (private communication) observed that the classical round elimination argument described above can be applied in the quantum setting as well: if A and B share EPR pairs in advance, then using teleportation [BBC$^+$93], B's messages can be assumed to be classical. Now, A can guess B's messages, and we can combine the classical round elimination argument above with the existing results for the index function problem in the quantum setting.

In this thesis, we generalise this result to a statement of the following form: if B 'leaks' only a small number of bits of information about his input, then A must send a large number of bits. Before we present our result, let us explain what we mean when we say that B 'leaks' only a small number of bits of information about his input. Fix a protocol for the index function problem. Assume that B's input J is a random index $i \in [n]$. Suppose B operates faithfully according to the protocol, but A deviates from it and manages to get her registers R entangled with J: we say that B leaks only b bits of information about his input if the mutual information between J and R, $I(J : R)$, is at most b. This must hold for all strategies adopted by A, which have the property that the reduced density matrix of Bob's qubits is at all times the same as in the original protocol. In other words, A wants to cheat and gather a lot of information about B's input, but B should not be able to figure out that A is cheating. Note that we do not assume that B's messages contain only b qubits, they can be arbitrarily long. In the quantum setting, A has a big bag of tricks she can use in order to extract information from B (see Section 1.4 for an example of a cheating A for the index function problem).

Result 1$'$ (informal statement) If there is a protocol for the index function problem where B leaks only b bits of information about his index i, then A must send $\Omega(n/2^{O(b)})$ bits.

This result should be compared with results on private information retrieval [CKGS98]. There, one requires that the party holding the database x know nothing about the index i. Result 1$'$ generalises this notion and shows a trade-off between the loss in privacy for the the database user B and the communication cost for A.

Klauck [Kla02] recently studied privacy in quantum communication protocols. In Klauck's setting, two players collaborate to compute a function, but at any point, one of the players might decide to terminate the protocol and try to infer something about the input of the other player using the bits in his possession. The players are *honest but curious*: in a sense, they don't deviate from the protocol in any way other than, perhaps, by stopping early. In this model, Klauck shows that there is a protocol for the *set disjointness* function where neither player reveals more than $O((\log n)^2)$ bits of information about his input, whereas in every classical protocol, at least, one of the players leaks $\Omega(\sqrt{n}/\log n)$ bits of information about his input. Klauck, however, proves no *lower bounds* for privacy loss in the quantum setting. Our model of privacy is more stringent. We allow malicious players who can deviate arbitrarily from the protocol, but with the restriction that the honest player does not realize the difference. Note that this precludes the malicious player from prematurely aborting the protocol. In this model (defined formally in Section 1.4), we can strengthen Result $1'$.

Result 1 (informal statement) If there is a protocol for the index function problem where B leaks only b bits of information about his input i, then A must leak $\Omega(n/2^{O(b)})$ bits of information about her input x. (Note that this implies Result $1'$.)

Corollary (informal statement) For the index function problem, one of the players must leak $\Omega(\log n)$ bits of information about his input.

Remark: For clean quantum protocols, it is easy to see that an $\Omega(\log n)$ privacy loss is inevitable for the index function problem (for details, see Section 1.4). But it is conceivable that Alice and Bob can protect their privacy by using unclean protocols. Our lower bound makes *no assumptions* about the quantum protocol, and thus, we show that an $\Omega(\log n)$ privacy loss in inevitable for any safe quantum protocol for the index function problem. We use the substate theorem in a central fashion to arrive at the above results.

General result and other problems: The index function problem is just one of several problems where a statement like Result 1 can be proved using our technique. In fact, it follows easily that if the communication matrix of the function has VC-dimension at least k, then one of the players must leak at least $\Omega(\log k)$ bits of information about his input. In particular, this implies an $\Omega(\log n)$ loss in privacy for the set disjointness and inner product modulo 2 problems.

1.4.2 The set disjointness problem

The set disjointness problem is a very central problem in communication complexity and has a long and interesting history.

Background: In the bounded error classical setting Babai, Frankl and Simon [BFS86] showed a lower bound of $\Omega(\sqrt{n})$. This was improved to an $\Omega(n)$ lower bound by Kalyanasundaram and Schnitger [KS92]; their proof was simplified by Razborov [Raz92]. There is a straightforward protocol with $n+1$ bits of communication where Alice sends her entire input to Bob, who computes the answer and returns it to Alice. Interest in the communication complexity of several problems related to the set disjointness function has been revived recently because of their connection to showing lower bounds in the classical data stream model [AMS99, FKS02, GGI$^+$02, Ind00, GMMO00, JKS03, SS02]. One of these problems is the $\mathcal{L}_\infty$ promise problem: Alice and Bob are given inputs $X_A, X_B \in \{0, 1, \ldots, m\}^n$, with the promise that either for all

$$i \in [n], |X_A[i] - X_B[i]| \leq 1$$

or there exists an $i \in [n]$, such that

$$|X[i] - Y[i]| = m;$$

they must communicate in order to distinguish between these two types of inputs. For this problem, Saks and Sun [SS02] showed a lower bound of $\Omega(n/m^2)$ in a restricted model; their lower bound was strengthened by Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02], who obtained the same lower bound without any restrictions.

The quantum communication complexity of set disjointness was first studied by Buhrman, Cleve and Wigderson [BCW98], who showed that there is a protocol for this problem with $O(\sqrt{n}\log n)$ qubits of communication. This bound was improved to $O(\sqrt{n}c^{\log^* n})$, where c is a small constant, by Hoyer and de Wolf [HdW02], and recently to $O(\sqrt{n})$ by Aaronson and Ambainis [AA03]. By a result of Razborov [Raz02] this last bound is optimal.

Therefore, if no restrictions are imposed on the number of rounds (i.e. the number of messages) in the protocol, the upper and lower bounds on the two-party quantum communication complexity of the set disjointness function are tight up to constant factors. The best upper bound uses $O(\sqrt{n})$ rounds of communication, and from it one can derive a k-round protocol where the parties exchange a total of at most $O(n/k)$ qubits. For $k = 1$, Buhrman and de Wolf [BdW01] observed that a lower bound of $\Omega(n)$ follows from the results of Nayak [Nay99] for the index-function problem. For $k \geq 2$, Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01a] showed a lower bound of $\Omega(n^{1/k})$, but this is subsumed by Razborov's [Raz02] lower bound of $\Omega(\sqrt{n})$ which holds even if there is no restriction on the number of rounds. However, for small k, Razborov's lower bound is far from the best upper bound known, namely $O(n/k)$. Our first result implies lower bounds for the two-party bounded error k-round quantum communication complexity of set disjointness that comes closer to the upper bound of $O(n/k)$.

Result 1.1 *The two-party k-round bounded error quantum communication complexity of the set disjointness problem is $\Omega(n/k^2)$.*

In fact, this result extends for a class of *disjointness like* problems in the multi-party quantum communication model. Below we define the multi-party quantum and classical communication models and subsequently state our results..

1.5 The t-party quantum communication model

We define t-party quantum communication protocols as a natural extension of two-party quantum communication protocols defined by Yao [Yao93]. Let

$$f : \mathcal{X}_1 \times \mathcal{X}_2 \cdots \mathcal{X}_t \to \mathcal{Z}$$

be a function. There are t parties $\mathcal{P}_1, \ldots, \mathcal{P}_t$ who hold qubits. When the quantum communication protocol Π starts, $\mathcal{P}_i$ holds $|x_i\rangle$ where $x_i \in \mathcal{X}_i$ together with some ancilla qubits ('work qubits') in the state $|0\rangle$. $\mathcal{P}_1, \ldots, \mathcal{P}_t$ may also share an input independent prior entanglement pure state (say $|\psi\rangle$). Different parties possess different qubits of $|\psi\rangle$. The parties take turns to communicate to compute $f(x_1, x_2, \cdots, x_t)$. Suppose it is $\mathcal{P}_1$'s turn to communicate. $\mathcal{P}_1$ can make an arbitrary unitary transformation on the qubits in her possession at this time and then send some of her qubits to $\mathcal{P}_2, \ldots, \mathcal{P}_t$. Whose turn it is to communicate, the unitary transformation applied by the active player and the qubits that the active player sends to the other players are predetermined by Π and independent of the input $(x_1, \ldots, x_t)$. A *round* of communication denotes the qubits that the active player sends to the other players. Sending qubits does not change the overall superposition, but rather changes the ownership of the qubits. At the end of the protocol Π, one of the parties performs a von Neumann measurement in the computational basis of some qubits in her possession (the 'answer qubits') to output an answer $\Pi(x_1, x_2, \cdots, x_t)$. The party performing the measurement as well as the qubits that she measures are predetermined by Π and independent of the input $(x_1, \ldots, x_t)$. We say that protocol Π computes f with error δ if

$$\max_{x_1, \ldots, x_t} \Pr[\Pi(x_1, \ldots, x_t) \neq f(x_1, \ldots, x_t)] \leq \delta.$$

The communication cost of Π is the number of qubits exchanged in Π between all the parties. The t-party k-round δ-error quantum communication complexity of f, denoted by $Q_\delta^{t,k}(f)$, is the minimum communication cost of a t-party k-round δ-error quantum protocol with prior entanglement for f. When δ is omitted, we mean that $\delta = 1/3$.

As in the two-party model, we require that the parties make a 'safe' copy of their inputs before beginning the protocol Π. We call such protocols *safe*, and henceforth, we will assume that all our protocols are safe.

1.5.1 The t-party classical communication model

In fact, there are several ways to generalise the two-party model to the multi-party model. In this thesis, we will consider the version where there are t parties $P_1, P_2, \ldots, P_t$ with respective inputs $X_1, X_2, \ldots, X_t \subseteq [n]$. In each round of communication some party sends a message to another party. The party who receives the last message can determine the desired value $F(X_1, X_2, \ldots, X_t)$ based on his current state at that point.

We consider the class of boolean valued functions whose value depends only on $X_1 \cap \cdots \cap X_t$; that is, for each F in this class there is an $f_F : 2^{[n]} \to \{0, 1\}$, such that

$$F(X_1, \ldots, X_t) = f_F(X_1 \cap \cdots \cap X_t).$$

We call such functions F *set disjointness-like.* Define the 'monotone sensitivity of f_F' as

$$s_m(f_F) \stackrel{\Delta}{=} \max_{S \subseteq [n]} |\{i : f_F(S \cup \{i\}) \neq f_F(S)\}|.$$

Result 1.2 *The t-party k-round bounded error quantum communication complexity of a set disjointness-like function F is $\Omega(s_m(f_F)/k^2)$.*

In fact, Result 1.2 follows from the following result via easy reductions.

Result 1.2' *The t-party k-round bounded error quantum communication complexity of the promise set disjointness problem is $\Omega(n/k^2)$. This lower bound also holds for Nisan's approximate set disjointness problem [Nis02].*

Recently, because of its connection to the problem of computing *frequency moments* in the data stream model [AMS99], the following *promise set disjointness* problem has been studied. Here, the parties are required to distinguish between two types of inputs: in the first type, $X_1, X_2, \ldots, X_t$ are pairwise disjoint; in the second type, $X_1, X_2, \ldots, X_t$ have exactly one element in common but are otherwise pairwise disjoint. For this problem, Chakrabarti, Khot and Sun [CKS03] show a lower bound of $\Omega(n/(t \log t))$, improving an earlier $\Omega(n/t^2)$ lower bound of Bar-Yossef, Jayram, Kumar and Sivakumar [BJKS02] and an $\Omega(n/t^4)$ lower bound of Alon, Matias and Szegedy [AMS99]. A slight variant of this problem, called the approximate set disjointness problem, was considered by Nisan [Nis02]; the lower bounds mentioned above apply to Nisan's version as well. The multi-party quantum communication complexity of these problems has not been considered before this work.

In a related work (not included in this thesis) we get the following lower bound for the $\mathcal{L}_\infty$ promise problem.

Result 1.3 *The two-party k-round quantum communication complexity of the $\mathcal{L}_\infty$ promise problem is $\Omega(n/(k^3 m^{k+1}))$.*

Remarks:
1. Observe that the lower bound in Result 1.2' is independent of t! This appears to contradict the $O((n \log n)/t)$ upper bound for the promise set disjointness problem in [BJKS02]. However, that upper bound is in the multi-party *simultaneous message* model, whereas in our definition of multi-party quantum protocols it is required to pass messages from one party to another. Thus, the simultaneous message protocol of [BJKS02] is actually a t-round protocol in our model.
2. For two-party quantum protocols with an unbounded number of rounds, we get a lower bound of $\Omega(n^{1/3})$ for the set disjointness problem.
3. All our lower bounds hold even if the parties start with arbitrary prior entanglement that is independent of the inputs.
4. Finally, we remark that our quantum communication complexity lower bounds imply space lower bounds for a natural model of 'quantum data stream computation', in exactly the same way as in the classical setting.

1.6 Organisation of the thesis

In Chapter 2 we give some information theoretic facts and definitions which will be used in later chapters. In the following chapter (Chapter 3), we prove the direct sum results for the simultaneous message and the two-party multiple round protocols. In Chapter 4 we prove the information theoretic result, which we believe is an important contribution of this work, and then use it to prove the privacy results. In Chapter 5 we present the results for both the full version and the bit version of the pointer chasing problem. The following chapter (Chapter 6) contains the proofs of the results on the set disjointness and related problems. Finally we conclude with Chapter 7 stating our conclusions and a few open problems.

Chapter 2

Information theory

In this chapter we give some definitions and state a few facts which will be used in subsequent chapters.

2.1 Classical information theory

In this thesis, 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 $P, Q : [k] \rightarrow \mathbb{R}$. The *total variation distance* (aka ℓ_1-*distance*) between P, Q is defined as

$$\|P - Q\|_t \stackrel{\Delta}{=} \sum_{i \in [k]} |P(i) - Q(i)|.$$

We say $P \leq Q$ iff $P(i) \leq Q(i)$ for all $i \in [k]$. Let X be a random variable. The *Shannon entropy* of X is defined as

$$H(X) \stackrel{\Delta}{=} -\sum_{x} \Pr[X = x] \log \Pr[X = x].$$

Suppose X, Y, Z are random variables with some joint distribution. The mutual information of X and Y is defined as

$$I(X : Y) \stackrel{\Delta}{=} H(X) + H(Y) - H(XY).$$

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 $I((X : Y) \mid Z) \stackrel{\Delta}{=} \mathrm{E}_Z\, I((X : Y) \mid Z = z)$. 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].

The following fact follows easily from the definitions.

Fact 2.1 *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) $I(XY : Z \mid W) \geq I(XY : Z) - H(W)$.

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

Definition 2.1 (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 fact follows easily from the definitions.

Fact 2.2 *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*

$$I(X : M) = \mathop{\mathrm{E}}_{X}[S(P_X\|P)],$$

where the expectation is taken by choosing X according to its marginal distribution.

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 2.3 (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,*

$$S(D_P\|D_Q) \leq S(P\|Q).$$

We recall here the definition of the Vapnik-Chervonenkis dimension (VC-dimension). Some of our results are stated in terms of VC-dimension.

Definition 2.2 (VC-dimension) *For a boolean valued function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, a set $S \subseteq \mathcal{Y}$ is* shattered*, if for all $R \subseteq S$ there is an $x \in \mathcal{X}$ such that*

$$\forall y \in S : f(x, y) = 1 \Leftrightarrow y \in R.$$

The VC-dimension *of f for $\mathcal{X}$, $\mathrm{VC}_{\mathcal{X}}(f)$, is the largest size of such a shattered set $S \subseteq \mathcal{Y}$. $\mathrm{VC}_{\mathcal{Y}}(f)$ is defined analogously.*

2.2 Quantum information theory

We now recall some basic definitions and facts from quantum information theory, which will be useful in stating and proving our main results. For excellent introduction to quantum information theory, see Nielsen and Chuang [NC00].

In this thesis all quantum systems are finite dimensional. We will use the notation $A \geq B$ for Hermitian operators A, B in the same finite dimensional Hilbert space $\mathbf{H}$ as a shorthand for the statement '$A - B$ is positive semidefinite'. Thus, $A \geq 0$ denotes that A is positive semidefinite. 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 density matrix ρ, then

$$S(A) \stackrel{\Delta}{=} S(\rho) \stackrel{\Delta}{=} -\mathrm{Tr}\ \rho \log \rho$$

is the *von Neumann entropy* of A.

Mutual information

Let A, B, C be three disjoint quantum systems. The *mutual information* of A and B is defined as

$$I(A : B) \stackrel{\Delta}{=} S(A) + S(B) - S(AB).$$

The *conditional von Neumann mutual information* of A and B given C is defined as

$$I((A : B) \mid C) \stackrel{\Delta}{=} S(AC) + S(BC) - S(C) - S(ABC).$$

If C is a classical random variable taking the classical value $|c\rangle$ with probability p_c, it is easy to see that

$$I((A : B) \mid C) = \sum_c p_c I(A^c : B^c),$$

where $(AB)^c$ denotes the joint density matrix of A and B when $C = |c\rangle$.
We also write $I((A : B) \mid C = c)$ for $I(A^c : B^c)$. Mutual information satisfies the following *monotonicity* property:

$$I(A : BC) \geq I(A : B).$$

The following facts are standard, and can be found in some form or the other in Nielsen and Chuang's book [NC00, Chapters 11, 12].

Fact 2.4 *Let $X_1, \ldots, X_n$ be classical random variables and let M be a quantum encoding of $X \stackrel{\Delta}{=} X_1 \ldots X_n$. Then,*

$$I(X : M) \geq \sum_{i=1}^{n} I(X_i : M) - \sum_{i=1}^{n} H(X_i) + H(X),$$

which implies that if $X_1, \ldots, X_n$ are independent random variables then,

$$I(X : M) \geq \sum_{i=1}^{n} I(X_i : M).$$

Also, if M is n qubits long, then $I(X : M) \leq n$.

Fact 2.5 (Monotonicity of mutual information) *Let A, B be two disjoint quantum systems and let $I(A : B)$ denote their mutual information. Consider completely positive trace preserving superoperators $\mathcal{F}_1, \mathcal{F}_2$ acting on A, B respectively. Let $I'(A : B)$ denote the mutual information after their action. Then,*

$$I(A : B) \geq I'(A : B).$$

Fact 2.6 (Fano's inequality) *Let X, Y be classical boolean random variables. Let*

$$\Pr[X = Y] = 1/2 + \delta,$$

where $-1/2 \leq \delta \leq 1/2$. Suppose $\Pr[X = 0] = \Pr[X = 1] = 1/2$. Then,

$$I(X : Y) \geq 1 - H(1/2 + \delta) \geq \delta^2.$$

The following fact can be implicitly found in Cleve et al [CvDNT98].

Fact 2.7 *Let Alice have a classical random variable X. Suppose Alice and Bob share a prior entanglement independent of X. Initially Bob's qubits have no information about X. Now let Alice and Bob run a quantum communication protocol, at the end of which Bob's qubits possess m bits of information about X. Then, Alice has to totally send at least $m/2$ qubits to Bob.*

POVM

Definition 2.3 (POVM element) *Let $\mathbf{H}$ be a finite dimensional Hilbert space. A POVM (positive operator valued measure) element F on $\mathbf{H}$ is a positive semidefinite operator on $\mathbf{H}$ such that $F \leq I$, where I is the identity operator on $\mathbf{H}$.*

If ρ is a density matrix in $\mathbf{H}$, the success probability of the mixed state ρ under POVM element F is $\mathrm{Tr}\ (F\rho)$.

Definition 2.4 (POVM) *Let $\mathbf{H}$ be a finite dimensional Hilbert space. A POVM $\mathcal{F}$ on $\mathbf{H}$ is a set of POVM elements $\{F_1, \ldots, F_k\}$ on $\mathbf{H}$ such that*

$$\sum_{i=1}^{k} F_i = I,$$

where I is the identity operator on $\mathbf{H}$.

If ρ is a density matrix in $\mathbf{H}$, $\mathcal{F}\rho$ denotes the probability distribution $\{p_1, \ldots, p_k\}$ on $[k]$, where $p_i \stackrel{\Delta}{=} \mathrm{Tr}\ (F_i\rho)$.

Trace distance

The trace norm of a linear operator A is defined as

$$\|A\|_t \triangleq \operatorname{Tr} \sqrt{A^\dagger A}.$$

The trace distance between two linear operators A, B is defined as $\|A - B\|_t$. The importance of trace distance as a metric on density matrices stems from the following fact.

Fact 2.8 (see [AKN98]) *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Let $\mathcal{F}$ be a measurement (POVM) on $\mathbf{H}$. Then,*

$$\|\mathcal{F}\rho - \mathcal{F}\sigma\|_t \leq \|\rho - \sigma\|_t .$$

Suppose A, B are disjoint quantum systems. Let ρ_{AB}, σ_{AB} be two density matrices of the joint quantum system AB. The trace distance satisfies the following property of *monotonicity*:

$$\|\rho_{AB} - \sigma_{AB}\|_t \geq \|\rho_A - \sigma_A\|_t .$$

In fact, Fact 2.8 can be derived from the monotonicity of trace distance.

Relative entropy

If ρ, σ are density matrices in the same Hilbert space, their *relative entropy* is defined as

$$S(\rho\|\sigma) \triangleq \operatorname{Tr} \left(\rho(\log \rho - \log \sigma)\right).$$

The following fact lists some useful properties of the relative entropy function. Proofs can be found in [NC00, Chapter 11]. The monotonicity property below is also called *Lindblad-Uhlmann monotonicity.*

Fact 2.9 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Then*

1. *$S(\rho\|\sigma) \geq 0$, with equality iff $\rho = \sigma$.*

2. *$S(\rho\|\sigma) < +\infty$ iff $\operatorname{supp}(\rho) \subseteq \operatorname{supp}(\sigma)$. Here $\operatorname{supp}(\rho)$ denotes the support of ρ i.e. the span of the eigenvectors corresponding to non-zero eigenvalues of ρ.*

3. *$S(\cdot\|\cdot)$ is continuous in its two arguments when it is not infinite.*

4. *(Joint convexity) Let $\rho_1, \rho_2, \sigma_1, \sigma_2$ be density matrices in $\mathbf{H}$. Let $\rho \triangleq \lambda\rho_1 + (1-\lambda)\rho_2$ and $\sigma \triangleq \lambda\sigma_1 + (1-\lambda)\sigma_2$, where $0 \leq \lambda \leq 1$. Then*

$$S(\rho\|\sigma) \leq \lambda S(\rho_1\|\sigma_1) + (1-\lambda) S(\rho_2\|\sigma_2).$$

5. *(Unitary invariance) If U is a unitary transformation on $\mathbf{H}$,*

$$S(U\rho U^\dagger\|U\sigma U^\dagger) = S(\rho\|\sigma).$$

6. *(Monotonicity) Suppose $\mathcal{K}$ is a finite dimensional Hilbert space, and ρ', σ' are density matrices in $\mathbf{H} \otimes \mathcal{K}$ such that $\mathrm{Tr}_{\mathcal{K}}\ \rho' = \rho$ and $\mathrm{Tr}_{\mathcal{K}}\ \sigma' = \sigma$. Then,*

$$S(\rho' \| \sigma') \geq S(\rho \| \sigma).$$

This implies, via unitary invariance and the Kraus representation theorem, that if $\mathcal{F}$ is a completely positive trace preserving superoperator, then

$$S(\mathcal{F}\rho \| \mathcal{F}\sigma) \leq S(\rho \| \sigma).$$

The following fact shows the connection between mutual information and relative entropy and is easy to derive.

Fact 2.10 *Let X be a classical random variable and M be a quantum encoding of X. Let X take the values $1, \ldots, l$ with probabilities $p_1, \ldots, p_l$ and let $\sigma_1, \ldots, \sigma_l$ be the respective density matrices of M. Let*

$$\sigma \triangleq \sum_{j=1}^{l} p_j \sigma_j$$

be the average density matrix of M. Then,

$$I(X : M) = \sum_{j=1}^{l} p_j S(\sigma_j \| \sigma).$$

Fidelity

Definition 2.5 (Fidelity) *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Their fidelity is defined as*

$$B(\rho, \sigma) \triangleq \sup_{\mathcal{K}, |\psi\rangle, |\phi\rangle} |\langle \psi | \phi \rangle|,$$

where $\mathcal{K}$ ranges over all finite dimensional Hilbert spaces and $|\psi\rangle, |\phi\rangle$ range over all purifications of ρ, σ respectively in $\mathbf{H} \otimes \mathcal{K}$.

The fidelity (or sometimes its square) is also known as the Bhattacharya distinguishability coefficient or the "transition probability" of Uhlmann.

Jozsa [Joz94] gave an elementary proof for finite dimensional Hilbert spaces of the following basic and remarkable property about fidelity.

Fact 2.11 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Then for any finite dimensional Hilbert space $\mathcal{K}$ such that $\dim(\mathcal{K}) \geq \dim(\mathbf{H})$, there exist purifications $|\psi\rangle, |\phi\rangle$ of ρ, σ in $\mathbf{H} \otimes \mathcal{K}$, such that*

$$B(\rho, \sigma) = |\langle \psi | \phi \rangle|.$$

Also,

$$B(\rho, \sigma) = \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_t.$$

We will also need the following result about fidelity, proved by Fuchs and Caves [FC95].

Fact 2.12 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Then*

$$B(\rho, \sigma) = \inf_{F_1, \ldots, F_k} \sum_{i=1}^{k} \sqrt{\mathrm{Tr}\ (F_i \rho)\ \mathrm{Tr}\ (F_i \sigma)},$$

where $\{F_1, \ldots, F_k\}$ ranges over POVMs on $\mathbf{H}$. In fact, the infimum above can be attained by a complete orthogonal measurement on $\mathbf{H}$.

The following relation is known between fidelity and trace distance between two density matrices [NC00].

Fact 2.13 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Then*

$$2(1 - B(\rho, \sigma)) \leq \|\rho - \sigma\|_t \leq 2\sqrt{1 - B(\rho, \sigma)^2}.$$

2.3 Miscellaneous

2.3.1 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 *predicate* or 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 2.14

(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,

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

2.3.2 A minimax theorem

We will require the following minimax theorem from game theory, which is a consequence of the Kakutani fixed point theorem in real analysis.

Fact 2.15 *Let A_1, A_2 be non-empty, convex and compact subsets of $\mathbb{R}^n$ for some n. Let $u : A_1 \times A_2 \to \mathbb{R}$ be a continuous function, such that*

- *$\forall a_2 \in A_2$, the set $\{a_1 \in A_1 : \forall a_1' \in A_1\, u(a_1, a_2) \geq u(a_1', a_2)\}$ is convex; and*
- *$\forall a_1 \in A_1$, the set $\{a_2 \in A_2 : \forall a_2' \in A_2\, u(a_1, a_2) \leq u(a_1, a_2')\}$ is convex.*

Then, there is an $(a_1^*, a_2^*) \in A_1 \times A_2$ such that

$$\max_{a_1 \in A_1} \min_{a_2 \in A_2} u(a_1, a_2) = u(a_1^*, a_2^*) = \min_{a_2 \in A_2} \max_{a_1 \in A_1} u(a_1, a_2).$$

Remark: The above statement follows by combining Proposition 20.3 (which shows the existence of Nash equilibrium a^* in strategic games) and Proposition 22.2 (which connects Nash equilibrium and the min-max theorem for games defined using a pay-off function such as u) of Osborne and Rubinstein's [OR94, pages 19–22] book on game theory.

Chapter 3

The direct sum problem

In this chapter we present our compression and direct sum results for the simultaneous message (Section 3.2) and two-party communication complexity (Section 3.3) models. The proof of these results follows the information theoretic framework of Chakrabarti et al. [CSWY01]. Let the k-round information complexity of f under distribution μ for the inputs, denoted by $\mathrm{IC}^k_{\mu,\delta}(f)$, be the minimum over all δ-error k-round private coin protocols of the mutual information between the random inputs of the players and the complete message transcript they generate. It can be easily seen that

$$R^k_{\mu^m,\delta}(f^m) \geq \mathrm{IC}^k_{\mu^m,\delta}(f^m).$$

In the rest of the proof we show a lower bound for $\mathrm{IC}^k_{\mu^m,\delta}(f^m)$. This argument has two parts. The first part is purely information theoretic, and is straightforward. Define the probability distribution μ^m on $\mathcal{X}^m \times \mathcal{Y}^m$ as

$$\mu(\langle x_1, \ldots, x_m \rangle, \langle y_1, \ldots, y_m \rangle) \stackrel{\Delta}{=} \mu(x_1, y_1) \cdot \mu(x_2, y_2) \cdots \mu(x_m, y_m).$$

Then it is easy to see that, for any product distribution μ,

$$IC^k_{\mu^m,\delta}(f^m) \geq m \cdot IC^k_{\mu,\delta}(f).$$

The second, more technical, part is to relate $IC^k_{\mu,\delta}(f)$ and $C^k_{\mu,\delta}(f)$. To prove such a result in the case $k = 1$ (and also for the simultaneous message model) Chakrabarti et al. [CSWY01] employed an interesting message compression result. Informally, it states that if the message contains at most a bits of information about the player's input, then one can modify the (one round or simultaneous message) protocol so that the *length* of the message is $O(a + \log n)$, where n is the total input size of both the players.

Comparison with previous work Chakrabarti et al. [CSWY01] used an ad hoc smoothening and sampling argument to show this compression result. Our approach on the other hand is more information theoretic. In particular, we make use of relative entropy of probability distributions and substate theorem. We know from the previous chapter that

$$I(X : M) = \mathop{\mathrm{E}}_{X}[S(P_x \| P)],$$

where P is the distribution of M and P_x is the distribution of M when we condition on the event $X = x$, and the expectation is over the (marginal) distribution of X. Thus, if $I(X : M) \leq a$, then typically the value of $S(P_x \| P)$ is at most a. For simplicity, let us consider for now the bounded error private coin simultaneous message model. Alice and Bob have to compute a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ in this model. Let X, M denote the random variables corresponding to Alice's input and her message respectively. Let $T_y^x(m)$ denote the correctness predicate (of the protocol when Alice's input is x, Bob's input is y, and Alice sends the message m to the referee. The main advantage of the substate theorem is that we can compress Alice's message for each $x \in \{0,1\}^n$ separately. We now indicate how this is done. Ideally on input x, Alice would like to pick messages according to the distribution P_x. It is easy to see via Chernoff-Hoeffding bounds that, for every $x \in \{0,1\}^n$, there exists a set S^x of $O(n)$ messages such that on input x, if Alice sends a message from S^x uniformly at random, the predicates T_y^x for all $y \in \{0,1\}^n$ will be satisfied with high probability. In fact, a random sample of $O(n)$ elements according to P_x serves as a suitable S^x with exponentially high probability. However, Alice needs to pick these sets S^x in such a way that their union is small, so that she and the referee can settle on a common succinct encoding for the messages. The substate theorem allows Alice to pick such correlated sets S^x. Informally stated, it allows her to argue, via a *rejection sampling* argument, that if she picks a sample of messages according to the distribution P, then one in every $2^{O(a)}$ messages can serve as a message sampled according to $\widetilde{P}_x$, where $\widetilde{P}_x$ is a distribution close to P_x in the total variation distance. Thus, if we pick a sample of size $n \cdot 2^{O(a)}$ according to P, then for most x we can get a sub sample of $O(n)$ elements according to $\widetilde{P}_x$. Since $\widetilde{P}_x$ is close to P_x, the $O(n)$-sized sub sample still serves as a suitable S^x with exponentially high probability. In particular, we can prove the following (see Lemma 3.2).

Theorem Let X and M be random variables (with some joint distribution), where X is uniformly distributed over $\{0,1\}^n$ and their mutual information $I(X : M) \leq a$. Let $[m]$ be the range of M. Let $S_y^x, x, y \in \{0,1\}^n$ be randomised predicates from $[m]$ to $[0,1]$. Then, there exists a random variable M' (correlated with X) such that

(a) M' takes values in a set of size $n \cdot 2^{O(a/\epsilon)}$;

(b) There exists $A \subseteq \{0,1\}^n$ of size at least $\frac{2}{3} \cdot 2^n$ such that for all $x \in A$ and $y \in \{0,1\}^n$,

$$|\Pr[S_y^x(M') \mid X = x] - \Pr[S_y^x(M) \mid X = x]| \leq \epsilon.$$

In our context, the above theorem states that if Alice's message contains only a bits of information about her input, she can compress it to $O(a + \log n)$ bits without changing the error probability of the protocol significantly. A similar message compression argument holds for Bob too. This gives us an alternative proof of the main result of Chakrabarti et al. [CSWY01], with better dependence on the parameters.

3.1 The rejection sampling argument

In this section we prove a few lemmas which build the rejection sampling argument. These lemmas are then used to prove the compression and direct sum results in the later sections.

Lemma 3.1 *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 choose $\mathbf{X}$ and $\mathbf{Y}$ using a standard rejection sampling idea (see e.g. [Ros97, Chapter 4, Section 4.4]): first pick the sequence $\mathbf{X}$; now if $X_i = \ell$ retain it in $\mathbf{Y}$ with probability

$$P(\ell)2^{-a}/Q(\ell).$$

We now give the formal details. Let (X, χ) be a pair of random variables taking values in $[k] \times \{0,1\}$, whose joint distribution is given by

$$\Pr[X = i] = Q(i) \quad \text{and} \quad \Pr[\chi = 1 \mid X_i = \ell] = P(\ell)2^{-a}/Q(\ell).$$

Now, pick t pairs of random variables $(X_i, \chi_i), i \in [t]$, where each (X_i, χ_i) has the same distribution as (X, χ) defined above, and (X_i, χ_i) is independent of $(X_j, \chi_j), j \neq i$. This specifies the joint distribution of the random variables $\mathbf{X} \triangleq \langle X_1, X_2, \ldots, X_t \rangle$ and $\chi \triangleq \langle \chi_1, \chi_2, \ldots, \chi_t \rangle$. Now, let $\mathbf{Y}$ be the subsequence $\langle X_i : \chi(i) = 1 \rangle$. That requirements (a) and (d) hold for $(\mathbf{X}, \mathbf{Y})$ constructed in this manner are immediate from our definition. For part (b), note that

$$R = \sum_{i=1}^{t} \chi_i,$$

where the χ_i are independent random variables, and

$$\begin{aligned} \Pr[\chi_i = 1] &= \sum_{\ell \in [k]} \Pr[X_i = \ell] \cdot \Pr[\chi_i = 1 \mid X_i = \ell] \\ &= \sum_{\ell \in [k]} Q(\ell) \cdot \left(\frac{P(\ell)2^{-a}}{Q(\ell)} \right) = 2^{-a}. \end{aligned}$$

We first show part (c) by conditioning on the event "$\chi = \sigma$" for an arbitrary $\sigma \in \{0,1\}^t$ such that

$$\sum_i \sigma_i = r.$$

Since the event $R = r$ is the disjoint union of events of the form $\chi = \sigma$, this implies (c) in general. For $j \in [r]$, let h_j be the position of the jth 1 in σ, that is,

$$h_j \triangleq \min\{h : \sum_{i=1}^{h} \sigma_i = j\}.$$

Then, under the condition $\chi = \sigma$,

$$\mathbf{Y} = \langle X_{h_1}, X_{h_2}, \ldots, X_{h_r} \rangle.$$

Since (X_i, χ_i) are independent for different i, we conclude that the Y_j's are independent under the condition $\chi = \sigma$. We only need to show that Y_j has the right distribution:

$$\begin{aligned}
\Pr[Y_j = \ell \mid \chi = \sigma] &= \Pr[X_{h_j} = \ell \mid \chi_{h_j} = 1] \\
&= \frac{\Pr[X_{h_j} = \ell \wedge \chi_{h_j} = 1]}{\Pr[\chi_{h_j} = 1]} \\
&= \frac{Q(\ell) \cdot P(\ell) 2^{-a} / Q(\ell)}{2^{-a}} \\
&= P(\ell).
\end{aligned}$$

■

Lemma 3.2 *Let Q and $P_1, P_2, \ldots, P_N$ be probability distributions on $[k]$. Define $a_i \triangleq S(P_i \| Q)$. Suppose $a_i < \infty$ for all $i \in [N]$. Let $S_1, S_2, \ldots, S_N$ be randomised predicates on $[k]$. Define*

$$p_{ij} \triangleq \Pr_{y \in_{P_i} [k]} [S_j(y)].$$

Fix $\epsilon \in (0,1]$. Then, there exists a sequence $\mathbf{x} \triangleq \langle x_1, \ldots, x_t \rangle$ 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

$$t_i \triangleq \left(\lceil \frac{8 \cdot 2^{(a_i+1)/\epsilon} \cdot \log(2N)}{(1-\epsilon)\epsilon^2} \rceil \right).$$

(b) For $i, j = 1, 2, \ldots, N$,

$$| \mathop{\mathrm{E}}_{\ell \in_U [R_i]} [S_j(\mathbf{y}^i[\ell])] - p_{ij} | \leq 2\epsilon,$$

where R_i is the length of $\mathbf{Y}^i$.

(c) $t \triangleq \max_i t_i$.

Proof: Using part (b) of Theorem 4.1, we obtain distributions $\widetilde{P}_i$ such that

$$\left\| P_i - \widetilde{P}_i \right\|_t \leq 2\epsilon$$

and

$$(1-\epsilon)2^{-(a_i+1)/\epsilon}\widetilde{P}_i \leq Q.$$

Using Lemma 3.1, 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 $t \triangleq \max_i t_i$ 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 2.14, we conclude that

$$\Pr[\exists i, R_i < \left(\frac{4}{\epsilon^2}\right)\log(2N)] < N \cdot \frac{1}{2N} = \frac{1}{2}. \tag{3.1}$$

Now, condition on the event

$$R_i \geq \left(\frac{4}{\epsilon^2}\right)\log(2N),$$

for all $1 \leq i \leq N$. Define

$$\widetilde{p}_{ij} \triangleq \Pr_{y \in_{\widetilde{P}_i}[k]}[S_j(y)].$$

We use part (a) of Fact 2.14 to conclude that for $i, j = 1, 2, \ldots, N$,

$$\Pr_{\mathbf{Y}^i}[| \mathop{\mathrm{E}}_{\ell \in_U [r_i]}[S_j(\mathbf{Y}^i[\ell])] - \widetilde{p}_{ij}| > \epsilon] < \frac{2}{(2N)^8},$$

implying,

$$\Pr_{\mathbf{Y}^1,\ldots,\mathbf{Y}^N}[\exists i, j, \mid \mathop{\mathrm{E}}_{\ell \in_U [r_i]}[S_j(\mathbf{Y}^i[l])] - \widetilde{p}_{ij}| > \epsilon] < \frac{1}{2}. \tag{3.2}$$

From (3.1) and (3.2) and the fact that

$$\forall i, j \quad |p_{ij} - \widetilde{p}_{ij}| \leq \epsilon$$

(since $\left\| P_i - \widetilde{P}_i \right\|_t \leq 2\epsilon$), 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. ■

Lemma 3.3 *Let P and Q be probability distributions on $[k]$, such that*

$$\mathsf{Good} \triangleq \{i \in [k] : \frac{P(i)}{2^a} \leq Q(i)\}$$

has probability exactly $1-\epsilon$ in P. Then, there exist correlated random variables $\mathbf{X} \triangleq \langle X_i \rangle_{i \in \mathbb{N}_+}$, 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

$$\beta \triangleq \epsilon 2^{-a}/(1-(1-\epsilon)2^{-a})$$

and

$$\gamma_i \triangleq P'(i)2^{-a}/Q(i).$$

The joint probability distribution of X and Z is given by

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

Note that this implies that

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

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

$$R \triangleq \min\{i : Z_i \neq \star\};$$

$$R \triangleq \infty \ \text{ if } \{i : Z_i \neq \star\} \text{ is the empty set.}$$

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,

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

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

■

Lemma 3.4 *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 $\ell \in [k] - \mathsf{Good}_j$,

$$\Pr[Y_j = \ell] = 0 \quad \textit{and} \quad \Pr[Y_j = 0] = 1 - P_j(\mathsf{Good}_j) \leq \epsilon;$$

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

Proof: Using Theorem 4.1, 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 for all $i \in \mathsf{Good}_j$

$$P_j(i)2^{-(a_j+1)/\epsilon} \leq Q(i).$$

Now from Lemma 3.3, 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. ■

3.2 Simultaneous message protocols

In this section we prove the compression and direct sum result for the simultaneous message model in communication complexity.

Theorem 3.1 (Compression result, simultaneous messages) *Suppose 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*

$$I(X : M_A) \leq a \quad \text{and} \quad I(Y : M_B) \leq b.$$

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 $\frac{3a}{\epsilon} + \log(n+1) + \log \frac{1}{\epsilon^2(1-\epsilon)} + \frac{1}{\epsilon} + 4$ bits and Bob sends messages of length at most $\frac{3b}{\epsilon} + \log(n+1) + \log \frac{1}{\epsilon^2(1-\epsilon)} + \frac{1}{\epsilon} + 4$ 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 2.2, we have

$$\mathop{\mathrm{E}}_{X}[S(P_x \| P)] \leq a,$$

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

$$t_a \triangleq \frac{8(n+1)2^{(3a+1)/\epsilon}}{\epsilon^2(1-\epsilon)}.$$

From Lemma 3.2, 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

$$t_b \triangleq \frac{8(n+1)2^{(3b+1)/\epsilon}}{\epsilon^2(1-\epsilon)}.$$

In Π', the error for an input $(x,y) \in \mathsf{Good}_A \times \mathsf{Good}_B$ is at most $\delta + 4\epsilon$. ■

The following corollary is immediate from Theorem 3.1.

Corollary 3.1 *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 main result of Chakrabarti et al. [CSWY01].

Theorem 3.2 (Direct sum, simultaneous messages) *Let $\delta, \epsilon > 0$. Let $f : \{0,1\}^n \times \{0,1\}^n \to \mathcal{Z}$ be a function. Define*

$$\tilde{R}^{\mathrm{sim}}_{\delta}(f) \triangleq \min_{f'} R^{\mathrm{sim}}_{\delta}(f'),$$

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 6.2, Fact 6.1 and Corollary 3.1. ■

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].
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.

3.3 Two-party multiple round protocols

In this section we prove the compression result and the direct sum result for the two-party multiple round protocols.

Theorem 3.3 (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 T denote the complete transcript of messages sent by Alice and Bob. Suppose $I(XY : T) \leq a$. 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 $\Pi'_k, \ldots, \Pi'_1$. 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 $(k-1)$st 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 2.1,

$$\begin{aligned} I(XY : M) &= I(XY : M_1) + \mathop{\mathrm{E}}_{M_1}[I((XY : M_2) \mid M_1 = m_1)] \\ &= I(XY : M_1) + \mathop{\mathrm{E}}_{M_1XY}[S(M_2^{xym_1} \| M_2^{m_1})], \end{aligned}$$

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

$$a_i \stackrel{\Delta}{=} \mathop{\mathrm{E}}_{M_1XY}[S(M_2^{xym_1} \| M_2^{m_1})].$$

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 $\mathbf{X}^{m_1} \triangleq \langle x_i^{m_1} \rangle_{i \in \mathbb{N}_+}$, where $(x_i^{m_1} : i \in \mathbb{N}_+)$ 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 $\mathbf{X}^{m_1}$. Using Lemma 3.4, 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). 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 $i+1$st to kth messages does not increase as compared to the expected length (averaged over the public coins of Π'_{i+1}) of the $i+1$st 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 $\frac{\epsilon}{k}$ 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 $\frac{\epsilon}{k}$.

By Fact 2.1,

$$\sum_{i=i}^{k} a_i = I(XY : T) \leq a,$$

where $I(XY : T)$ is the mutual information in the original protocol Π. This is because the quantity $\mathrm{E}_{M_1XY}[S(M_2^{xym_1} \| M_2^{m_1})]$ 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 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 is at most $\delta+2\epsilon$. ■

The following corollary is immediate from Theorem 3.3.

Corollary 3.2 *Let $f:\mathcal{X}\times\mathcal{Y}\rightarrow\mathcal{Z}$ be a function. Let μ be a probability 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 3.4 (Direct sum, k-rounds) *Let m,k be positive integers, and $\epsilon,\delta>0$. Let $f:\mathcal{X}\times\mathcal{Y}\rightarrow\mathcal{Z}$ 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 6.2, Fact 6.1 and Corollary 3.2. ■

Corollary 3.3 *Let m,k be positive integers, and $\epsilon,\delta>0$. Let $f:\mathcal{X}\times\mathcal{Y}\rightarrow\mathcal{Z}$ 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 chasing problem with the 'wrong' player starting (both the bit version and the full pointer version), since the product distributional complexity (in fact, for the uniform distribution) of pointer chasing is the same as its randomised complexity [NW93, PRV01b].

3.4 Impossibility of quantum compression

In this section, we show that the information cost based message compression approach does not work in the quantum setting.

3.5 Sampling uniformly random orthonormal sets

To prove our result about the incompressibility of quantum information (section 3.4) we need to define the notion of a uniformly random set of size d of orthonormal vectors from $\mathbb{C}^m$.

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

$$x_i \triangleq y_i / \sqrt{y_1^2 + \cdots + y_m^2}$$

(note that any $y_i = 0$ with zero probability).

What follows is a formal justification of why the above mentioned method does indeed generate a uniformly random unit vector in $\mathbb{R}^m$. Reader can skip it if she is already convinced.

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

$$V \triangleq \mathrm{Span}(v_1, \ldots, v_d),$$

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 $\mathbb{C}^m$ and $\mathbb{R}^{2m}$ in mind.

Now it is easily seen that the resulting distribution on unit vectors generated by the above mentioned method 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 3.1

(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

$$w'' \triangleq y - \langle x|y\rangle x,$$

and set $w \triangleq x$ and $w' \triangleq \frac{w''}{\|w''\|}$ (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

$$\widehat{x} \triangleq \frac{Px}{\|Px\|}, \qquad \widehat{y} \triangleq \frac{Py}{\|Py\|},$$

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 3.2 ([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 3.3 *Let $f : \mathbf{U}_{m,1} \to \mathbb{R}$ be 1-Lipschitz. Then for all $0 \leq t \leq 1$,*

$$\Pr[f > \mathrm{E}[f] + t + 12/\sqrt{2m}] \leq 2\exp(-t^2 m).$$

The following definitions will be needed in our proof of the incompressibility result.

Definition 3.1 *Consider a quantum system with Hilbert space $\mathbb{C}^m$. For a POVM element M over $\mathbb{C}^m$ and a subspace W of $\mathbb{C}^m$, define*

$$M(W) \triangleq \max_{w \in W : \|w\|=1} \langle w|M|w\rangle w.$$

For subspaces W, W' of $\mathbb{C}^m$, define

$$\Delta(W, W') \triangleq \max_{M} |M(W) - M(W')|,$$

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.

The following fact can be proved from the results in [AKN98] and will be needed later.

Fact 3.4 *Let M be a POVM element over $\mathbb{C}^m$ and $w, \widehat{w} \in \mathbb{C}^m$ be unit vectors. Then*

$$|\langle w|M|w\rangle w - \langle \widehat{w}|M|\widehat{w}\rangle \widehat{w}| \leq \|w - \widehat{w}\|.$$

We first need a few lemmas.

Lemma 3.5 *Fix positive integers d, m and $\epsilon > 0$. Then there is a set $\mathcal{S}$ of at most d-dimensional subspaces of $\mathbb{C}^m$ such that*

(a) $|\mathcal{S}| \leq \exp(O(md \log(d/\epsilon)))$.

(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 3.2. For a unit vector $v \in \mathbb{C}^m$, let $\widetilde{v}$ denote the vector in $\mathcal{N}$ closest to it. 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 $\sum_{i=1}^{t} |\alpha_i|^2 = 1$. Define

$$w' \triangleq \sum_{i=1}^{d} \alpha_i \widetilde{w}_i$$

and

$$\widehat{w} \triangleq \frac{w'}{\|w'\|}.$$

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

$$\delta \triangleq \frac{\epsilon}{2\sqrt{d}}.$$

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 3.4 and (c) above, $\Delta(W, \widehat{W}) \leq \epsilon$. Define

$$\mathcal{S} \triangleq \{\widehat{W} : W \text{ subspace of } \mathbb{C}^m \text{ of dimension } d\}.$$

$\mathcal{S}$ satisfies Part (b) of the present lemma. Also,

$$|\mathcal{S}| \leq (4/\delta)^{2md} \leq \exp(O(md \log(d/\epsilon))),$$

proving Part (a) of the present lemma. ■

We next prove the following two propositions using Fact 3.3.

Proposition 3.1 *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,*

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

$$\Pr[\|Px\| \geq 2/\sqrt{l}] \leq 2\exp\left(-\frac{m}{100l}\right), x = w, w'.$$

and

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

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 $1/\sqrt{m}$ by symmetry. By Fact 3.3,

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

proving the first inequality of the present proposition.

The argument for the second inequality is similar. By symmetry,

$$\mathrm{E}[\|Pw\|] = \mathrm{E}[\|Pw'\|] = 1/\sqrt{l}.$$

Since the map $w \mapsto \|Pw\|$ is 1-Lipschitz, by Fact 3.3 we get that

$$\begin{aligned}\Pr[\|Px\| \geq 2/\sqrt{l}] &\leq \Pr[\|Px\| > 1/\sqrt{l} + 12/\sqrt{2m} + 0.1/\sqrt{l}] \\ &\leq 2\exp\left(-\frac{m}{100l}\right), x = w, w',\end{aligned}$$

proving the second inequality of the present proposition.

We now prove the third inequality of the present proposition. Let

$$\widehat{w} \stackrel{\Delta}{=} \frac{Pw}{\|Pw\|}$$

and

$$\widehat{w'} \stackrel{\Delta}{=} \frac{Pw'}{\|Pw'\|}$$

(note that $\|Pw\| = 0$ and $\|Pw'\| = 0$ are each zero probability events). By Fact 3.1, $\widehat{w}, \widehat{w'}$ are random independently chosen unit vectors in V. By the argument used in the proof of the first inequality 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,

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

proving the third equality of the present proposition. ■

Proposition 3.2 *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,*

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

Proof: By Fact 3.1, 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

$$w'' \stackrel{\Delta}{=} y - \langle x|y\rangle x,$$

and set $w \stackrel{\Delta}{=} x$ and

$$w' \stackrel{\Delta}{=} \frac{w''}{\|w''\|}.$$

Now,

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

By Proposition 3.1 we see that,

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

proving the present proposition. ■

Lemma 3.6 *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,*

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

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 3.1, 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 3.1 and 3.2 we have with probability at least

$$1 - 2d\exp\left(-\frac{m}{100l}\right) - 10d^2\exp\left(-\frac{m}{100d^4 l}\right),$$

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

We show that whenever this happens

$$|\langle w|P|w\rangle w| \leq 6/l$$

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,

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

Thus,

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

completing the proof of the present lemma. ■

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

Theorem 3.5 (Quantum incompressibility) *Let the underlying Hilbert space be $\mathbb{C}^m$. There exist n states ρ_{ij} and n orthogonal projections M_{ij}, $1 \leq i \leq \frac{n}{2^k}$, $1 \leq j \leq 2^k$, such that*

(a) $\forall i, j \operatorname{Tr} M_{ij}\rho_{ij} = 1$.

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

(c) $\forall i, j\ S(\rho_{ij} \| \rho) = k$.

(d) Suppose $d \geq 2^k$, $n = \Omega(d^5 \log d 2^{2k})$ and $2^k d \log d = \Omega(n/m)$. Then for all subspaces W of dimension d,

$$|\{M_{ij} : M_{ij}(W) \leq 1/10\}| \geq n/4.$$

Proof: For $1 \leq i \leq \frac{n}{2^k}$, choose $\mathcal{B}^i = (|b_1^i\rangle, \ldots, |b_m^i\rangle)$ to be a random orthonormal basis of $\mathbb{C}^m$. $\mathcal{B}^i$ is chosen independently of $\mathcal{B}^k$, $k \neq i$. Partition the sequence $\mathcal{B}^i$ into 2^k equal parts; call these parts $\mathcal{B}^{ij}$, $1 \leq j \leq 2^k$. Define

$$\rho_{ij} \stackrel{\Delta}{=} \frac{2^k}{m} \cdot \sum_{v \in \mathcal{B}^{ij}} |v\rangle\langle v|.$$

Define

$$M_{ij} \stackrel{\Delta}{=} \sum_{v \in \mathcal{B}^{ij}} |v\rangle\langle v|.$$

It is easy to see that ρ_{ij}, M_{ij} satisfy parts (a), (b) and (c).

To prove part (d), we reason as follows. Let W be a fixed subspace of $\mathbb{C}^m$ of dimension d. Let V be a random subspace of $\mathbb{C}^m$ of dimension 2^k. Let P denote the orthogonal projection operator on V. By the $\mathbf{U}(m)$-invariance of the distribution $\mu_{m,d}$ and from Lemma 3.6,

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

Define the set

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

Hence for a fixed $i \in [n/2^k]$,

$$\Pr[i \in \mathsf{Bad}] \leq 2^k \exp\left(-\frac{m}{200 \cdot 2^k d^4}\right) \leq \exp\left(-\frac{m}{300 \cdot 2^k d^4}\right).$$

Since the events $i \in \mathsf{Bad}$ are independent,

$$\begin{aligned}\Pr\left[|\mathsf{Bad}| \geq \frac{3n}{4 \cdot 2^k}\right] &\leq \binom{\frac{n}{2^k}}{\frac{3n}{4 \cdot 2^k}} \exp\left(-\frac{3mn}{1200 \cdot 2^{2k} d^4}\right).\\ &\leq (4e/3)^{3n/2^{k+2}} \exp\left(-\frac{3mn}{1200 \cdot 2^{2k} d^4}\right).\end{aligned}$$

So

$$\Pr[|\{M_{ij} : M_{ij}(W) \leq 6/2^k\}| \geq 3n/4] \leq (4e/3)^{3n/2^{k+2}} \exp\left(-\frac{3mn}{1200 \cdot 2^{2k} d^4}\right).$$

By setting $\epsilon = 1/20$ in Lemma 3.5, we get

$$\begin{aligned}&\Pr[\exists W \text{ subspace of } \mathbb{C}^m, \dim(W) = d, |\{M_{ij} : M_{ij}(W) \leq 1/10\}| \geq 3n/4]\\ &\leq (4e/3)^{3n/2^{k+2}} \exp(O(md \log d)) \exp\left(-\frac{3mn}{1200 \cdot 2^{2k} d^4}\right)\\ &< 1,\end{aligned}$$

for the given constraints on the parameters. This completes the proof of part (d) of the present lemma. ■

Remark: The above theorem 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.

Chapter 4

Substate theorem and the index function problem

4.1 Substate theorem

In this chapter we prove a fundamental theorem about relative entropy of quantum states, which roughly states that if the relative entropy, of two quantum states ρ and σ is at most c, then $\rho/2^{O(c)}$ 'sits inside' σ. Using this substate theorem, we give tight lower bounds for the privacy loss of bounded error quantum communication protocols for the index function problem. In the next chapter we will the pointer chasing problem, in whose solution also substate theorem plays a crucial role. We will explain our information theoretic result, by first considering its classical analogue. Let P and Q be probability distributions on the set $[n]$ such that their relative entropy is bounded by c, that is

$$S(P\|Q) \stackrel{\Delta}{=} \sum_{i\in[n]} P(i)\log_2 \frac{P(i)}{Q(i)} \;\leq\; c. \tag{4.1}$$

When $c \ll 1$, this implies that P and Q are close to each other; indeed, one can show that (see [CT91, Lemma 12.6.1])

$$\|P-Q\|_t \stackrel{\Delta}{=} \sum_{i\in[n]} |P(i)-Q(i)| \;\leq\; \sqrt{(2\ln 2)c}. \tag{4.2}$$

That is, the probability of an event $\mathcal{E} \subseteq [n]$ in P is close to its probability in Q:

$$|P(\mathcal{E}) - Q(\mathcal{E})| \leq \sqrt{(c\ln 2)/2}.$$

We are, however, sometimes concerned with a situation when $c \gg 1$. In that case, (4.2) becomes weak: we cannot even infer from it that an event $\mathcal{E}$ with probability $3/4$ in P has positive probability in Q. But is it true that when $S(P\|Q) < +\infty$ $P(\mathcal{E}) > 0$, then $Q(\mathcal{E}) > 0$? Yes! To see this, let us reinterpret the expression in (4.1) as the expectation of $\log P(i)/Q(i)$ as i is chosen according to P. Thus, one is lead to believe that if $S(P\|Q) \leq$

$c < +\infty$, then $\log P(i)/Q(i)$ is typically bounded by c, that is, $P(i)/Q(i)$ is typically bounded by 2^c. One can formalise this intuition and show, for all $r > 1$,

$$\Pr_{i \in P}\left[\frac{P(i)}{Q(i)} \geq 2^{r(c+1)}\right] \leq \frac{1}{r}.$$

Let

$$\mathsf{Good} \triangleq \{i : P(i)/2^{rc} \leq Q(i)\},$$

$$P'(i) \triangleq P(i \mid i \in \mathsf{Good}).$$

That is in P' we just discard the bad values of i, and renormalised. Now, $\frac{r-1}{r2^{r(c+1)}}P'$ is dominated by Q everywhere. We have thus proved the following.

Proposition 4.1 *If $S(P\|Q) \leq c$, then for all $r > 1$, there exists a distribution P' such that*

$$|P - P'|_1 \leq \frac{2}{r}$$

and

$$Q = \alpha P' + (1-\alpha)P'',$$

where P'' is some other distribution and $\alpha = 2^{-O(rc)}$.

Let us return to our event $\mathcal{E}$ that occurred with some small probability p in P. Now, if we take r to be $2/p$, then $\mathcal{E}$ occurs with probability at least $p/2$ in P', and hence appears with probability $p/2^{O(rc)}$ in Q. Thus, we have shown that even though P and Q are far apart as distributions, events that have positive probability (no matter how small) in P, continue to have positive probability in Q.

We prove the following quantum analogue of Proposition 4.1.

Result 3 (Substate theorem) Suppose ρ and σ are quantum states with $S(\rho\|\sigma) \leq c$. Then, for all $r > 1$, there are states ρ' and ρ'' such that

$$\|\rho - \rho'\|_t \leq 2/\sqrt{r}$$

and

$$\sigma = \alpha\rho' + (1-\alpha)\rho'',$$

where $\alpha = 2^{O(rc)}$.

(This has been stated here in a form that brings out the analogy with the classical statement above. Below (Theorem 4.1) we give a more nuanced statement which is better suited for our applications.)

The ideas used to arrive at Proposition 4.1 do not immediately generalise to get this statement, because ρ and σ may not be simultaneously diagonalisable. As it turns out, our proof of the substate theorem takes an indirect route. First, by exploiting the Fuchs and Caves [FC95] characterisation of fidelity and a minimax theorem of game theory, we obtain a 'lifting' theorem about an 'observational' version of relative entropy; this statement is interesting on its own. Using this 'lifting' theorem, and a connection between the 'observational' version of relative entropy and actual relative entropy, we argue that it is enough to verify the original statement when ρ and σ reside in a two-dimensional space and ρ is a pure state. The two dimensional case is then established by a direct computation.

We now state the substate theorem as it is actually used in our lower bound proofs for the index function problem and the pointer chasing problem.

Theorem 4.1 (Substate theorem) *Consider two finite dimensional Hilbert spaces* $\mathbf{H}$ *and* $\mathcal{K}$*, where* $\dim(\mathcal{K}) \geq \dim(\mathbf{H})$*. Let* $\mathbb{C}^2$ *denote the two dimensional complex Hilbert space. Let* ρ, σ *be density matrices in* $\mathbf{H}$*. Let* $r > 1$ *be any real number. Let* $|\psi\rangle$ *be a purification of* ρ *in* $\mathbf{H} \otimes \mathcal{K}$*. Then there exist pure states* $|\phi\rangle, |\theta\rangle \in \mathbf{H} \otimes \mathcal{K}$ *(depending on* r*) and* $|\zeta\rangle \in \mathbf{H} \otimes \mathcal{K} \otimes \mathbb{C}^2$ *such that* $|\zeta\rangle$ *is a purification of* σ *and*

$$\left\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \right\|_t \leq 2/\sqrt{r},$$

where

$$|\zeta\rangle \triangleq \sqrt{\frac{r-1}{r2^{rk}}}\,|\phi\rangle|0\rangle + \sqrt{1 - \frac{r-1}{r2^{rk}}}\,|\theta\rangle|1\rangle \quad \text{and} \quad k \triangleq 8S(\rho\|\sigma) + 14.$$

Remarks:

1. Note that Result 3 follows from above by tracing out $\mathcal{K} \otimes \mathbb{C}^2$.
2. From Result 3, one can easily see that

$$\|\rho - \sigma\|_t \leq 2 - 2^{-O(k)}.$$

This implies a $2^{-O(k)}$ lower bound on the fidelity of ρ and σ.

As stated earlier, to prove the substate theorem, it is useful for us to define a new notion of distinguishability between density matrices. We shall call this notion *observational divergence.*

Definition 4.1 (Observational divergence) *Let* ρ, σ *be density matrices in the same finite dimensional Hilbert space* $\mathbf{H}$*. Their observational divergence is defined as*

$$D(\rho\|\sigma) \triangleq \sup_F \left(\mathrm{Tr}\ (F\rho) \log \frac{\mathrm{Tr}\ (F\rho)}{\mathrm{Tr}\ (F\sigma)} \right),$$

where F *above ranges over POVM elements on* $\mathbf{H}$ *such that* $\mathrm{Tr}\ (F\sigma) \neq 0$*.*

The following properties of observational divergence follow easily from the definition.

Lemma 4.1 *Let* ρ, σ *be density matrices in the same finite dimensional Hilbert space* $\mathbf{H}$*. Then*

1. $D(\rho\|\sigma) \geq 0$*, with equality iff* $\rho = \sigma$*.*

2. *$D(\rho\|\sigma) < +\infty$ iff $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$. If $D(\rho\|\sigma) < +\infty$, then there is a POVM element F which achieves equality in Definition 4.1.*

3. *$D(\cdot\|\cdot)$ is continuous in its two arguments when it is not infinite.*

4. *(Joint convexity) Let $\rho_1, \rho_2, \sigma_1, \sigma_2$ be density matrices in $\mathbf{H}$. Let $\rho \stackrel{\Delta}{=} \lambda\rho_1 + (1-\lambda)\rho_2$ and $\sigma \stackrel{\Delta}{=} \lambda\sigma_1 + (1-\lambda)\sigma_2$, where $0 \leq \lambda \leq 1$. Then*
$$D(\rho\|\sigma) \leq \lambda D(\rho_1\|\sigma_1) + (1-\lambda)D(\rho_2\|\sigma_2).$$

5. *(Unitary invariance) If U is a unitary transformation on $\mathbf{H}$,*
$$D(U\rho U^\dagger\|U\sigma U^\dagger) = D(\rho\|\sigma).$$

6. *(Monotonicity) Suppose $\mathcal{K}$ is a finite dimensional Hilbert space, and ρ', σ' are density matrices in $\mathbf{H} \otimes \mathcal{K}$ such that $\mathrm{Tr}_{\mathcal{K}}\ \rho' = \rho$ and $\mathrm{Tr}_{\mathcal{K}}\ \sigma' = \sigma$. Then,*
$$D(\rho'\|\sigma') \geq D(\rho\|\sigma).$$
This implies, via unitary invariance and the Kraus representation theorem, that if $\mathcal{F}$ is a completely positive trace preserving superoperator, then
$$D(\mathcal{F}\rho\|\mathcal{F}\sigma) \leq D(\rho\|\sigma).$$

Fact 2.9 and Lemma 4.1 seem to suggest that observational divergence and relative entropy are very similar quantities. In fact, the relative entropy is an upper bound on the observational divergence to within an additive constant.

Lemma 4.2 *Let ρ, σ be density matrices in the same finite dimensional Hilbert space $\mathbf{H}$. Then,*
$$D(\rho\|\sigma) < S(\rho\|\sigma) + 1.$$

Proof: By Fact 2.9 and Lemma 4.1, $D(\rho\|\sigma) = +\infty$ iff $\mathrm{supp}(\rho) \not\subseteq \mathrm{supp}(\sigma)$ iff $S(\rho\|\sigma) = +\infty$. Thus, we can henceforth assume without loss of generality that $D(\rho\|\sigma) < +\infty$. By Lemma 4.1, there is a POVM element F such that
$$D(\rho\|\sigma) = p\log(p/q),$$
where $p \stackrel{\Delta}{=} \mathrm{Tr}\ (F\rho)$ and $q \stackrel{\Delta}{=} \mathrm{Tr}\ (F\sigma)$. We now have
$$\begin{aligned} S(\rho\|\sigma) &\geq p\log\frac{p}{q} + (1-p)\log\frac{(1-p)}{(1-q)} \\ &> p\log\frac{p}{q} + (1-p)\log\frac{1}{(1-q)} - 1 \\ &\geq p\log\frac{p}{q} - 1 \\ &= D(\rho\|\sigma) - 1. \end{aligned}$$

The first inequality follows from the Lindblad-Uhlmann monotonicity of relative entropy (Fact 2.9), and the second inequality follows because

$$(1-p)\log(1-p) \geq (-\log e)/e > -1,$$

for $0 \leq p \leq 1$. This completes the proof of the lemma. ■

We now prove the following lemma, which can be thought of as a substate theorem when the first density matrix is in fact a pure state.

Lemma 4.3 *Let $|\psi\rangle$ be a pure state and σ be a density matrix in the same finite dimensional Hilbert space $\mathbf{H}$. Let*

$$k \triangleq D\left((|\psi\rangle\langle\psi|)\|\sigma\right).$$

If $k > 0$, then for all $r > 1$, there exists a pure state $|\phi\rangle$ (depending on r) such that

$$\left\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \right\|_t \leq \frac{2}{\sqrt{r}} \quad \text{and} \quad \left(\frac{r-1}{r2^{rk}}\right) |\phi\rangle\langle\phi| \leq \sigma.$$

Proof: We assume without loss of generality that $0 < k < +\infty$. Consider

$$M \triangleq \sigma - (|\psi\rangle\langle\psi|/2^{rk}).$$

Since $-(|\psi\rangle\langle\psi|/2^{rk})$ has exactly one non-zero eigenvalue and this eigenvalue is negative viz. $-1/2^{rk}$, and σ is positive semidefinite, M is a hermitian matrix with at most one negative eigenvalue.

If $M \geq 0$ we take $|\phi\rangle$ to be $|\psi\rangle$. The lemma trivially holds in this case.

Otherwise, let $|w\rangle$ be the eigenvector corresponding to the unique negative eigenvalue $-\alpha$ of M. Thinking of $|w\rangle\langle w|$ as a POVM element, we get

$$\begin{aligned} 0 > -\alpha = \mathrm{Tr}\ (M|w\rangle\langle w|) &= \langle w|\sigma|w\rangle - \frac{|\langle\psi|w\rangle|^2}{2^{rk}} \\ \Rightarrow \langle w|\sigma|w\rangle &< \frac{|\langle\psi|w\rangle|^2}{2^{rk}}. \end{aligned}$$

Hence

$$\begin{aligned} k = D(|\psi\rangle\langle\psi|\|\sigma) \geq |\langle\psi|w\rangle|^2 \log \frac{|\langle\psi|w\rangle|^2}{\langle w|\sigma|w\rangle} &> rk|\langle\psi|w\rangle|^2 \\ \Rightarrow |\langle\psi|w\rangle|^2 < \frac{1}{r} &< 1. \end{aligned}$$

In particular, this shows that $|\psi\rangle, |w\rangle$ are linearly independent.

Let $n \triangleq \dim(\mathbf{H})$. Let $\{|v\rangle, |w\rangle\}$ be an orthonormal basis for the two dimensional subspace of $\mathbf{H}$ spanned by $\{|\psi\rangle, |w\rangle\}$. Extend it to $\{|v_1\rangle, \ldots, |v_{n-2}\rangle, |v\rangle, |w\rangle\}$, an orthonormal basis for the entire space $\mathbf{H}$. In this basis we have the following matrix equation,

$$\begin{bmatrix} F & e & d \\ e^* & a & b \\ d^* & b^* & c \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0^* & x & y \\ 0^* & y^* & z \end{bmatrix} = \begin{bmatrix} P & l \\ l^* & -\alpha \end{bmatrix}, \tag{4.3}$$

where the first, second and third matrices are σ, $|\psi\rangle\langle\psi|/2^{rk}$ and M respectively. F is an $(n-2) \times (n-2)$ matrix, P is an $(n-1) \times (n-1)$ matrix, d, e are $(n-2) \times 1$ matrices and l is an $(n-1) \times 1$ matrix. a, c, x, z, α are non-negative real numbers and b, y are complex numbers. The zeroes above denote all zero matrices of appropriate dimensions. The asterisk denotes conjugate transpose.

Claim 4.1 *We have the following properties.*

1. $b, y \in \mathbb{C}$, $a, c, x, z, \alpha \in \mathbb{R}$.
2. $b = y \neq 0$, $1/(r2^{rk}) > z = c + \alpha > c > 0$, $\alpha > 0$, $a > 0$, $0 < x < 1/2^{rk}$, $x + z = 1/2^{rk}$, $l = 0$ *and* $d = 0$.
3. $0 < \frac{xc}{|b|^2} < \frac{xz}{|y|^2} = 1$.

Proof: The first part of the claim has already been mentioned above. Since $|w\rangle$ is an eigenvector of M corresponding to eigenvalue $-\alpha$, $l = 0$. By inspection, we have $b = y, z = c + \alpha, d = 0$. $x > 0$ since $|\psi\rangle$, $|w\rangle$ are linearly independent, and $z > c \geq 0$ since $\alpha > 0$.

$$x + z = \mathrm{Tr}\ (|\psi\rangle\langle\psi|/2^{rk}) = 1/2^{rk}$$

$$\Rightarrow x < 1/2^{rk}.$$

Also,

$$z = |\langle\psi|w\rangle|^2/2^{rk} < 1/(r2^{rk}).$$

Since $\sigma \geq 0$, $F \geq 0$ and $\begin{bmatrix} a & b \\ b^* & c \end{bmatrix} \geq 0$. Hence,

$$\det \begin{bmatrix} a & b \\ b^* & c \end{bmatrix} = ac - |b|^2 \geq 0.$$

Since $|\psi\rangle\langle\psi|/2^{rk}$ has one dimensional support,

$$\det \begin{bmatrix} x & y \\ y^* & z \end{bmatrix} = xz - |y|^2 = 0.$$

If $c = 0$ then $y = b = 0$ which implies that $xz = 0$, which is a contradiction. Hence, $c > 0$ and $b \neq 0$. Similarly, $a > 0$. This proves the second part of the claim. The third part now follows easily. ■

We can now write $\sigma = \sigma_1 + \sigma_2$, where

$$\sigma_1 \triangleq \begin{bmatrix} F & e & 0 \\ e^* & a - \frac{|b|^2}{c} & 0 \\ 0^* & 0^* & 0 \end{bmatrix} \quad \text{and} \quad \sigma_2 \triangleq \begin{bmatrix} 0 & 0 & 0 \\ 0^* & \frac{|b|^2}{c} & b \\ 0^* & b^* & c \end{bmatrix}.$$

Note that

$$|\xi\rangle = (0, \ldots, 0, 1, (-b^*)/c)$$

is an eigenvector of σ_2 corresponding to the eigenvalue 0. $\sigma_2 \geq 0$, and in fact, σ_2 has one dimensional support. We now claim that $\sigma_1 \geq 0$. For otherwise, since $F \geq 0$, there is a vector $|\theta\rangle$ of the form $(a_1, \ldots, a_{n-2}, 1, 0)$ such that

$$\langle\theta|\sigma_1|\theta\rangle < 0.$$

Now consider the vector

$$|\theta'\rangle \stackrel{\Delta}{=} (a_1, \ldots, a_{n-2}, 1, (-b^*)/c).$$

We have,

$$\langle\theta'|\sigma|\theta'\rangle = \langle\theta'|\sigma_1|\theta'\rangle + \langle\theta'|\sigma_2|\theta'\rangle = \langle\theta|\sigma_1|\theta\rangle + \langle\xi|\sigma_2|\xi\rangle < 0,$$

contradicting $\sigma \geq 0$. This shows that $\sigma_1 \geq 0$, and hence, $\sigma \geq \sigma_2$.

We are now finally in a position to define the pure state $|\phi\rangle$. $|\phi\rangle\langle\phi|$ is nothing but σ_2 normalised to have unit trace.

$$|\phi\rangle\langle\phi| \stackrel{\Delta}{=} \frac{\sigma_2}{\frac{|b|^2}{c} + c}.$$

Using Claim 4.1 we get,

$$\mathrm{Tr}\ \sigma_2 = \frac{|b|^2}{c} + c \geq \frac{|b|^2}{z} + c = x + z - \alpha \geq \frac{r-1}{r2^{rk}}.$$

Hence,

$$\frac{r-1}{r2^{rk}} |\phi\rangle\langle\phi| < \sigma_2 \leq \sigma.$$

This shows the second assertion of the lemma.

To complete the proof of the lemma, we still need to show that $\left\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \right\|_t$ is small. Up to global phase factors, one can write $|\psi\rangle, |\phi\rangle$ as follows.

$$|\psi\rangle = \frac{\frac{b}{\sqrt{z}}|v\rangle + \sqrt{z}|w\rangle}{\sqrt{\frac{|b|^2}{z} + z}} \quad |\phi\rangle = \frac{\frac{b}{\sqrt{c}}|v\rangle + \sqrt{c}|w\rangle}{\sqrt{\frac{|b|^2}{c} + c}}.$$

We now lower bound $|\langle\phi|\psi\rangle|$ as follows, using Claim 4.1.

$$\begin{aligned} |\langle\phi|\psi\rangle| &\geq \frac{\frac{|b|^2}{\sqrt{cz}} + \sqrt{cz}}{\sqrt{\frac{|b|^2}{c} + c}\sqrt{\frac{|b|^2}{z} + z}} \\ &= \frac{|b|^2 + cz}{\sqrt{(|b|^2 + c^2)(|b|^2 + z^2)}} \\ &\geq \frac{|b|^2 + cz}{\sqrt{(|b|^2 + cz)(|b|^2 + z^2)}} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{\frac{|b|^2 + cz}{|b|^2 + z^2}} \\
&= \sqrt{\frac{x + c}{x + z}} \\
&= \sqrt{1 - \frac{\alpha}{x + z}} \\
&\geq \sqrt{1 - 1/r}.
\end{aligned}
$$

This proves that

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_t = 2\sqrt{1 - |\langle\phi|\psi\rangle|^2} \leq \frac{2}{\sqrt{r}},$$

establishing the first assertion of the lemma and completing its proof. ■

We next prove the following lemma, which can be thought of as an 'observational substate' lemma.

Lemma 4.4 *Consider two finite dimensional Hilbert spaces* $\mathbf{H}$ *and* $\mathcal{K}$, $\dim(\mathcal{K}) \geq \dim(\mathbf{H})$. *Let* ρ, σ *be density matrices in* $\mathbf{H}$. *Let* $|\psi\rangle$ *be a purification of* ρ *in* $\mathbf{H} \otimes \mathcal{K}$. *Let* F *be a POVM element on* $\mathbf{H} \otimes \mathcal{K}$. *Then there exists a purification* $|\phi\rangle$ *of* σ *in* $\mathbf{H} \otimes \mathcal{K}$ *such that*

$$q \geq \frac{p}{2^{k'/p}},$$

where

$$p \stackrel{\Delta}{=} \mathrm{Tr}\ (F|\psi\rangle\langle\psi|), q \stackrel{\Delta}{=} \mathrm{Tr}\ (F|\phi\rangle\langle\phi|)$$

and

$$k' \stackrel{\Delta}{=} 4D(\rho\|\sigma) + 2.$$

Proof: We assume without loss of generality that $0 < D(\rho\|\sigma) < +\infty$ and that $p > 0$. Let $n \stackrel{\Delta}{=} \dim(\mathbf{H})$ and $\{|\alpha_i\rangle\}_{i=1}^n$ be the orthonormal eigenvectors of F with corresponding eigenvalues $\{\lambda_i\}_{i=1}^n$. Note that $0 \leq \lambda_i \leq 1$ and $|\alpha_i\rangle \in \mathbf{H} \otimes \mathcal{K}$. We have,

$$p = \sum_{i=1}^n \lambda_i |\langle\alpha_i|\psi\rangle|^2 \quad \text{and} \quad q = \sum_{i=1}^n \lambda_i |\langle\alpha_i|\phi\rangle|^2.$$

Define,

$$|\theta'\rangle \stackrel{\Delta}{=} \frac{\sum_{i=1}^n \lambda_i \langle\alpha_i|\psi\rangle |\alpha_i\rangle}{\sqrt{p}} \quad \text{and} \quad |\theta\rangle \stackrel{\Delta}{=} \frac{|\theta'\rangle}{\| |\theta'\rangle \|}.$$

Note that

$$p = |\langle\psi|\theta\rangle|^2 \| |\theta'\rangle \|^2$$

and

$$0 < \| |\theta'\rangle \|^2 \leq 1.$$

Using the Cauchy-Schwarz inequality, we see that

$$\begin{aligned} |\langle\phi|\theta\rangle|^2 \| |\theta'\rangle\|^2 = |\langle\phi|\theta'\rangle|^2 &= \frac{\left|\sum_{i=1}^n \lambda_i \langle\alpha_i|\psi\rangle\langle\phi|\alpha_i\rangle\right|^2}{\sum_{i=1}^n \lambda_i |\langle\alpha_i|\psi\rangle|^2} \\ &\leq \sum_{i=1}^n \lambda_i |\langle\alpha_i|\phi\rangle|^2 = q. \end{aligned}$$

Thus,

$$\frac{p}{2^{k'/p}} = \frac{|\langle\psi|\theta\rangle|^2 \| |\theta'\rangle\|^2}{2^{k'/(|\langle\psi|\theta\rangle|^2 \| |\theta'\rangle\|^2)}} \leq \frac{|\langle\psi|\theta\rangle|^2 \| |\theta'\rangle\|^2}{2^{k'/|\langle\psi|\theta\rangle|^2}}.$$

Hence, it will suffice to show that there exists a purification $|\phi\rangle$ of σ in $\mathbf{H}\otimes\mathcal{K}$ such that

$$|\langle\phi|\theta\rangle|^2 \geq \frac{|\langle\psi|\theta\rangle|^2}{2^{k'/|\langle\psi|\theta\rangle|^2}}.$$

Define the density matrix τ in $\mathbf{H}$ as

$$\tau \triangleq \mathrm{Tr}_{\mathcal{K}}\ |\theta\rangle\langle\theta|.$$

By Facts 2.11 and 2.12, there is a purification $|\phi\rangle$ of σ in $\mathbf{H}\otimes\mathcal{K}$ and a POVM $\{F_1, \ldots, F_l\}$ in $\mathbf{H}$ such that,

$$|\langle\phi|\theta\rangle| = B(\tau, \sigma) = \sum_{i=1}^l \sqrt{c_i b_i},$$

where $c_i \triangleq \mathrm{Tr}\ (F_i\tau)$ and $b_i \triangleq \mathrm{Tr}\ (F_i\sigma)$.
Let $a_i \triangleq \mathrm{Tr}\ (F_i\rho)$. We know from Facts 2.11 and 2.12 that

$$0 < \sqrt{p} \leq |\langle\psi|\theta\rangle| \leq B(\tau, \rho) \leq \sum_{i=1}^l \sqrt{c_i a_i}.$$

Note that the a_i's are non-negative real numbers summing up to 1, and so are the b_i's and the c_i's.
Define the set S as

$$S \triangleq \left\{ i \in [l] : a_i > b_i 2^{4k/B(\tau,\rho)^2} \right\},$$

where

$$k \triangleq D(\rho\|\sigma).$$

Note that $\forall i \in S, b_i \neq 0$ as $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\sigma)$, k being finite. Define the POVM element G on $\mathbf{H}$ as

$$G \triangleq \sum_{i\in S} F_i.$$

Let $a \stackrel{\Delta}{=} \mathrm{Tr}\ (G\rho)$ and $b \stackrel{\Delta}{=} \mathrm{Tr}\ (G\sigma)$. Then

$$a = \sum_{i \in S} a_i, \ b = \sum_{i \in S} b_i,$$

$$b > 0 \text{ and } a > b\, 2^{4k/B(\tau,\rho)^2}.$$

We have that

$$D(\rho \| \sigma) = k \geq a \log \frac{a}{b} > \frac{4ka}{B(\tau,\rho)^2}$$
$$\Rightarrow a < \frac{B(\tau,\rho)^2}{4}.$$

Now, by the Cauchy-Schwarz inequality and the other inequalities proved above, we get

$$\begin{aligned}
B(\tau,\rho) &\leq \sum_{i=1}^{l} \sqrt{c_i a_i} \\
&= \sum_{i \in S} \sqrt{c_i a_i} + \sum_{i \notin S} \sqrt{c_i a_i} \\
&\leq \sqrt{\sum_{i \in S} c_i} \sqrt{\sum_{i \in S} a_i} + 2^{2k/B(\tau,\rho)^2} \sum_{i \notin S} \sqrt{c_i b_i} \\
&\leq 1 \cdot \sqrt{a} + 2^{2k/B(\tau,\rho)^2} B(\tau,\sigma) \\
&< \frac{B(\tau,\rho)}{2} + 2^{2k/B(\tau,\rho)^2} B(\tau,\sigma).
\end{aligned}$$

This shows that

$$B(\tau,\rho)^2 < 4 \cdot 2^{4k/B(\tau,\rho)^2} B(\tau,\sigma)^2$$
$$\Rightarrow |\langle\psi|\theta\rangle|^2 < 4 \cdot 2^{4k/|\langle\psi|\theta\rangle|^2} |\langle\phi|\theta\rangle|^2.$$

Since $k' = 4k + 2$, we get

$$|\langle\phi|\theta\rangle|^2 \geq \frac{|\langle\psi|\theta\rangle|^2}{2^{k'/|\langle\psi|\theta\rangle|^2}},$$

completing the proof of the lemma. ■

In the previous lemma, the purification $|\phi\rangle$ of σ was a function of the POVM element F. We now prove a lemma which, for any fixed $0 \leq p \leq 1$, removes the dependence on F, for

$$\mathrm{Tr}\ (F|\psi\rangle\langle\psi|) \geq p,$$

at the expense of having a mixed extension of σ in the place of a pure extension (i.e. purification).

Lemma 4.5 *Consider two finite dimensional Hilbert spaces* $\mathbf{H}$ *and* $\mathcal{K}$, $\dim(\mathcal{K}) \geq \dim(\mathbf{H})$. *Let* ρ, σ *be density matrices in* $\mathbf{H}$ *and* $|\psi\rangle$ *be a purification of* ρ *in* $\mathbf{H} \otimes \mathcal{K}$. *Let* $0 \leq p \leq 1$. *There exists a density matrix* ω *in* $\mathbf{H} \otimes \mathcal{K}$ *such that*

$$\mathrm{Tr}_{\mathcal{K}}\ \omega = \sigma,$$

and for all POVM elements F *on* $\mathbf{H} \otimes \mathcal{K}$ *such that*

$$\mathrm{Tr}\ (F|\psi\rangle\langle\psi|) \geq p,\ \ \mathrm{Tr}\ (F\omega) \geq p/2^{k'/p},$$

where

$$k' \triangleq 4D(\rho\|\sigma) + 2.$$

Proof: We assume without loss of generality that $0 < D(\rho\|\sigma) < +\infty$ and that $p > 0$. Consider the set A_1 of all extensions ω of σ in $\mathbf{H} \otimes \mathcal{K}$ i.e.

$$\mathrm{Tr}_{\mathcal{K}}\ \omega = \sigma.$$

A_1 is a non-empty, compact, convex set. Consider the set A_2 of all POVM operators F in $\mathbf{H} \otimes \mathcal{K}$ such that

$$\mathrm{Tr}\ (F|\psi\rangle\langle\psi|) \geq p.$$

A_2 is a compact convex set. Without loss of generality, A_2 is non-empty. The conditions of Fact 2.15 are trivially satisfied (note that we think of our matrices, which in general have complex entries, as vectors in a larger real vector space). For every $F \in A_2$, we have a purification $|\phi^F\rangle \in \mathbf{H} \otimes \mathcal{K}$ of σ such that

$$\mathrm{Tr}\ \left(F|\phi^F\rangle\langle\phi^F|\right) \geq \frac{\mathrm{Tr}\ (F|\psi\rangle\langle\psi|)}{2^{k'}/\mathrm{Tr}\ (F|\psi\rangle\langle\psi|)} \geq \frac{p}{2^{k'/p}}.$$

Using Fact 2.15, we see that there exists a density matrix ω in $\mathbf{H} \otimes \mathcal{K}$ such that,

$$\mathrm{Tr}_{\mathcal{K}}\ \omega = \sigma$$

and

$$\mathrm{Tr}\ (F\omega) \geq \frac{p}{2^{k'/p}}$$

for all $F \in A_2$. This completes the proof. ■

The previous lemma depends on the parameter p. We now remove this restriction, to get an 'observational substate' theorem.

Theorem 4.2 (Observational divergence lifting theorem) *Consider two finite dimensional Hilbert spaces* $\mathbf{H}, \mathcal{K}$, $\dim(\mathcal{K}) \geq \dim(\mathbf{H})$. *Let* ρ, σ *be density matrices in* $\mathbf{H}$. *Let* $|\psi\rangle$ *be a purification of* ρ *in* $\mathbf{H} \otimes \mathcal{K}$. *Then there exists a density matrix* ω *in* $\mathbf{H} \otimes \mathcal{K}$ *such that*

$$\mathrm{Tr}_{\mathcal{K}}\ \omega = \sigma$$

and

$$D((|\psi\rangle\langle\psi|)\ \|\omega) < 8D(\rho\|\sigma) + 6.$$

Proof: We assume without loss of generality that $0 < D(\rho\|\sigma) < +\infty$ and that $p > 0$. Define the function $f : [0,1] \to [0,1]$ as follows.

$$f(p) \triangleq \frac{p}{2^{k/p}} \quad \text{where} \quad 0 \leq p \leq 1 \quad \text{and} \quad k \triangleq 4D(\rho\|\sigma) + 2.$$

For a fixed positive integer l, define the density matrix ω_l in $\mathbf{H} \otimes \mathcal{K}$ as

$$\omega_l \triangleq (1/l) \sum_{i=1}^{l} \omega(i/l),$$

where for $0 \leq p \leq 1$, $\omega(p)$ is a density matrix in $\mathbf{H} \otimes \mathcal{K}$ such that

$$\mathrm{Tr}_{\mathcal{K}}\ \omega(p) = \sigma$$

and

$$\mathrm{Tr}\ (F\omega(p)) \geq f(p)$$

for all POVM elements F on $\mathbf{H} \otimes \mathcal{K}$ satisfying

$$\mathrm{Tr}\ (F|\psi\rangle\langle\psi|) \geq p.$$

Such an $\omega(p)$ exists by Lemma 4.5. Then,

$$\mathrm{Tr}_{\mathcal{K}}\ \omega(l) = \sigma.$$

Suppose F is a POVM element on $\mathbf{H} \otimes \mathcal{K}$. Let

$$j/l \leq p \triangleq \mathrm{Tr}\ (F|\psi\rangle\langle\psi|) < (j+1)/l,$$

where $0 \leq j \leq l$. We assume without loss of generality that $p > 0$. Then,

$$\begin{aligned} \mathrm{Tr}\ (F\omega_l) \geq \frac{1}{l}\sum_{i=1}^{j} \mathrm{Tr}\ \left(F\omega\left(\frac{i}{l}\right)\right) &\geq \frac{1}{l}\sum_{i=1}^{j} f\left(\frac{i}{l}\right) \\ &\geq \frac{j}{l} f\left(\frac{j+1}{2l}\right) \geq \left(p - \frac{1}{l}\right) f\left(\frac{p}{2}\right). \end{aligned}$$

The third inequality above follows from the convexity of $f(\cdot)$. By compactness, the set $\{\omega_l : l \in \mathbb{N}\}$ has a limit point ω. By standard continuity arguments, $\mathrm{Tr}_{\mathcal{K}}\ \omega = \sigma$ and

$$q \triangleq \mathrm{Tr}\ (F\omega) \geq pf\left(\frac{p}{2}\right) = \frac{p^2}{2 \cdot 2^{2k/p}}.$$

Hence, $q > 0$ and

$$p \log \frac{p}{q} \leq -p\log p + p + 2k < 2k + 2 \leq 8D(\rho\|\sigma) + 6.$$

The second inequality follows because

$$-p \log p + p < 2$$

for $0 \leq p \leq 1$. This completes the proof of the lemma. ■

The above theorem relates the observational divergence of a pair of density matrices to the observational divergence of their extensions in an extended Hilbert space, where the extension of the first density matrix is a pure state. Using this, we are now finally in a position to prove the substate theorem.

Proof:(Substate theorem, Theorem 4.1) By Lemma 4.2 and Theorem 4.2, there exists a density matrix ω in $\mathbf{H} \otimes \mathcal{K}$ such that

$$\mathrm{Tr}_{\mathcal{K}}\ \omega = \sigma$$

and

$$D\left(\left(|\psi\rangle\langle\psi| \right) \| \, \omega \right) \leq 8D(\rho\|\sigma) + 6 < 8\left(S(\rho\|\sigma) + 1 \right) + 6 = 8S(\rho\|\sigma) + 14 = k.$$

By Lemma 4.3, there exists a pure state $|\phi\rangle$ such that

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_t \leq \frac{2}{\sqrt{r}} \quad \text{and} \quad \left(\frac{r-1}{r2^{rk'}} \right) |\phi\rangle\langle\phi| \leq \omega.$$

Let

$$\tau_1 \triangleq \mathrm{Tr}_{\mathcal{K}}\ |\phi\rangle\langle\phi|.$$

By above, there exists a density matrix τ_2 in $\mathbf{H}$ such that

$$\sigma = \left(\frac{r-1}{r2^{rk}} \right) \tau_1 + \left(1 - \frac{r-1}{r2^{rk}} \right) \tau_2.$$

Let $|\theta\rangle \in \mathbf{H} \otimes \mathcal{K}$ be a canonical purification of τ_2. Then $|\zeta\rangle$ defined above is a purification of σ in $\mathbf{H} \otimes \mathcal{K} \otimes \mathbb{C}^2$. This completes the proof. ■

4.2 Privacy tradeoffs

In this section, we prove a tradeoff between privacy loss of Alice and privacy loss of Bob for the index function problem INDEX_n. We then indicate how similar tradeoffs can be proved for disjointness DISJ_n and inner product mod 2 IP_n, and formalise the idea in terms of VC-dimension. From these tradeoffs, the logarithmic privacy loss for INDEX_n, DISJ_n and IP_n trivially follows. We also mention some corollaries of these results.

Before proving our results let us briefly mention why substate theorem is crucial in our proofs.

Need for substate theorem We know (by Nayak's observation described above) that if B sends only b qubits, then A must send $n/2^{O(b)}$ qubits to solve the index function problem. Our results appear similar, but unfortunately, the old proof is not applicable now. The old argument relied on the fact that A can generate a distribution on messages, so that every potential message of B is well-represented in this distribution: if the messages are classical and only b bits long, the uniform distribution is such a distribution—each b bit message appears in it with probability 2^{-b}. This argument depends crucially on the fact that the message contains only b classical bits. *In our problem, we are not assuming that messages of B have at most b qubits*, only that they reveal less than b bits of information about B's input. So, A cannot just guess B's messages. This is where the substate theorem comes in handy. Let ρ_i be the state that A and B reach when A's input is a uniform superposition and B's input is i. The substate theorem allows us to show that (roughly) A and B can exchange messages (independent of i) and arrive at a state ρ which contains $\rho_i/2^{O(b)}$ as a 'substate'. After this, a standard argument of Cleve, van Dam, Nielsen and Tapp [CvDNT98] takes over.

Lemma 4.6 *Consider a safe quantum protocol $\mathcal{P}$ for* INDEX$_n$*. Let μ denote the uniform probability distribution on Alice's and Bob's inputs. Suppose $\mathcal{P}$ has error at most* $1/4$ *with respect to μ. Suppose*

$$L^{\mathcal{P}}(\mathrm{INDEX}_n, \mu, B, A) \leq k.$$

Then,

$$L^{\mathcal{P}}(\mathrm{INDEX}_n, \mu, A, B) \geq n/2^{O(k)}.$$

Proof: Let registers A, X, B, Y denote Alice's work qubits, Alice's input qubits, Bob's work qubits and Bob's input qubits respectively, at the end of protocol $\mathcal{P}$. We can assume without loss of generality that the last round of communication in $\mathcal{P}$ is from Alice to Bob, since otherwise, we can add an extra round of communication at the end wherein Alice sends the answer qubit to Bob. This process increases $L^{\mathcal{P}}(\mathrm{INDEX}_n, \mu, A, B)$ by at most two (by Fact 2.7), and does not increase $L^{\mathcal{P}}(\mathrm{INDEX}_n, \mu, B, A)$. Thus, at the end of $\mathcal{P}$, Bob measures the answer qubit (which is a qubit in the register B) in the computational basis to determine $f(x, y)$. In the proof, subscripts of pure and mixed states will denote the registers which are in those states.

Let $|\psi_i\rangle_{XAYB}$ be the state vector of Alice's and Bob's qubits and $(\rho_i)_{XA}$ the density matrix of Alice's qubits at the end of the protocol $\mathcal{P}$, when Alice is fed a uniform superposition over bit strings in her input register X and Bob's is fed the index $|i\rangle$ in his input register Y. Let $1/2 + \epsilon_i$ be the success probability of $\mathcal{P}$ in this case. Without loss of generality, $\epsilon_i \geq 0$. Consider a run, Run 1, of $\mathcal{P}$ when a uniform mixture of indices is fed to register Y, and a uniform superposition over bit strings is fed to register X. Let $1/2 + \epsilon$ be the success probability of $\mathcal{P}$ for Run 1, which is also the success probability of $\mathcal{P}$ with respect to μ. Then

$$1/4 \leq \epsilon = (1/n) \sum_{i=1}^{n} \epsilon_i.$$

Let $I(Y:AX)$ denote the mutual information of register Y with registers AX at the end of Run 1 of $\mathcal{P}$. We know that

$$I(Y:AX) = L^{\mathcal{P}}(\mathrm{INDEX}_n, \mu, B, A) \leq k.$$

Let

$$\rho \triangleq (1/n) \sum_{i=1}^{n} \rho_i \text{ and } k_i \triangleq S(\rho_i \| \rho).$$

By Fact 2.10,

$$k \geq I(Y:AX) = \frac{1}{n} \sum_{i=1}^{n} S(\rho_i \| \rho) = \frac{1}{n} \sum_{i=1}^{n} k_i.$$

Let

$$k'_i \triangleq 8k_i + 14, r_i \triangleq (4/\epsilon_i)^2 \text{ and } k' \triangleq (1/n) \sum_{i=1}^{n} k'_i.$$

Then,

$$k' \leq 8k + 14.$$

Let us now consider a run, Run 2, of $\mathcal{P}$ with uniform superpositions fed to registers X, Y. Let $|\phi\rangle_{XAYB}$ be the state vector of Alice's and Bob's qubits at the end of Run 2 of $\mathcal{P}$. Since $\mathcal{P}$ is a safe protocol,

$$\mathrm{Tr}_{YB}\, |\phi\rangle\langle\phi| = \rho_{XA},$$

and the success probability of $\mathcal{P}$ for Run 2 is $1/2 + \epsilon$. Let Q be an additional qubit. By the substate theorem (Theorem 4.1), there exists a state

$$|\phi_i\rangle_{XAYBQ} \triangleq \sqrt{\frac{r_i - 1}{r_i 2^{r_i k'_i}}}\, |\psi'_i\rangle_{XAYB} |0\rangle_Q + \sqrt{1 - \frac{r_i - 1}{r_i 2^{r_i k'_i}}}\, |\theta'_i\rangle_{XAYB} |1\rangle_Q,$$

where

$$\| |\psi_i\rangle\langle\psi_i| - |\psi'_i\rangle\langle\psi'_i| \|_t \leq 4/\sqrt{r_i} = \epsilon_i$$

and

$$\mathrm{Tr}_{YBQ}\, |\phi_i\rangle\langle\phi_i| = \rho_{XA}.$$

In fact, there exists a unitary transformation U_i on registers YBQ, transforming the state $|\phi\rangle_{XAYB}|0\rangle_Q$ to the state $|\phi_i\rangle_{XAYBQ}$.

For each $i \in [n]$, let X'_i denote the classical random variable got by measuring the ith bit of register X in state $|\phi\rangle_{XAYB}$. We now prove the following claim.

Claim For each $i \in [n]$, there is a boolean valued POVM $\mathcal{M}_i$ acting on YB such that, if Z'_i is the result of $\mathcal{M}_i$ on $|\phi\rangle_{XAYB}$, then

$$\Pr[Z'_i = X'_i] \geq 1/2 + \epsilon_i / 2^{O(k_i)/\epsilon_i^2}.$$

Proof: $\mathcal{M}_i$ proceeds by first applying U_i to $|\phi\rangle_{XAYB}|0\rangle_Q$ to get $|\phi_i\rangle_{XAYBQ}$. It then measures Q in the computational basis. If it sees $|0\rangle_Q$, it measures the answer qubit in the computational basis and declares the result as Z'_i. If it measures $|1\rangle_Q$, it tosses a fair boolean coin and declares the result as Z'_i.

Let us now analyse the probability that $X'_i = Z'_i$. In the case when $\mathcal{M}_i$ measures $|0\rangle$ for qubit Q, which happens with probability $(r_i - 1)/(r_i 2^{r_i k'_i})$, the state vector of $XAYB$ collapses to $|\psi'_i\rangle$. In this case, by Fact 2.8,

$$\Pr[Z'_i = X'_i | Q = 0] \geq \frac{1}{2} + \epsilon_i - \frac{1}{2} \left\| |\psi_i\rangle\langle\psi_i| - |\psi'_i\rangle\langle\psi'_i| \right\|_t \geq \frac{1}{2} + \frac{\epsilon_i}{2}.$$

In the case when $\mathcal{M}_i$ measures $|1\rangle$ for qubit Q, which happens with probability $1 - (r_i - 1)/(r_i 2^{r_i k'_i})$,

$$\Pr[Z'_i = X'_i | Q = 1] = 1/2.$$

Thus,

$$\Pr[Z'_i = X'_i] \geq \frac{1}{2} + \frac{(r_i - 1)\epsilon_i}{2 r_i 2^{r_i k'_i}} \geq \frac{1}{2} + \frac{\epsilon_i}{2^{O(k_i)/\epsilon_i^2}}.$$

■

Consider now a run, Run 3, of $\mathcal{P}$ when a uniform mixture over bit strings is fed to register X and a uniform superposition over indices is fed to register Y. Let ρ_{XAYB} denote the density matrix of the registers $XAYB$ at the end of Run 3 of $\mathcal{P}$. In fact, measuring in the computational basis the register X in the state $|\phi\rangle_{XAYB}$ gives us ρ_{XAYB}. Let $I(X : YB)$ denote the mutual information between register X and registers YB in the state ρ_{XAYB}. For each $i \in [n]$, let X_i denote the classical random variable corresponding to the ith bit of register X in state ρ_{XAYB}. Then $X_1, \ldots, X_n$ are independent random variables and $X = X_1 \ldots X_n$. Let Z_i denote the result of POVM $\mathcal{M}_i$ on ρ_{XAYB}. Then,

$$\Pr[Z_i = X_i] = \Pr[Z'_i = X'_i] \geq 1/2 + \epsilon_i / 2^{O(k_i)/\epsilon_i^2}.$$

By Facts 2.5 and 2.6,

$$I(X_i : YB) \geq I(X_i : Z_i) \geq \epsilon_i^2 / 2^{O(k_i)/\epsilon_i^2}.$$

By Fact 2.4 and convexity,

$$\begin{aligned} I(X : YB) &\geq \sum_{i=1}^{n} I(X_i : YB) \\ &\geq \sum_{i=1}^{n} \frac{\epsilon_i^2}{2^{O(k_i)/\epsilon_i^2}} \geq \frac{n\epsilon^2}{2^{O(k)/\epsilon^2}} \geq \frac{n}{2^{O(k)}}. \end{aligned}$$

This shows that

$$L^{\mathcal{P}}(\mathrm{INDEX}_n, \mu, A, B) = I(X : YB) \geq n/2^{O(k)}.$$

■

Remark: This lemma is the formal version of Result 1 stated in the introduction.

The same privacy tradeoff result holds for functions f into which INDEX_n can be 'embedded'. Examples of such functions f are DISJ_n and IP_n. This idea can be formalised in terms of VC-dimension as follows.

Theorem 4.3 *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ be a boolean valued function. Let*

$$\mathrm{VC}_{\mathcal{X}}(f) = n.$$

Then there is a product distribution μ on $\mathcal{X} \times \mathcal{Y}$ such that, if $\mathcal{P}$ a safe quantum protocol for f with average error at most $1/4$ with respect to μ,

$$L^{\mathcal{P}}(f, \mu, B, A) \leq k \Leftrightarrow L^{\mathcal{P}}(f, \mu, A, B) \geq n/2^{O(k)}.$$

An analogous statement holds for $\mathrm{VC}_{\mathcal{Y}}(f)$.

Proof: Since $\mathrm{VC}_{\mathcal{X}}(f) = n$, there is a set $S \subseteq \mathcal{Y}$, $|S| = n$ which is shattered. Without loss of generality, $S = [n]$. Let $r \in \{0, 1\}^n$. r can be thought of as the characteristic vector of a subset $R \subseteq S$. There is an $x \in \mathcal{X}$ such that

$$\forall y \in S : f(x, y) = 1 \Leftrightarrow y \in R.$$

We now give a reduction from INDEX_n to f as follows: In INDEX_n, Alice is given an $r \in \{0, 1\}^n$ and Bob is given a $y \in [n]$. Alice and Bob run the protocol $\mathcal{P}$ for f on inputs x and y respectively, to solve INDEX_n. The theorem now follows from Lemma 4.6. ■

The following corollaries are now immediate, using Fact 2.7.

Corollary 4.1 INDEX_n, DISJ_n *and* IP_n *suffer from $\Omega(\log n)$ privacy loss.*

Corollary 4.2 *Let $\mathcal{P}$ be a safe quantum protocol for* INDEX_n *with average error at most $1/4$ with respect to the uniform distribution μ. Suppose*

$$L^{\mathcal{P}}(f, \mu, B, A) \leq k.$$

Let m be the number of qubits communicated by Alice to Bob. Then,

$$m = n/2^{O(k)}.$$

Corollary 4.3 *Let $\mathcal{P}$ be a safe quantum protocol for* INDEX_n *with worst case error at most $1/4$. Suppose Bob sends at most k qubits to Alice. Let m be the number of qubits communicated by Alice to Bob. Then,*

$$m = n/2^{O(k)}.$$

Chapter 5

The pointer chasing problem

In this chapter we prove our lower bounds for both the full and the bit versions of the pointer chasing problem.

5.1 The pointer chasing problem P_k

In this section, we prove our lower bound for the pointer chasing problem P_k. Below we define it again.

> Let V_A and V_B be disjoint subsets of size n. Player A is given a function $F_A : V_A \to V_B$ and player B is given a function $F_B : V_B \to V_A$. Let $F = F_A \cup F_B$. There is a fixed vertex s in V_B. A and B need to communicate to determine $t = F^{(k+1)}(s)$ with probability of correctness being at least 3/4; k and s are known to both parties in advance.

Before proving the lower bound, we discuss the difficulty one encounters while applying the existing techniques for this problem, and explain at an intuitive level why our new tool namely substate theorem is useful.

Comparison with previous work Let us review the idea behind the $\Omega(n/2^{2^{O(k)}})$ lower bound for P_k proved in Klauck et al. [KNTZ01a]. Assume that A and B are given uniformly random functions, and there is a protocol that solves P_k using k messages with ϵn qubits each, $\epsilon \ll 1$. In particular, the first message of the protocol (from A) has at most ϵn bits. Then, for a typical $s' \in V_A$ the information contained in this message about $F_A[s']$ must be $O(\epsilon)$. That is, the first message is 'roughly the same' when $F_A[s']$ takes different values. So, we should be able to eliminate the first message and assume that $F_B[s] = s'$, to obtain a solution for the problem P_{k-1}. This, is the (simplistic) intuition behind the proof in [KNTZ01a]. However, much depends on the notion of 'roughly the same'. For [KNTZ01a], two messages are deemed to be 'roughly the same' if their trace distance (see section 2.2 for definition) is small, which implies that an observation can distinguish them by probability at most δ (which goes to zero as ϵ goes to zero). In particular, this implies that any event in one state will occur in the other state with probability roughly the same (within $\delta/2$, in fact). This connection between trace distance of the messages and information (i.e. the connection between δ and ϵ) is not powerful enough when the information is not $\epsilon \ll 1$ but some large number (perhaps even growing with n). Consider, for example, two random $\log n$ bit numbers which share their first ten bits. If we take two typical (with probability $1 - 2^{-10}$) instances of the first number, the two conditional distributions on the second number have trace distance 2. Thus, the quantity δ above can be close to 2 when the information is allowed to grow above 1 (in the above example, the information is 10). In particular, an event that occurs with probability 3/4 in one case might not happen at all in the other. We are precisely in such a situation. If the messages have rn bits, all we can guarantee is that A's first message contains at most r bits of information about $F_A[s']$, for typical s', and we wish to allow r to grow with n (or at least take values larger than 1).

To get around this limitation in the existing techniques, which exploit only the connection between information and trace distance, we consider a different notion of 'roughly the same' for quantum states. This notion is that of *relative entropy.* Using substate theorem, this allows us to conclude that if an event occurs with probability p in one case, then it occurs with probability $p/2^{O(r/p)}$ in the other. This is good enough for eliminating the first message and for obtaining a solution to the problem P_{k-1}.

Our proof technique: The underlying information theoretic tools we use are, in fact, mainly taken from the paper Klauck, Nayak, Ta-Shma and Zuckerman [KNTZ01b]. Our proofs use the round elimination method, stated explicitly in the classical communication complexity setting by Miltersen, Nisan, Safra and Wigderson [MNSW98a]. This technique was applied in the quantum setting by Klauck *et al.*, who developed several tools, notably the *average encoding theorem* and the *local transition theorem.* Their argument was refined further by Sen and Venkatesh [SV98]. In this thesis, we adapt this argument but we consider a slightly different pointer chasing problem, where the two players are allowed to generate their own inputs and then proceed to compute the answer. To keep this problem non-trivial we must impose some restrictions on the way the players behave. First, we insist that the inputs they generate must be sufficiently rich. Second, the amount of communication before the input is generated, is limited. In previous round elimination arguments, the inputs were supplied to the two players from 'outside'. While this worked well for many problems, for the pointer chasing problem it made things difficult. However, letting the players generate their inputs gives rise to new technical difficulties, because the inputs they generate are not exactly what we want, but only close to it. So, we need to apply a correction step, that converts a protocol whose inputs have a distribution close to the one we desire into one where the inputs are exactly what we want. Overall, we believe, the main contribution of this work is in showing how existing information theoretic tools can be better exploited for round elimination in quantum communication protocols.

Below we mention an improved version of the average encoding theorem and the local transition theorem which we will be using in our proofs.

5.1.1 Improved average encoding and local transition theorem

In this section, we observe that the following lemma from [DHR78] can be used to improve the average encoding and local transition arguments of [KNTZ01a]. If Lemmas 5.2 and 5.3 are used in their place, the factor k^4 in the denominator of some existing lower bounds (e.g. [KNTZ01a] and [JRS02b]) can be replaced by k^2.

Lemma 5.1 *Let ρ and σ be two density matrices such that $S(\rho\|\sigma)$ is finite. Then,*

$$B(\rho,\sigma) \geq 2^{-S(\rho\|\sigma)/2}.$$

Proof: Let M be the complete orthogonal measurement which achieves the infimum as in the Fact 2.12. Let P and Q be the classical distributions resulting after the measurement M is performed. From Fact 2.9 and concavity of the log function it follows that:

$$\begin{aligned} -(1/2)S(\rho\|\sigma) \leq -(1/2)S(P\|Q) &= \sum_i p_i \log \sqrt{q_i/p_i} \\ &\leq \log \sum_i \sqrt{q_i p_i} \\ &= \log B(P,Q) = \log B(\rho,\sigma). \end{aligned}$$

■

Corollary 5.1 *Let ρ and σ be two density matrices such that $S(\rho\|\sigma)$ is finite. Then,*

$$1 - B(\rho,\sigma) \leq ((\ln 2)/2)S(\rho\|\sigma).$$

Proof: If $((\ln 2)/2)S(\rho\|\sigma) \geq 1$ then the inequality is trivial since $B(\ , \ ) \geq 0$. Therefore when $((\ln 2)/2)S(\rho\|\sigma) \leq 1$,

$$\begin{aligned} B(\rho,\sigma) &\geq 2^{-S(\rho\|\sigma)/2} \\ &\geq \exp(-((\ln 2)/2)S(\rho\|\sigma)) \\ &\geq 1 - ((\ln 2)/2)S(\rho\|\sigma) \quad (\text{since} \quad \exp(-x) \geq 1 - x, \quad \text{for} \quad 0 \leq x \leq 1) \\ \Rightarrow 1 - B(\rho,\sigma) &\leq ((\ln 2)/2)S(\rho\|\sigma). \end{aligned}$$

■

The following lemma follows immediately from the above corollary and Fact 2.10.

Lemma 5.2 (Average encoding theorem) *Suppose X, Q are two disjoint quantum systems, where X is a classical random variable which takes value x with probability p_x, and Q is a quantum encoding $x \mapsto \sigma_x$ of X. Let the density matrix of the average encoding be $\sigma \stackrel{\Delta}{=} \sum_x p_x \sigma_x$. Then,*

$$\sum_x p_x(1 - B(\rho,\rho_x)) \leq (\ln 2/2)I(X:Q).$$

The following lemma follows immediately from Fact 2.11 and Fact 2.13 and Corollary 5.1

Lemma 5.3 (Local transition theorem) *Let ρ_1, ρ_2 be two density matrices in the same finite dimensional Hilbert space $\mathbf{H}$, $\mathcal{K}$ any Hilbert space of dimension at least the dimension of $\mathbf{H}$, and $|\phi_i\rangle$ any purifications of ρ_i in $\mathbf{H}\otimes\mathcal{K}$. Then, there is a local unitary transformation U on $\mathcal{K}$ that maps $|\phi_2\rangle$ to $|\phi_2'\rangle \overset{\Delta}{=} (I \otimes U)|\phi_2\rangle$ (I is the identity operator on $\mathbf{H}$) such that*

$$\begin{aligned} \||\phi_1\rangle\langle\phi_1| - |\phi_2'\rangle\langle\phi_2'|\|_t &= 2\sqrt{1 - B(\rho_1,\rho_2)^2} \\ &\le 2\sqrt{2(1 - B(\rho_1,\rho_2))} \le 2\sqrt{\ln 2(S(\rho_1\|\rho_2))}. \end{aligned}$$

Fact 5.1 ([Lin91]) *Suppose X and Q are two classical correlated random variables, where X is uniformly distributed over $\{0,1\}$ and Q is an encoding $x \to P_x$ of X. Then,*

$$1 - B(P_1, P_2) \le I(X : Q).$$

Following corollary is immediate from Fact 2.12 and monotonicity of information,

Corollary 5.2 *Let $x \to \sigma_x$ be a quantum encoding, where $x \in \{0,1\}$. Let X be a random variable uniformly distributed in $\{0,1\}$. Let $\sigma = (\sigma_1 + \sigma_2)/2$. Then,*

$$1 - B(\sigma_1, \sigma_2) \le I(X : \sigma).$$

We will also use the following elementary fact.

Fact 5.2 *Suppose D, D' are two probability distributions on the same finite set X, whose total variation distance is $\|D - D'\|_1 = \delta$. Then, there exists a stochastic matrix $P = (p_{xx'})_{xx' \in X}$, such that $D = PD'$ and*

$$\sum_{x' \in X} P(x', x')D(x') = 1 - \frac{\delta}{2}.$$

Let $\mathcal{H}$ be a Hilbert space with basis $(|x\rangle : x \in X)$. Let C be a unitary transformation on $\mathcal{H} \otimes \mathcal{H}$ that maps basis vectors of the form $|x'\rangle|\mathbf{0}\rangle$ (where $\mathbf{0}$ is a special element of X) according to the rule

$$|x'\rangle|\mathbf{0}\rangle \to |x'\rangle \otimes \sum_{x \in X} \sqrt{p_{xx'}}|x\rangle,$$

and maps other standard basis vectors suitably. Suppose R' and R are registers that can hold states in $\mathcal{H}$, where R' contains a mixture of basis states with distribution D' and R is in the state $|\mathbf{0}\rangle$. Apply C to (R', R), and then measure the registers in the computational basis. Let the resulting random variables (taking values in X) be Z' and Z. Then, Z' has distribution D', Z has distribution D and

$$\Pr[Z \neq Z'] \le \frac{\delta}{2}.$$

Note, that C acts safely on R'.

We prove our lower bound using an inductive argument. It will be convenient to state our induction hypothesis by means of predicates Q_k^A and Q_k^B, defined below. Roughly, the induction proceeds as follows. We show that if there is an efficient protocol for P_k, then Q_k^A is true. We then show independently that Q_ℓ^A implies $Q_{\ell-1}^B$ and Q_ℓ^B implies $Q_{\ell-1}^A$, and that Q_0^A and Q_0^B are false. Thus, there is no efficient protocol for P_k.

We first define

$$Q_k^A(c_1, \ldots, c_k, n_a, n_b, \epsilon),$$

for $k \geq 1$. Then, separately, we define $Q_0^A(\epsilon)$. For $k \geq 0$, Q_k^B is the same as Q_k^A, with the roles of Alice and Bob reversed. Consequently, all our statements involving Q_k^A and Q_k^B have two forms, where one is obtained from the other by reversing the roles of Alice and Bob. We will typically state just one of them, and let the reader infer the other.

From now on, the term *measurement* means a von Neumann measurement in the computational basis.

5.1.2 The predicate $Q_k^A, k \geq 1$

The predicate $Q_k^A(c_1, \ldots, c_k, n_a, n_b, \epsilon)$ holds if there is a quantum protocol $\mathcal{P}$ of the following form.

Input generation: In $\mathcal{P}$, Alice and Bob 'generate' most of their inputs themselves. Alice has n input registers $(F_A[u] : u \in V_A)$ and Bob has n input registers $(F_B[v] : v \in V_B)$. There is a fixed vertex $s \in V_B$, that is known to both players. Each of Alice's registers has $\log n$ qubits so that it can hold a description of a vertex in V_B; similarly, each of Bob's registers can hold a description of a vertex in V_A. In addition, Alice and Bob have registers for their 'work' qubits W_A and W_B.

When $\mathcal{P}$ starts, Alice's registers are all initialised to zero. On Bob's side, the register $F_B[s]$ starts off with the uniform superposition

$$\frac{1}{\sqrt{n}} \sum_{a \in V_A} |a\rangle;$$

his other registers are all zero.

Alice starts by generating a pure state in $\widetilde{M_1} M_1$, where $\widetilde{M_1}, M_1$ are each $c_1 n$-qubit registers. Then she applies a unitary transformation U_A on $\widetilde{M_1}$ plus some ancilla qubits to generate a pure state in registers F_A, W_A and M_1. After the application of U_A, F_A holds the 'generated input' to Alice for the pointer chasing problem, and W_A holds Alice's 'work qubits'. Alice then sends M_1 to Bob.

Now, Bob generates his input by applying a unitary transformation U_B on the registers M_1, F_B and W_B. U_B operates "safely" on $F_B[s]$. After the application of U_B, F_B holds the 'generated input' to Bob for the pointer chasing problem, and W_B holds Bob's 'work qubits'.

We will use F_A, F_B to refer to the actual states of the respective registers; f_A, f_B will denote the states that would result, were we to measure F_A, F_B. Thus, typically F_A, F_B

will be parts of a pure state (the global state of Alice's and Bob's qubits), whereas f_A, f_B will be mixtures of computational basis states.

For our predicate $Q_k^A(c_1, \ldots, c_k, n_a, n_b, \epsilon)$ to hold for a $k > 0$, this input generation process must satisfy some conditions.

> **Requirement 1(a):** There is a subset $X_A \subseteq V_A$ of size at most n_a such that the variables $(f_A(u) : u \in V_A)$ are independent, and for $u \in V_A - X_A$, $f_A(u)$ is uniformly distributed.
>
> **Requirement 1(b):** There is a subset $X_B \subseteq V_B - \{s\}$ of size at most n_b such that the random variables $(f_B(v) : v \in V_B)$ are independent, and $f_B(v)$ for $v \in V_B - X_B$ is uniformly distributed. Note that $f_B[s]$ is automatically uniformly distributed, because initially $F_B[s]$ contains the uniform superposition, and U_B acts safely on $F_B[s]$.

Communication: After U_A, U_B have been applied, Alice and Bob follow a quantum protocol exchanging further messages $M_2, \ldots, M_k$ of lengths $c_2 n, \ldots, c_k n$. Bob sends the message M_2. The rest of the protocol $\mathcal{P}$ (after the application of U_A and U_B) is required to act safely on registers F_A, F_B. At the end of $\mathcal{P}$, the player who receives M_k places $\log n$ qubits in a special register Ans. $\mathcal{P}$ then terminates.

The success probability: Once $\mathcal{P}$ has terminated, all registers are measured. Let ans denote the value observed in Ans, and let f_A and f_B be the values observed in F_A and F_B; we treat f_A and f_B as functions from V_A to V_B and V_B to V_A respectively. Let

$$f \triangleq f_A \cup f_B.$$

Requirement 4: $\Pr[\mathsf{ans} = f^{(k+1)}(s)] \geq \epsilon$.

5.1.3 The predicate Q_0^A

The predicate $Q_0^A(\epsilon)$ holds if there is a quantum protocol $\mathcal{P}$ of the following form. At the start of $\mathcal{P}$, Alice's registers are all initialised to zero. On Bob's side, the register $F_B[s]$ starts off with the uniform superposition

$$\frac{1}{\sqrt{n}} \sum_{a \in V_A} |a\rangle;$$

his other registers are all zero.

Alice starts by generating a pure state in $\widetilde{M_1} M_1$, where $\widetilde{M_1}, M_1$ are each $c_1 n$-qubit registers. Then she applies a unitary transformation U_A on $\widetilde{M_1}$ plus some ancilla qubits to generate a pure state in registers F_A, W_A and M_1. Alice then sends M_1 to Bob.

Now, Bob generates his input by applying a unitary transformation U_B on the registers M_1, F_B and W_B. U_B operates "safely" on $F_B[s]$.

Alice now places $\log n$ qubits in a special register Ans and $\mathcal{P}$ terminates. All the registers of Alice and Bob are measured. We require that

$$\mathsf{ans} = f_B[s]$$

with probability at least ϵ.

Following two lemmas are immediate from definitions.

Lemma 5.4 *If there is a safe quantum protocol for P_k^A with $v_0 = s \in V_B$, messages of lengths $c_1 n, \ldots, c_k n$, and success probability at least $3/4$ in the worst case, then $Q_k^A(c_1, \ldots, c_k, 0, 0, 3/4)$ is true.*

Lemma 5.5 *If $Q_0^A(\epsilon)$ is true, then*

$$\epsilon \leq n^{-1}.$$

The following lemma is the key to our inductive argument.

Lemma 5.6 (Round elimination) *(a) For $k \geq 2$, suppose that $Q_k^A(c_1, \ldots, c_k, n_A, n_B, \epsilon)$ holds (with $n_A < n$). Then, $Q_{k-1}^B(c_1 + c_2, c_3, \ldots, c_k, n_A, n_B + 1, \epsilon')$ holds with*

$$\epsilon' \triangleq 2^{\frac{-128\left(\frac{nc_1}{n-a}+2\right)}{\left(\epsilon-\frac{n_a}{n}\right)^2}}.$$

(b) If $Q_1^A(c_1, n_A, n_B, \epsilon)$ holds (with $n_A < n$), then $Q_0(\epsilon')$ holds, where ϵ' is the same as in (a).

The next subsection is devoted to the proof of this lemma. Now, assuming this lemma and Lemma 5.5, we get a lower bound on Q_k^A.

Theorem 5.1 *Let us suppose that k is constant and that $Q_k^A(c_1, \ldots, c_k, 0, 0, 3/4)$ holds. Then*

$$c_1 + c_2 + \cdots + c_k = \Omega(\log^{(k)} n).$$

Now, by using Lemma 5.4, we can derive our lower bound for P_k.

Corollary 5.3 *In any protocol for P_k, k constant, Alice and Bob must exchange a total of $\Omega(n \log^{(k)} n)$ qubits.*

5.1.4 Round elimination: proof of Lemma 5.6

We consider Part (a) first. Part (b) follows by using similar arguments, and we do not describe them explicitly. Suppose $Q_k^A(c_1, c_2, \ldots, c_k, n_A, n_B, \epsilon)$ is true and let protocol $\mathcal{P}$ satisfy the requirements. We will show that there is a protocol $\mathcal{Q}$ that satisfies the requirements for Q_{k-1}^B with parameters stated in Lemma 5.6(a).

In what follows, subscripts of pure and mixed states will denote the registers which are in those states. For $u \in V_A$, we use the subscript u instead of $F_A[u]$. Similarly, for $v \in V_B$, we use the subscript v instead of $F_B[v]$. For example, we say that the register $F_B[s]$ is initially in the state

$$|\mu\rangle_s \triangleq \frac{1}{\sqrt{n}} \sum_{u \in V_A} |u\rangle_s.$$

For $u \in V_A$, $F_{A,u}$ is a shorthand for registers $(F_A[w] : w \in V_A - \{u\})$. For $v \in V_B$, $F_{B,v}$ denotes likewise.

Suppose $a \in V_A$. Let T denote the registers M_1, $F_{A,a}$ and W_A. Let $|\psi^A_{a\to b}\rangle$ denote the (pure) state of T in protocol $\mathcal{P}$ just before Alice sends M_1 to Bob, if $f_A[a] = b$. (If $\Pr[f_A[a] = b] = 0$, then $|\psi^A_{a\to b}\rangle$ is defined to be the zero vector.) Let R denote the registers $F_{B,s}$ and W_B. Let $\ell_a \triangleq 1$ if $a \in X_A$ and $\ell_a \triangleq n$ otherwise. The global state vector of Alice and Bob in $\mathcal{P}$ just before Alice sends M_1 to Bob is

$$|\psi_{\text{in}}\rangle \triangleq \frac{1}{\sqrt{n}} \sum_{a \in V_A} \frac{1}{\sqrt{\ell_a}} \sum_{b \in V_B} |b\rangle_a |\psi^A_{a\to b}\rangle_T |a\rangle_s |\mathbf{0}\rangle_R.$$

At this point in $\mathcal{P}$, the first message M_1 is sent to Bob. Let the rest of the protocol starting from this point be $\mathcal{P}'$.

Let $\epsilon_{a\to b}$ be the probability of success when $\mathcal{P}'$ is run starting from the state

$$|b\rangle_a |\psi^A_{a\to b}\rangle_T |a\rangle_s |\mathbf{0}\rangle_R.$$

Since $\mathcal{P}'$ is safe, we have

$$\epsilon_{a\to b} = \Pr[\mathsf{ans} = f^{(k+1)}(s) \mid f_B[s] = a \text{ and } f_A[a] = b].$$

Also, we have

$$\epsilon = \mathop{\mathrm{E}}_{a,b}[\epsilon_{a\to b}] \leq \mathop{\mathrm{E}}_{a \in V_A - X_A, b \in V_B}[\epsilon_{a\to b}] + \frac{n_a}{n}. \tag{5.1}$$

In the first expectation, (a, b) are chosen with the same distribution as $(f_B[s], f_A[f_B[s]])$ of the given protocol $\mathcal{P}$; in the second, they are chosen uniformly and independently from the sets specified.

Let $\sigma_{1,a\to b}$ be the density matrix of register M_1 in $|\psi^A_{a\to b}\rangle$, and σ_1 be the density matrix of register M_1 in $|\psi_{\text{in}}\rangle$. Note that σ_1 is independent of the contents of register $F_B[s]$. Let $(M_1, \widetilde{M_1})$ contain the canonical purification of σ_1, where $\widetilde{M_1}$ is a $c_1 n$-qubit register. Then by the substate theorem (Theorem 4.1), there exists a unitary transformation $U_{a\to b}$ that when applied to $\widetilde{M_1}$ together with ancilla qubits, takes the pure state $(M_1, \widetilde{M_1})$ to a pure state $|\tilde{\theta}^A_{a\to b}\rangle$ on registers F_A, W_A, M_1, Q, where Q is a single ancilla qubit, such that

$$\begin{aligned} |\tilde{\theta}^A_{a\to b}\rangle_{F_A W_A M_1 Q} \quad = \quad & \sqrt{\delta_{a\to b}} |\tilde{\psi}^A_{a\to b}\rangle_{F_A W_A M_1} |0\rangle_Q + \\ & \sqrt{1 - \delta_{a\to b}}\, |\tilde{\phi}^A_{a\to b}\rangle_{F_A W_A M_1} |1\rangle_Q. \end{aligned}$$

Above, $|\tilde{\psi}^A_{a\to b}\rangle$, $|\tilde{\phi}^A_{a\to b}\rangle$ are pure states on registers F_A, W_A and M_1,

$$\delta_{a\to b} \triangleq \frac{1 - \frac{\epsilon^2_{a\to b}}{16}}{2^{(16/\epsilon^2_{a\to b})(8S(\sigma_{1,a\to b} \| \sigma_1)+14)}}, \tag{5.2}$$

and

$$\left\| |b\rangle\langle b| \otimes |\psi^A_{a\to b}\rangle\langle\psi^A_{a\to b}| - |\tilde{\psi}^A_{a\to b}\rangle\langle\tilde{\psi}^A_{a\to b}| \right\|_t \leq \frac{\epsilon_{a\to b}}{2}.$$

In particular, if $\mathcal{P}'$ is started from $|\tilde{\psi}^A_{a\to b}\rangle_{F_A W_A M_1}|a\rangle_s|\mathbf{0}\rangle_R$ (instead of $|b\rangle_a|\psi^A_{a\to b}\rangle_T|a\rangle_s|\mathbf{0}\rangle_R$), the probability of success is at least

$$\epsilon_{a\to b} - (\epsilon_{a\to b}/4) = (3\epsilon_{a\to b}/4).$$

The protocol $\mathcal{P}_{a\to b}$

Let us now fix $a \in V_A$ and $b \in V_B$ and consider the case when $f_B[s] = a$ and $f_A[a] = b$. We now describe a protocol $\mathcal{P}_{a\to b}$. This is just an intermediate protocol. Later we will describe how we obtain our final protocol $\mathcal{Q}$ (satisfying the requirements of Q^B_{k-1}) from $\mathcal{P}_{a\to b}$. It will be helpful, meanwhile, to keep in mind that in $\mathcal{Q}$, the roles of Alice and Bob will be reversed, $F_B[s]$ will be fixed at $|a\rangle$ (thus, we will add s to X_B), a will be our new s, and the state of $F_A[a]$ will be the uniform superposition

$$\frac{1}{\sqrt{n}} \sum_{v \in V_B} |v\rangle.$$

Step 1: Alice generates the canonical purification of σ_1 in registers $(M_1, \widetilde{M}_1)$. Alice applies $U_{a\to b}$ to $\widetilde{M}_1$ plus some ancilla qubits initialised to zero. She sends M_1 to Bob.
Step 2: Bob sets $F_B[s]$ to $|a\rangle$.
Step 3: Alice and Bob now proceed according to the protocol $\mathcal{P}'$. $\mathcal{P}'$ does not 'touch' the ancilla qubit Q described above.

Remark on the inputs generated: Let $\tilde{f}_{A,a\to b}$ be the random variable with distribution $\tilde{D}_{a\to b}$, resulting on measuring F_A in the state $|\tilde{\psi}^A_{a\to b}\rangle$. Let $f_{A,a\to b}$ be the random variable with distribution $D_{a\to b}$, resulting on measuring F_A in the state $|b\rangle|\psi^A_{a\to b}\rangle$. Then,

$$\|D_{a\to b} - \tilde{D}_{a\to b}\|_1 \leq \epsilon_{a\to b}/2.$$

Let $\tilde{\epsilon}_{a\to b}$ denote the probability of success of $\mathcal{P}_{a\to b}$. Since the ancilla qubit Q is zero with probability $\delta_{a\to b}$, we get that

$$\tilde{\epsilon}_{a\to b} \geq (3\delta_{a\to b}\epsilon_{a\to b})/4.$$

In $\mathcal{P}_{a\to b}$, f_B satisfies Requirements 1 and 2 with respect to $X_B \cup \{s\}$ (note that we want to eventually switch the roles of Alice and Bob). However, we cannot say that f_A satisfies Requirements 1 and 3. To get over this hurdle, we have to do a "correction process" on Alice's input registers as described below, leading us to protocol $\mathcal{P}'_{a\to b}$.

The protocol $\mathcal{P}'_{a\to b}$

We now describe the protocol $\mathcal{P}'_{a\to b}$. In $\mathcal{P}'_{a\to b}$, in addition to her registers in $\mathcal{P}_{a\to b}$, Alice has a fresh set of registers $\hat{F}_A$ of $n \log n$ qubits, initialised to zero. After $\mathcal{P}'_{a\to b}$ terminates, $\hat{F}_A$ is treated as Alice's input register while determining whether $\mathcal{P}'_{a\to b}$ succeeded or not.

Step 1(a): Alice generates the canonical purification of σ_1 in registers $(M_1, \widetilde{M_1})$. Alice applies $U_{a\to b}$ to $\widetilde{M_1}$ plus some ancilla qubits initialised to zero.

Step 1(b) (Correcting Alice's input registers): Alice now does a "correction process" on her input registers in order to satisfy Requirements 1 and 3. Let $C_{a\to b}$ be the unitary transformation corresponding to $\tilde{D}_{a\to b}$ and $D_{a\to b}$ according to Fact 5.2. If the ancilla qubit Q is zero, Alice applies $C_{a\to b}$ to registers $F_A, \hat{F}_A$. If Q is one, Alice sets $\hat{F}_A$ to

$$\sum_{y\in[n]^{V_A}} \sqrt{D_{a\to b}(y)}|y\rangle.$$

The input generation for Alice is now complete.

Step 1(c): Alice sends M_1 to Bob.

Step 2: Bob sets $F_B[s]$ to $|a\rangle$.

Step 3: From this point on, Alice and Bob just follow $\mathcal{P}'$. The registers $\hat{F}_A$ and the ancilla qubit Q are not 'touched' by $\mathcal{P}'$.

While executing $\mathcal{P}'$, Alice's old input registers (in protocol $\mathcal{P}$) F_A are used. Alice's new input registers $\hat{F}_A$ are not touched by any unitary transformation in $\mathcal{P}'$. At the end of $\mathcal{P}'_{a\to b}$ however, we will check the correctness of the answer with respect to $\hat{F}_A$ and F_B. Let $\hat{f}_{A,a\to b}$ denote the random variable got by measuring $\hat{F}_A$ at the end of $\mathcal{P}'_{a\to b}$. Note that if we measure $(F_A, \hat{F}_A)$ at the end of $\mathcal{P}'_{a\to b}$, the resulting random variables $(\tilde{f}_{A,a\to b}, \hat{f}_{A,a\to b})$ have distribution precisely $\tilde{D}_{a\to b}$ and $D_{a\to b}$. Furthermore (see Fact 5.2), if Q is zero then

$$\Pr[\tilde{f}_{A,a\to b} \neq \hat{f}_{A,a\to b}] \leq (1/2)\cdot(\epsilon_{a\to b}/2) = \epsilon_{a\to b}/4.$$

Let $\epsilon'_{a\to b}$ be the success probability of $\mathcal{P}'_{a\to b}$. Then one can see that

$$\begin{aligned} \epsilon'_{a\to b} &\geq (1-\epsilon_{a\to b}/4)(3\delta_{a\to b}\epsilon_{a\to b}/4) \\ &\geq (9\delta_{a\to b}\epsilon_{a\to b}/16). \end{aligned} \tag{5.3}$$

The protocol $\mathcal{P}'_a$

In $\mathcal{P}'_{a\to b}$, $\hat{f}_A[a] = b$ and $f_B[s] = a$. We now describe a protocol $\mathcal{P}'_a$ where $F_B[s]$ is fixed to $|a\rangle$, but $\hat{F}_A[a]$ contains the uniform superposition

$$\frac{1}{\sqrt{n}}\sum_{v\in V_B}|v\rangle.$$

In $\mathcal{P}'_a$, Alice starts the communication and $k-1$ rounds take place. $\hat{F}_A, F_B$ are Alice's and Bob's input registers respectively in $\mathcal{P}'_a$. Alice's old input registers F_A continue to exist,

but they count as her work qubits now. In $\mathcal{P}'_a$, in addition to her registers in $\mathcal{P}'_{a\to b}$, Alice has a fresh register Z of $\log n$ qubits. Initially, all qubits are initialised to zero.
Step 1: Bob generates the canonical purification of σ_1 in registers $(M_1, \widetilde{M}_1)$. He sets $F_B[s]$ to $|a\rangle$, and using the transformation U_B, generates his inputs F_B and work qubits W_B. Then he generates the first message M'_1 of protocol $\mathcal{P}'$ (this corresponds to message M_2 of $\mathcal{P}$), and sends M'_1 along with $\widetilde{M}_1$ to Alice.
Step 2(a): Alice places the uniform superposition state

$$\frac{1}{\sqrt{n}} \sum_{v \in V_B} |v\rangle$$

in register $\hat{F}_A[a]$. On receiving $\widetilde{M}_1$, Alice applies a unitary transformation on $\widetilde{M}_1$ plus some ancilla qubits initialised to zero. This unitary transformation is nothing but $U_{a\to b}$ if $\hat{f}_A[a] = b$. Note that it is safe on $\hat{F}_A[a]$.
Step 2(b): Alice does an 'input correction process' as follows. Let Y denote the registers $\hat{F}_{A,a}$ and Z. If the ancilla qubit Q is zero, Alice applies $C_{a\to b}$ to registers F_A, U if $\hat{f}_A[a] = b$. If Q is one, she sets $F_{A,a}$ to

$$\sum_{y \in [n]^{V_A - \{a\}}} \sqrt{D_{a\to b}(y)}|y\rangle.$$

Note that this 'input correction process' is safe on $\hat{F}_A[a]$.
Step 3: Alice resumes the protocol $\mathcal{P}'$. Note that Bob has already executed the first step of $\mathcal{P}'$ and sent M'_1. Alice responds to M'_1 as before. $\mathcal{P}'$ does not 'touch' $\hat{F}_A, Z, Q$.

Let $\hat{\epsilon}'_a$ be the success probability of $\mathcal{P}'_a$. Then, one can see that $\hat{\epsilon}'_a = \mathrm{E}_{b \in V_B}[\hat{\epsilon}'_{a\to b}]$, where b is chosen uniformly from V_B in the expectation.

The final protocol $\mathcal{Q}$

We first describe a protocol $\mathcal{P}_a$ with $k-1$ rounds of communication and Alice starting, which satisfies Requirements 1, 2 and 3 with respect to X_A and $X_B \cup \{s\}$, the roles of Alice and Bob being reversed, with a as the new 'starting vertex'. The only difference between $\mathcal{P}_a$ and $\mathcal{P}'_a$ is in Step 1. Let $\widehat{M}_1$ denote the first message of $\mathcal{P}'_a$, that is,

$$\widehat{M}_1 = (\widetilde{M}_1, M'_1).$$

$\widehat{M}_1$ is $(c_1+c_2)n$-qubits long. Let $(\widetilde{\widehat{M}}_1, \widehat{M}_1)$ contain a canonical purification of $\widehat{M}_1$, where $\widetilde{\widehat{M}}_1$ is $(c_1+c_2)n$-qubits long. In Step 1 of $\mathcal{P}_a$, Bob applies an appropriate unitary transformation on $\widetilde{\widehat{M}}_1$ to generate the same state as in at the end of Step 1 of $\mathcal{P}'_a$. After this, $\mathcal{P}_a$ proceeds in the same fashion as $\mathcal{P}'_a$. The success probability $\hat{\epsilon}_a$ of $\mathcal{P}_a$ is the same as the success probability $\hat{\epsilon}'_a$ of $\mathcal{P}_a$.

It can be shown that in $\mathcal{P}$,

$$\mathop{\mathrm{E}}_{a \in V_A - X_A}[I(f_A[a] : M_1)] \leq \frac{nc_1}{n - n_a}.$$

From this and (5.1), (5.2), (5.3) and joint convexity, we get that

$$\mathop{\mathrm{E}}_{a \in V_A - X_A} [\hat{\epsilon}_a] \geq \epsilon' \triangleq 2^{\frac{-128\left(\frac{nc_1}{n-a}+2\right)}{\left(\epsilon - \frac{n_a}{n}\right)^2}} .$$

In the expectations above, a is chosen uniformly from $V_A - X_A$.

Thus, there exists an $a \in V_A - X_A$ such that $\hat{\epsilon}_a \geq \epsilon'$. Our final protocol $\mathcal{Q}$ is nothing but $\mathcal{P}_a$ for this a. It can be verified that $\mathcal{Q}$ satisfies the requirements for $Q_{k-1}^B(c_1 + c_2, c_3, \ldots, c_k, n_A, n_B + 1, \epsilon')$. This completes the proof of Lemma 5.6.

5.2 The pointer chasing problem P_k^{bit}

In this section we prove our result concerning the bit version of the pointer chasing problem. We prove our result using a similar inductive argument as in the proof of the full version of the problem. First let us recall the definition of the problem

The input: Alice's input is a function

$$F_A : V_A \to V_B.$$

Bob's input is a function

$$F_B : V_B \to V_A.$$

V_A and V_B are disjoint sets of size n each. We assume that $n = 2^r$ for some $r \geq 1$.

The golden path: There is a fixed vertex $s \in V_B$. Let

$$F \triangleq F_A \cup F_B;$$

let

$$\mathsf{ans} \triangleq \mathrm{lsb}(F^{(k+1)}(s)).$$

Here lsb(x) is the least significant bit of x; we assume that vertices in V_A and V_B have binary encodings of length $\log n$.

The communication: Alice and Bob exchange messages $M_1, \ldots, M_k$, having lengths $c_1 n, \ldots, c_k n$, via a safe quantum protocol in order to determine ans. Alice starts the communication, that is, she sends M_1. The player receiving M_k places a guess for ans in the register Ans. We require that the bit obtained by measuring Ans in the computational basis[1] should be the correct answer (i.e. equal to lsb($F^{(k+1)}(s)$) with probability at least $\frac{3}{4}$, for all F_A, F_B.

[1] From now on, all measurements are to be performed using the computational basis.

5.2.1 The predicate Q_k^A

We will show our lower bound for P_k^{bit} using an inductive argument. It will be convenient to state our induction hypothesis by means of a predicates Q_k^A and Q_k^B, defined below, which are again very similar to the ones defined for P_k. In the inductive argument we show that if there is an efficient protocol for P_k^{bit}, then Q_k^A is true. We then show that Q_ℓ^A implies $Q_{\ell-1}^B$ and Q_ℓ^B implies $Q_{\ell-1}^A$, and that Q_0^A and Q_0^B are false. Thus, there is no efficient protocol for P_k^{bit}.

We now define $Q_k^A(c_1, \ldots, c_k, n_a, n_b, \epsilon)$ for $k \geq 1$. Then, separately, we define Q_0^A. For $k \geq 0$, Q_k^B is the same as Q_k^A, with the roles of Alice and Bob reversed. Consequently, all our statements involving Q_k^A and Q_k^B have two forms, where one is obtained from the other by reversing the roles of Alice and Bob. As earlier, we will typically state just one of them, and let the reader infer the other.

The predicate $Q_k^A(c_1, \ldots, c_k, n_a, n_b, \epsilon)$ holds if there is a quantum protocol of the following form.

Input generation: The input generation process is the same as in P_k; we repeat here for completeness. Alice and Bob 'generate' most of their inputs themselves. Alice has n input registers $(F_A[u] : u \in V_A)$ and Bob has n input registers $(F_B[v] : v \in V_B)$. There is a fixed vertex $s \in V_B$, that is known to both players. Each of Alice's registers has $\log n$ qubits so that it can hold a description of a vertex in V_B; similarly, each of Bob's registers can hold a description of a vertex in V_A. In addition, Alice and Bob have registers for their 'work' qubits W_A and W_B.

When the protocol starts, Alice's registers are all initialised to 0. On Bob's side, the register $F_B[s]$ starts off with the uniform superposition

$$|\mu\rangle \triangleq \frac{1}{\sqrt{n}} \sum_{a \in V_A} |a\rangle;$$

the other registers are all 0.

Alice starts by generating a pure state in $\widetilde{M}_1 M_1$, where $\widetilde{M}_1, M_1$ are each $c_1 n$ qubit registers. Then she applies a unitary transformation U_A on her registers other than M_1 to generate a state in registers F_A and W_A. Alice then sends M_1 to Bob.

Now, Bob generates his input using the message M_1 as follows. He applies a unitary transformation U_B on the registers that he owns at this point:

- M_1, the message registers just received from Alice;
- $F_B[s]$ the register holding the start pointer, which is in the state $|\mu\rangle$ in tensor with the other register;
- $(F_B[b] : b \in V_B - \{s\})$ and the registers W_B holding the work qubits of B, which contain 0.

U_B must operate “safely” on $F_B[s]$. F_B holds the ‘generated input’ to Bob for the pointer chasing problem, and W_B Bob’s ‘work qubits’.

We will use F_A, F_B also to refer to the actual states of the respective registers; f_A, f_B will denote the states that would result, were we to measure F_A, F_B. Thus, typically F_A, F_B will be parts of a pure state (the global state of Alice’s and Bob’s qubits) whereas f_A, f_B will be mixtures of computational basis states.

For our predicate $Q_k^A(c_1, \ldots, c_k, n_a, n_b, \epsilon)$ to hold, this input generation process must satisfy some conditions.

> **Requirement 1(a):** There is a subset $X_A \subseteq V_A$ of size at most n_a such that the variables $(f_A(u) : u \in V_A)$ are independent, and for $u \in V_A - X_A$, $f_A(u)$ is uniformly distributed.
>
> **Requirement 1(b):** There is a subset $X_B \subseteq V_B - \{s\}$ of size at most n_b such that the random variables $(f_B(v) : v \in V_B)$ are independent, and $f_B(v)$ for $v \in V_B - X_B$ is uniformly distributed. Note that $f_B[s]$ is automatically uniformly distributed, because initially $F_B[s]$ contains the uniform superposition, and U_B acts safely on $F_B[s]$.

Communication: After U_A, U_B have been applied, Alice and Bob follow a quantum protocol exchanging further messages $M_2, \ldots, M_k$ of lengths $c_2 n, \ldots, c_k n$. Bob sends the message M_2. The rest of the protocol is required to act safely on registers F_A, F_B. At the end of the protocol, the player who receives M_k places a qubit in a special register Ans. The protocol then terminates.

The probability of error: Once the protocol has terminated, all registers are measured. Let ans denote the value observed in Ans, and let f_A and f_B be the values observed in F_A and F_B; we treat f_A and f_B as functions (from V_A to V_B and V_B to V_A respectively). Let

$$f \triangleq f_A \cup f_B.$$

Note that ans and f are random variables.

> **Requirement 2:** $\Pr[\mathsf{ans} = \mathrm{lsb}(f^{(k+1)}(s))] \geq 1 - \epsilon$.

Base case: In $Q_0^A(\epsilon)$, there is no input generation phase or communication. Bob and Alice start as before, with $|\mu\rangle$ in Bob’s register $F_B[s]$. Alice produces a guess ans for $\mathrm{lsb}(f_B(s))$, which must be correct with probability at least $1 - \epsilon$. Clearly, we have the following base case for our induction.

Proposition 5.1 *If $Q_0^A(\epsilon)$ is true then $\epsilon \geq \frac{1}{2}$.*

Our goal is to show that if Q_k^A holds, then $c_1 + c_2 + \ldots + c_k = \Omega(k^{-4})$. By the following lemma, this implies a lower bound $\frac{n}{k^2}$ for P_k^A.

Lemma 5.7 *If there is a safe quantum protocol for* P_k^A *with* $v_0 = s \in V_B$*, messages of lengths* $c_1 n, \ldots, c_k n$*, and worst case error at most* $\frac{1}{4}$*, then* $Q_k^A(c_1, \ldots, c_k, n_A = 0, n_B = 0, \frac{1}{4})$ *is true.*

Proof: We are given a safe quantum protocol $\mathcal{P}$ for P_k^{bit}, where Alice sends the first message M_1. Consider the operation of $\mathcal{P}$ when uniform superpositions are fed for F_A and F_B. Consider the state of Alice just before M_1 is sent to B. This state has two parts.

1. The qubits that Alice keeps with herself, $F_A W_A$, where F_A is $n \log n$ qubits long.

2. The $c_1 n$ qubits that constitute the message M_1.

Let $\widetilde{M_1} M_1$ contain a canonical purification of M_1, where $\widetilde{M_1}$ is $c_1 n$ qubits long. Clearly, it is within Alice's powers to first generate the canonical purification in $\widetilde{M_1} M_1$, and then apply a unitary transformation U_A on $\widetilde{M_1}$ plus some initially zero ancilla qubits in order to generate the correct state of $F_A W_A M_1$. Alice then sends M_1

In our protocol, on Bob's side, $F_B[s]$ already has a uniform superposition in tensor with the rest of Alice's and Bob's qubits. Then, Bob generates the rest of his "input", $F_B[v], v \neq s$ as a uniform superposition in tensor with everything else. The registers W_B are set to $|0\rangle$. At this point, the state of $F_A W_A M_1 F_B W_B$ is exactly the same as it would be in $\mathcal{P}$ after Bob receives the first message. From now on, Alice and Bob operate exactly as in $\mathcal{P}$, which is "safe" on F_A, F_B. The above parameters for Q_k^A can now be verified easily. ■

The following lemma is the key to our inductive argument.

Lemma 5.8 (Round elimination) *(a) For* $k \geq 2$*, if* $Q_k^A(c_1, \ldots, c_k, n_A, n_B, \epsilon)$ *holds (with* $n_A < n$*) then* $Q_{k-1}^B(c_1 + c_2, c_3, \ldots, c_k, n_A, n_B + 1, \epsilon')$ *holds with*

$$\epsilon' = \left(\frac{n}{n - n_a}\right) \left[\epsilon + 3((\ln 2) c_1)^{\frac{1}{2}}\right].$$

(b) If $Q_1^A(c_1, n_A, n_B, \epsilon)$ *holds (with* $n_A < n$*), then* $Q_0^A(\epsilon')$ *holds, where* ϵ' *is exactly as in part (a).*

The next subsection is devoted to the proof of this lemma. Now, let us assume this lemma and prove our main lower bound.

Theorem 5.2 *Suppose* $k \leq n^{\frac{1}{4}}$ *and* $Q_k^A(c_1, \ldots, c_k, 0, 0, \frac{1}{4})$ *holds. Then*

$$c_1 + c_2 + \cdots + c_k = \Omega(k^{-4}).$$

Proof: (Sketch) By $k - 1$ applications of Part (a) of Lemma 5.8 (a) and one application of Part (b), we conclude that either $Q_0^A(\epsilon')$ or $Q_0^B(\epsilon')$ holds with

$$\epsilon' \leq \left(\frac{n}{n-k}\right)^k \left[\frac{1}{4} + 3k((2\ln 2)(c_1 + c_2 + \cdots + c_k))^{\frac{1}{4}}\right].$$

Our theorem follows immediately from this and Proposition 5.1. ■

Now, by using Lemma 5.7, we can derive from this our lower bound for P_k^{bit}.

Corollary 5.4 (Main result) *In any protocol for P_k^{bit}, Alice and Bob must exchange a total of $\Omega(\frac{n}{k^2})$ qubits.*

5.2.2 Round elimination: proof of Lemma 5.8

We consider Part (a) first. Part (b) follows using similar argument, and we do not describe them explicitly. Suppose $Q_k^A(c_1, c_2, \ldots, c_k, n_A, n_B, \epsilon)$ is true. That is, there is a protocol $\mathcal{P}$ satisfying Requirements 1 and 2 in the definition of Q_k^A. We need to show that there is a protocol that satisfies the requirements for Q_{k-1}^B with parameters stated in Lemma 5.8 (a).

In what follows, subscripts of pure and mixed states will denote the registers which are in those states. For $u \in V_A$, we use the subscript u instead of $F_A[u]$. Similarly, for $v \in V_B$, we use the subscript v instead of $F_B[v]$. For example, we say that the register $F_B[s]$ is initially in the state

$$|\mu\rangle_s = \frac{1}{\sqrt{n}} \sum_{u \in V_A} |u\rangle_s.$$

Let $|\psi^A\rangle$ be the (pure) state of Alice's registers just before she sends M_1 to Bob. At this point the state of all the registers taken together is the pure state

$$|\psi_{\text{in}}\rangle = |\psi^A\rangle \otimes \frac{1}{\sqrt{n}} \sum_{a \in V_A} |a\rangle_s |\mathbf{0}\rangle_R, \tag{5.4}$$

where R is the set of registers corresponding to the rest of B's input $(F_B[v] : v \in V_B - \{s\})$, and work qubits W_B. For $a \in V_A$, we may expand $|\psi^A\rangle$ as

$$|\psi^A\rangle = \frac{1}{\sqrt{\ell_a}} \sum_{b \in V_B} |b\rangle_a |\psi^A_{a\to b}\rangle, \tag{5.5}$$

where $\ell_a = 1$ if $a \in X_A$ and $\ell_a = n$ otherwise. Here, $|\psi^A_{a\to b}\rangle$ is a pure state of Alice's registers $(F_A(v) : v \in V_A - \{a\})$ and W_A. Note that $|\psi^A_{a\to b}\rangle$ is precisely the state of these registers when $F_A[a]$ is measured and found to be in state $|b\rangle$. (If $\Pr[f_A[a] = b] = 0$, then $|\psi^A_{a\to b}\rangle \triangleq 0$.) From (5.4) and (5.5), we have

$$|\psi_{\text{in}}\rangle = \frac{1}{\sqrt{n}} \sum_{a \in V_A} \frac{1}{\sqrt{\ell_a}} \sum_{b \in V_B} |b\rangle_a |\psi^A_{a\to b}\rangle |a\rangle_s |\mathbf{0}\rangle_R \tag{5.6}$$

At this point the first message M_1 is sent to Bob. Let the rest of the protocol starting from this point be $\mathcal{P}'$; that is, in $\mathcal{P}'$ Bob starts by generating his input from M_1 and $F_B[s]$, sends the message M_2 to A, to which Alice responds with M_3, and so on. At the end of $\mathcal{P}'$ we have a register containing the answer which we measure to find ans, and the input registers of Alice and Bob, which when measured yield f_A and f_B.

Let $\epsilon_{a\to b}$ be the probability of error when $\mathcal{P}'$ is run starting from the state

$$|b\rangle_a |\psi^A_{a\to b}\rangle |a\rangle_s |\mathbf{0}\rangle_R.$$

Thus, we have

$$\epsilon_{a\to b} = \Pr[\mathsf{ans} \neq \mathrm{lsb}(f^{(k+1)}(s)) \mid f_B[s] = a \text{ and } f_A[a] = b],$$

in the original protocol $\mathcal{P}$ (or in $\mathcal{P}'$, when it is run starting from $|\psi_{\mathrm{in}}\rangle$). In particular, we have

$$\epsilon = \mathop{\mathrm{E}}_{a,b}[\epsilon_{a\to b}] \geq \frac{n-n_a}{n} \mathop{\mathrm{E}}_{a\in_u V_A - X_A, b\in_u V_B}[\epsilon_{a\to b}]. \tag{5.7}$$

In the first expectation, (a,b) are chosen with the same distribution as $(f_B[s], f_A[f_B[s]])$ of the given protocol $\mathcal{P}$; in the second, they are chosen uniformly from the sets specified.

Overview: We want to eliminate the first message sent by Alice, at the cost of increasing the probability of error slightly, but preserving the total length of the communication. This is based on the following idea (taken from [KNTZ01b]). Let $M_{1,a\to b}$ be the state of the registers holding the first message when the entire state of Alice's registers is $\psi^A_{a\to b}$; that is, $M_{1,a\to b}$ is the state of the message registers corresponding to message M_1, when we measure $F_A[a]$ and observe $|b\rangle$ there. Note, that $\psi^A_{a\to b}$ is a purification of $M_{1,a\to b}$. Also, the state of the first message in $\mathcal{P}$, M_1 is the average, taken over the choices of b, of $M_{1,a\to b}$.

Suppose there is an $a \in V_A - X_A$ such that for all b, the message $M_{1,a\to b}$ is independent of b, that is, it is always the fixed sate M^*. Then, we can eliminate the first message. Informally stated, this amounts to restricting ourselves to the sub case of the protocol when Bob's first pointer $F_B[s]$ is fixed at $|a\rangle$, and Bob generates M^* himself, and sends some small advice along with his message M_2, to enable Alice to reproduce the right entanglement between her registers and Bob's. Unfortunately, we will not be able to show that there is an a and an M^* such that $M_{1,a\to b} = M^*$, for all b. Instead, we will show that there is an M^* that will be close to $M_{1,a\to b}$ for typical b. In fact, the message M_1 (which is the average of $M_{1,a\to b}$ as b varies) will be our M^*.

Let $(M_1, \widetilde{M_1})$ be the canonical purification of the first message of the protocol $\mathcal{P}$. Our first goal is to show that if M_1 is close to $M_{1,a\to b}$, then Alice can create a state close to $|\psi^A_{a\to b}\rangle$ from $(M_1, \widetilde{M_1})$ by applying a unitary transformation on $\widetilde{M_1}$. More precisely, suppose

$$1 - B(M_{1,a\to b}, M_1) \triangleq \delta_{a\to b}.$$

Then, by the Local Transition Theorem(Theorem 5.3), there is a unitary transformation $U_{a\to b}$ that when applied to $\widetilde{M_1}$ (together with ancilla qubits initialised to zero) takes the pure state $(M_1, \widetilde{M_1})$ to a state $\tilde{\psi}^A_{a\to b}$ such that

$$\left\| |\psi^A_{a\to b}\rangle\langle\psi^A_{a\to b}| - |\tilde{\psi}^A_{a\to b}\rangle\langle\tilde{\psi}^A_{a\to b}| \right\|_t \leq 2\sqrt{2\delta_{a\to b}}. \tag{5.8}$$

In particular, if the protocol $\mathcal{P}'$ is run starting from the state

$$|\tilde{\psi}^A_{a\to b}\rangle|\mu\rangle_s|\mathbf{0}\rangle_R$$

(instead of $|\psi^A_{a\to b}\rangle|\mu\rangle_s|\mathbf{0}\rangle_R$), the probability of error is at most

$$\epsilon_{a\to b} + 2\sqrt{2\delta_{a\to b}}.$$

5.2.3 The protocol $\mathcal{P}_{a\to b}$

Now, we fix $a \in V_A$ and $b \in V_B$ and consider the case when $f_B(s) = a$ and $f_A(a) = b$. We now describe a protocol that functions for this situation (see Figure 5.1) . This is just an intermediate protocol. Later we will describe how we obtain our final protocol (satisfying the requirements of Q^B_{k-1}) from this. It will be helpful, meanwhile, to keep in mind that in our final protocol, the roles of A and B will be reversed, $F_B[s]$ will be fixed at $|a\rangle$ (we will add s to X_B), a will be our new s, and the state of $F_A[a]$ will not be fixed at $|b\rangle$ but will be the uniform superposition $|\mu\rangle$.

Step 1: Alice generates the canonical purification $(M_1, \widetilde{M_1})$. Alice applies $U_{a\to b}$ to $\widetilde{M_1}$ (plus some ancilla) to produce the state $|\tilde{\psi}^A_{a\to b}\rangle$ in the registers (M_1, F_A, W_A).
Step 2: Alice and Bob proceed according to the protocol $\mathcal{P}'$ starting from the state

$$|\tilde{\psi}_{a\to b}\rangle = |\tilde{\psi}^A_{a\to b}\rangle|a\rangle_s|\mathbf{0}\rangle_R,$$

where, as before, R is the set of registers of Bob corresponding to $(F_B[v] : v \in V_B - \{s\})$ and work qubits W_B .

Figure 5.1: The intermediate protocol $\mathcal{P}_{a\to b}$

Remark on the inputs generated: Suppose we measure registers F_A just after $U_{a\to b}$ has been applied in the above protocol. Let $f'_{A,a\to b}$ be the resulting random variable with distribution $D'_{a\to b}$. On the other hand, if we were to measure the same registers in the state $|\psi^A_{a\to b}\rangle$, then the resulting random variable is $f_{A,a\to b}$ whose distribution is D; that is, D is the distribution of f_A conditioned on the event $f_A[a] = b$. Then, it follows from (5.8) and Theorem 3.4 that

$$\|D_{a\to b} - D'_{a\to b}\|_1 \leq 2\sqrt{2\delta_{a\to b}}. \tag{5.10}$$

We will want Alice's input registers to satisfy Requirement 1(b). Unfortunately, the distribution D' may not satisfy this requirement automatically, but (5.10) will help us 'correct' this.

Next consider Bob's input registers. In $\mathcal{P}$, Bob's register $F_B[s]$ contained the uniform superposition μ and he generated the input in the rest of the registers himself form M_1 using the unitary transformation U_B. The input he generated satisfied Requirement 1(b). In $\mathcal{P}_{a\to b}$, Bob applies the same transformation U_B on M_1, but $F_B[s]$ is now $|a\rangle$ and not $|\mu\rangle$. Suppose F_B is measured at this stage resulting in the random variable $f_{B,a\to b} : V_B \to V_A$. Note that $f_{B,a\to b}$ has the same distribution as f_B conditioned on the event $f_B(s) = a$. Thus,

B1. $f_{B,a\to b}$ is constant on $X_A \cup \{s\}$ (in fact, $f_B[s] = a$), and

B2. the set of random variables $(f_{B,a\to b}[v] : v \in V_B - X_B - \{s\})$ are independent and uniformly distributed over V_A.

Probability of error in $\mathcal{P}_{a\to b}$: By (5.8) and Theorem 3.4, the probability of error of $\mathcal{P}_{a\to b}$, which we denote by $\tilde{\epsilon}_{a\to b}$, is at most

$$\epsilon_{a\to b} + 2\sqrt{2\delta_{a\to b}}.$$

Correcting Alice's input registers: The random variable $f'_{A,a\to b}$ that results from measuring F_A has a distribution $D'_{a\to b}$ which is close to the desired distribution $D_{a\to b}$ of $f_{A,a\to b}$ (by (5.10) above). It will be easier to satisfy Requirement 1(b), however, if we could arrange that the distribution of Alice's inputs is exactly $D_{a\to b}$. To do this, we use Fact 5.2; let $C_{a\to b}$ be the unitary transformation corresponding to $D'_{a\to b}$ and $D_{a\to b}$. We revise the protocol $\mathcal{P}_{a\to b}$ by including this operation (see Figure 5.2).

Error probability of the revised protocol: At that end of the protocol, we measure all registers and obtain the answer ans, and the inputs $\hat{f}_{A,a\to b}$ and $f_{B,a\to b}$. We also have $f_{A,a\to b}$ corresponding to Alice's old input registers F_A. Let

$$\hat{f}_{a\to b} = \hat{f}_{A,a\to b} \cup f_{B,a\to b}$$

and

$$f'_{a\to b} = f'_{A,a\to b} \cup f_{B,a\to b}.$$

This revised protocol makes an error whenever

$$\mathsf{ans} \neq \mathrm{lsb} \hat{f}^{(k+1)}_{a\to b}(s).$$

We then have

$$\begin{aligned}
\hat{\epsilon}_{a\to b} &\triangleq \Pr[\mathsf{ans} \neq \mathrm{lsb} \hat{f}^{(k+1)}_{a\to b}(s)] \\
&\leq \Pr[\hat{f}_{a\to b} \neq f'_{a\to b}] + \Pr[\mathsf{ans} \neq \mathrm{lsb} f'^{(k+1)}_{a\to b}(s)] \\
&\leq \frac{1}{2}\cdot 2\sqrt{2\delta_{a\to b}} + \epsilon_{a\to b} + 2\sqrt{2\delta_{a\to b}} \\
&= \epsilon_{a\to b} + 3\sqrt{2\delta_{a\to b}}. \qquad (5.11)
\end{aligned}$$

5.2.4 The final protocol: $\mathcal{P}_a$

A small modification now gives us our final protocol, which will satisfy the requirements for Q^{k-1}. We make two changes to the revised version of $\mathcal{P}_{a\to b}$. First, instead of Alice sending M_1 and retaining $\widetilde{M_1}$, now Bob creates the canonical purification $(M_1, \widetilde{M_1})$ and sends Alice $\widetilde{M_1}$, while retaining M_1. Second, in $\mathcal{P}_{a\to b}$, the register $\hat{F}_A[a]$ is fixed to the value $|b\rangle$. Now, however, Alice starts with $|\mu\rangle$ in $\hat{F}[a]$. With these modifications, Alice's role in the input generation phase of the new protocol is similar to Bob's role in the protocol we started with. The resulting protocol $\mathcal{P}_a$ (see Figure 5.3) depends on the choice of a. Using an averaging argument we will conclude that there is a choice for $a \in V_A$ so that $\mathcal{P}_a$ satisfies the requirements for Q^B_{k-1} as needed in Lemma 5.8(a).

The probability of error of $\mathcal{P}_a$: For $a \in V_A - X_A$, let $\hat{\epsilon}_a$ be the probability of error of $\mathcal{P}_a$. Then, by (5.11), we have

$$\begin{aligned} \hat{\epsilon}_a &= \mathop{\mathrm{E}}_{b \in_u V_B} [\hat{\epsilon}_{a \to b}] & (5.12) \\ &\leq \mathop{\mathrm{E}}_{b \in_u V_B} [\epsilon_{a \to b} + 3\sqrt{2\delta_{a \to b}}]. & (5.13) \end{aligned}$$

We need to show that there exists an a such that $\hat{\epsilon}_a$ is small. For this we consider the average of $\hat{\epsilon}_a$ as a is chosen uniformly from $V_A - X_A$:

$$\mathop{\mathrm{E}}_{a \in_u V_A - X_A} [\hat{\epsilon}_a] \leq \mathop{\mathrm{E}}_{a,b}[\epsilon_{a \to b} + 3\sqrt{2\delta_{a \to b}}], \tag{5.14}$$

where on the right a is chosen uniformly from $V_A - X_A$ and b is chosen independently and uniformly from V_B. (From now on, when we average over a and b, we will assume that they are chosen in this manner.) By (5.7), we have

$$\mathop{\mathrm{E}}_{a,b}[\epsilon_{a \to b}] \leq \left(\frac{n}{n - n_a}\right) \epsilon. \tag{5.15}$$

It remains to bound

$$\mathop{\mathrm{E}}_{a,b}[\sqrt{\delta_{a \to b}}].$$

Consider the state obtained by measuring Alice's input registers F_A just before M_1 is sent to Bob in the original protocol. As stated earlier, if the value b is observed for $F_A[a]$, then the state of the message registers will be $M_{1,a \to b}$; also, M_1 is the average of these states, that is,

$$M_1 = \frac{1}{n} \sum_{b \in V_B} M_{1,a \to b}.$$

Claim 5.1 *For $a \in V_A - X_A$,*

$$\mathop{\mathrm{E}}_{b}[\delta_{a \to b}] \leq (\ln 2/2) I(f_A[a] : M_1).$$

Proof: Consider the encoding of elements of V_B given by $b \mapsto M_{1,a \to b}$ by restricting attention the registers $F_A[a]$ and M_1. Our claim now follows from the Average Encoding Theorem (Theorem 5.2) and the definition of $\delta_{a \to b}$. ■

Claim 5.2

$$\mathop{\mathrm{E}}_{a \in_u V_A - X_A} [I(f_A[a] : M_1)] \leq \left(\frac{n}{n - n_a}\right) c_1.$$

Proof: Using Fact 2.4 and (5.10), we have

$$\begin{aligned} c_1 n &\geq I(f_A : M_1) \\ &\geq \sum_{a \in V_A} I(f_A[a] : M_1) \geq \sum_{a \in V_A - X_A} I(f_A[a] : M_1). \end{aligned}$$

■

By combining these two claims, and noting that the square-root function is concave, we obtain

$$\begin{aligned} \mathop{\mathrm{E}}_{a,b}[\delta_{a \to b}] &\leq \mathop{\mathrm{E}}_{a}[(\ln 2/2) I(f_A[a] : M_1)] \\ &\leq (\ln 2/2) \mathop{\mathrm{E}}_{a}[I(f_A[a] : M_1)] \leq \left(\frac{n}{n - n_a}\right)(\ln 2/2) c_1. \end{aligned}$$

This implies, again because the square root is concave, that

$$\mathop{\mathrm{E}}_{a,b}[\sqrt{\delta_{a \to b}}] \leq \left[\left(\frac{n}{n - n_a}\right)(\ln 2/2) c_1\right]^{\frac{1}{2}}. \tag{5.16}$$

Now we return to (5.14), and use (5.15) and (5.16) to obtain

$$\mathop{\mathrm{E}}_{a}[\hat{\epsilon}_a] \leq \left(\frac{n}{n - n_a}\right)\left[\epsilon + 3((\ln 2) c_1)^{\frac{1}{2}}\right].$$

Thus, there exists an $a \in V_A - X_A$ such that

$$\hat{\epsilon}_a \leq \left(\frac{n}{n - n_a}\right)\left[\epsilon + 3((\ln 2) c_1)^{\frac{1}{2}}\right].$$

Now, it can be verified, the protocol $\mathcal{P}_a$ satisfies the requirements for $Q_{k-1}^B(c_1 + c_2, c_3, \ldots, c_k, n_A, n_B + 1, \hat{\epsilon}_a)$. This shows Part (a) of Lemma 5.8. Part (b) can be established similarly.

Revised Step 1:

- Alice generates the canonical purification $(M_1, \widetilde{M_1})$. Alice applies $U_{a\to b}$ to $\widetilde{M_1}$ (plus some ancilla) to produce the state $|\psi^A_{a\to b}\rangle$ in the registers (M_1, F_A, W_A). Alice sends M_1 to Bob.

- Next, to produce input registers satisfying Requirement 1(a), Alice uses a fresh set of registers $\hat{F}_A$ and sets $\hat{F}_A[a] = |b\rangle$. Next, Alice applies a unitary transformation to registers $(\hat{F}_A[a], F_A, \tilde{F}_A)$ defined by

$$|b\rangle_{\hat{F}[a]}|\psi\rangle_{F_A,\tilde{F}_A} \to |b\rangle_{\hat{F}[a]}C_{a\to b}|\psi\rangle_{F_A,\tilde{F}_A}.$$

Before the application of this the registers $\tilde{F}_A$ are initialised to $|\mathbf{0}\rangle$ (as in the statement of Fact 5.2). Alice then copies $(\tilde{F}_A[u] : u \in V_A - \{a\})$ into $(\hat{F}_A[u] : u \in V_A - \{a\})$. The input generation for Alice is now complete.

Note that at this point if we measure $(F_A, \hat{F}_A)$, the resulting random variables $(f'_{A,a\to b}, \hat{f}_{A,a\to b})$ have distribution precisely $D'_{a\to b}$ and $D_{a\to b}$. Furthermore, (see Fact 5.2),

$$\Pr[f'_{A,a\to b} \neq \hat{f}_{A,a\to b}] \leq \frac{1}{2} \cdot 2\sqrt{\delta_{a\to b}}. \tag{5.9}$$

Step 2: From this point on, Alice and Bob just follow $\mathcal{P}'$ described above. On receiving M_1, Bob generates his input and work qubits by appropriately applying the unitary transformation U_B. He then generates message M_2 and sends it to Alice.

Let $|\phi_{a\to b}\rangle$ denote the state of the entire system just after M_2 is sent to Alice.

After this, Alice and Bob continue as before. In particular, the Alice continues to use her old input register F_A (safely) as before. The registers $\hat{F}$ are not used until the end, when they are measured in order to decide if the answer returned by the protocol is correct.

Figure 5.2: The revised protocol $\mathcal{P}_{a\to b}$

The new input registers for Alice will be denoted by $\hat{F}_A$. The old input registers will continue to exist, but they will count as work qubits of Alice. Initially, in the register $\hat{F}_A[a]$ we place a uniform superposition $|\mu\rangle$. All other registers are initialised to 0.

Step 1: Bob generates the canonical purification $(M_1, \widetilde{M_1})$ of the first message of $\mathcal{P}$. He sets his register $F_B[s]$ to the state $|a\rangle$, and using the transformation U_B generates his inputs F_B and work qubits W_B. Then, he generates the first message of protocol $\mathcal{P}'$ (this corresponds message M_2 of the $\mathcal{P}$), and sends this message along with $\widetilde{M_1}$ to Alice.

Step 2: (a) One receiving $\widetilde{M_1}$, Alice applies a unitary transform on registers $(\hat{F}_A[a], \widetilde{M_1}, A)$ to generate a state in registers F_A (the old input registers) and W_A (the work qubits of the original protocol). Here, A is a set of ancilla qubits initialised to 0. This unitary transformation acts according to the rule

$$|b\rangle_{\hat{F}[a]}|\theta\rangle_{\widetilde{M_1},A} \mapsto |b\rangle_{\hat{F}[a]}U_{a\to b}|\theta\rangle_{\widetilde{M_1},A}.$$

Note that this transformation is safe on $\hat{F}[a]$.

(b) Since F_A is not in the desired state, Alice applies the correction used in the revised Step 1 of $\mathcal{P}_{a\to b}$. That is, she applies a unitary transformation to registers $(\hat{F}_A[a], F_A, \tilde{F}_A)$ defined by

$$|b\rangle_{\hat{F}[a]}|\psi\rangle_{F_A,\tilde{F}_A} \mapsto |b\rangle_{\hat{F}[a]}C_{a\to b}|\psi\rangle_{F_A,\tilde{F}_A}.$$

Before the application of this the registers $\tilde{F}_A$ are initialised to 0. Alice then copies $(\tilde{F}_A[u] : u \in V_A - \{a\})$ into $(\hat{F}_A[u] : u \in V_A - \{a\})$. For the purpose of satisfying Requirement 1(b), $\hat{F}_A$ are to be treated as A's input register.

(continued on next page)

Figure 5.3: The protocol $\mathcal{P}_a$

The state of entire system at this point is precisely

$$\frac{1}{\sqrt{n}} \sum_{b \in V_B} |\phi_{a \to b}\rangle,$$

where $|\phi_{a \to b}\rangle$ is the state at the corresponding point in the revised protocol $\mathcal{P}_{a \to b}$ (see Figure 5.2). The rest of the protocol operates safely on F_A,$\hat{F}_A$ and F_B. In fact, no unitary transform will now be applied to registers $\hat{F}_A$.

Step 3: Alice resumes the protocol $\mathcal{P}'$. Note that Bob has already executed the first step of $\mathcal{P}'$ and sent the first message (which corresponds to message M_2 of the original protocol). Alice responds to this message as before.

While executing $\mathcal{P}'$, the old input registers F_A are used. The new registers $\hat{F}_A$ are not touched by any unitary transformation from now on. At the end, however, when we try to decide if an error has been made, we will measure all registers, and check if the answer ans' agrees with the answer $\mathsf{ans}(\hat{f}_A, f_B)$, where $\hat{f}_A$ is the random variable obtained by measuring the new input registers $\hat{F}_A$.

Figure 5.4: The protocol $\mathcal{P}_a$

Chapter 6

The set disjointness problem

6.1 Lower bound for set disjointness

In this chapter we prove a lower bound for the quantum communication complexity of set disjointness. We first broadly mention the techniques we have used to arrive at the lower bound.

Techniques used

The original lower bounds for set disjointness in the classical two-party communication model are based on deep analyses of the communication matrix and can be said to be based on the *discrepancy method* (see e.g. [KN97]). Razborov's recent $\Omega(\sqrt{n})$ lower bound [Raz02] for the bounded error two-party quantum communication complexity of set disjointness also uses the discrepancy method. The discrepancy method for quantum protocols was formulated explicitly by Kremer [Kre95] (see also Klauck [Kla01] and Yao [Yao93]), but Razborov's proof extends it substantially by developing interesting and powerful tools based on the spectral theory of matrices.

Recently however, Bar-Yossef et al. [BJKS02] proposed an information-theoretic approach for studying set disjointness-like problems in the classical setting. Using a refinement of the notion of *information cost* of a communication protocol originally defined by Chakrabarti, Shi, Wirth and Yao [CSWY01], they showed that a linear lower bound for the bounded error two-party randomised communication complexity of set disjointness follows from an $\Omega(1)$ lower bound on a certain *information cost* of a two-party communication protocol computing the AND $a \wedge b$ of just two bits a, b! The information-theoretic machinery essentially allowed them to treat the set disjointness function like a direct sum of n two-bit AND's. Their work provided a compelling and beautiful illustration of information-theoretic tools in the analysis of communication protocols. Interestingly, the idea of proving lower bounds for set disjointness by treating it like a direct sum of n two-bit AND's was earlier employed in [KKN95] in the setting of two-party nondeterministic classical communication complexity; however, their approach was not information-theoretic and does not seem to be suitable for bounded error classical randomised or quantum communication protocols.

We adapt their approach to the quantum setting. In order to bring out the contribution of this work more clearly, we will now informally describe the information-theoretic argument underlying the proof of [BJKS02] and discuss how we adapt it to the quantum setting. The argument has two parts: in the first part, using a direct-sum property for information cost of a communication protocol one reduces the communication problem DISJ to the communication problem AND of two bits (one with Alice and one with Bob); in the second part, one shows that any communication protocol for AND of two bits needs to have high information cost.

The information cost approach: The first part of the argument is based on the notion of *information cost* of private coin randomised communication protocols, defined to be the mutual information between the inputs (which are assumed to come from some distribution) and the entire message transcript of the protocol. Bar-Yossef et al. [BJKS02] examine the information cost of the protocol for several distributions. Let the number of bits transmitted by the protocol be c. Then, the information cost is also bounded by c for each distribution.

At this point it will be convenient to view the inputs X_A and X_B of Alice and Bob as elements of $\{0,1\}^n$ and the set disjointness function DISJ as

$$\bigvee_{i=1}^{n} X_A[i] \wedge X_B[i].$$

A typical distribution considered by Bar-Yossef et al. is defined as follows. For each coordinate i, independently, one party is given the input 0 and the other party is given a uniformly random bit. Using the sub-additivity property of mutual information, one concludes that the sum over i of the mutual information between the transcript and $X_A[i]$ is bounded by c; a similar statement holds for Bob's inputs. It is then not hard to argue using a standard averaging argument that there is an i and a probability distribution D^* on $(X_A[j], X_B[j] : j \neq i)$ such that the following conditions hold:

- $X_A[j], X_B[k], j \neq i, k \neq i$ are independent random variables under D^*;
- For all $j \neq i$, $X_A[j] \wedge X_B[j] = 0$ (with probability 1);
- If $X_A[i]$ is set to 0, $X_B[i]$ is chosen uniformly at random from $\{0,1\}$ and $(X_A[j], X_B[j] : j \neq i)$ are chosen according to D^*, then the mutual information between the message transcript and $X_B[i]$ is at most $2c/n$; similarly, if $X_B[i]$ is set to 0, $X_A[i]$ is chosen uniformly at random from $\{0,1\}$ and $(X_A[j], X_B[j] : j \neq i)$ are chosen according to D^*, then the mutual information between the message transcript and $X_A[i]$ is at most $2c/n$.

From the first condition, by viewing $(X_A[j], X_B[j] : j \neq i)$ as private coins of the two parties, we obtain from the protocol for DISJ a protocol that computes the AND of the two bits $X_A[i]$ and $X_B[i]$. The stage is thus set for analysing the information cost of a protocol computing the AND of two bits: a lower bound of ϵ on this quantity translates to a lower bound of $\Omega(\epsilon n)$ on the communication complexity of the set disjointness function.

In order to implement this programme in the quantum setting, one has to define a notion of information cost for quantum protocols. It is not immediately clear how this can be done, because quantum operations are notorious for destroying the states on which they act; in particular, it is not reasonable to expect that the complete transcript of all messages is part of the final global state of the algorithm. Even if the complete transcript is available in the final global state of the algorithm, it may not contain any information about the inputs of either party. For example if the parties are allowed prior entanglement, then using quantum teleportation, one can implement any protocol such that the messages are classical and uniformly random. So, the transcript will just be a uniformly random string of length c independent of the actual inputs!

The definition of information loss for quantum protocols: We address these difficulties as follows. Assume that the players' inputs come from some classical probability distribution. Without loss of generality, the players make a 'safe' copy of their (classical) inputs before proceeding with the quantum protocol. Instead of considering the information carried by a particular message, we examine the the context in which the message is received i.e. we consider the von Neumann mutual information between the sender's input and all the qubits in the possession of the receiver at that time, including the qubits of the message just received. The *information loss* (we use the term loss instead of cost) of the protocol for the given input distribution is defined to be a certain weighted sum of these mutual informations taken over all rounds. With this definition of information loss, the arguments of [BJKS02] are easily carried over to the quantum setting. We can then conclude that if the information loss of computing the AND of two bits is ϵ then the communication complexity of DISJ is $\Omega(n\epsilon/k)$.

We have arrived at the second part of our programme, that is, to show non-trivial lower bounds on the information loss of a quantum protocol computing the AND of two bits. In their original argument, [BJKS02] showed a lower bound on the information cost of a classical private coin protocol computing the AND of two bits via a direct argument using *Hellinger distances* between certain probability distributions. Since we are working with our different notion of information loss, this argument does not appear to be immediately applicable to us; so instead of reviewing it, we will now directly describe our new argument for showing a lower bound on the information loss of a quantum protocol computing the AND of two bits. We consider two input distributions: in the first distribution, Alice has 0 and Bob has a uniformly random bit; in the second distribution, Bob has 0 and Alice has a uniformly random bit. Suppose we are given that for these distributions at no stage do the qubits of the receiver of a message contain more than ϵ bits of information about the input of the sender. We wish to show that if ϵ is very small, then this leads to a contradiction. Our argument can be understood at an intuitive level in the framework of *round elimination* in communication protocols [MNSW98b, KNTZ01a, Sen03]. Suppose Alice sends the first message of the protocol. We know that when Bob's input is 0 the state of his qubits after receiving the first message is essentially the same whether Alice's input is 0 or 1. So no matter what her actual input is, Alice might as well send her first message assuming that her input is 0. Using standard arguments (see below), we can eliminate the first message of Alice and obtain a protocol with one fewer round of communication, increasing the error probability of the protocol by a small amount. Now it is Bob's turn. Our hypothesis says that when Alice's input is 0 the state of her qubits after receiving the first message from Bob is essentially the same whether Bob's input is 0 or 1. But the modified protocol so far has proceeded as if Alice's input is 0 (even though her actual input might be something else). We can thus eliminate Bob's first message as well. If ϵ is small, then the increase in error probability on account of this manoeuvre is also small. Proceeding in this manner we eliminate all rounds. But it is obvious that if the parties exchange no messages they cannot compute any non-trivial function unless one allows error probability greater than or equal to 1/2. Since there are at most k rounds of communication, this gives us a lower bound of the form $\epsilon \geq \epsilon(k)$. Using these ideas one can show an $\Omega(n/k^2)$ lower bound on the two-party quantum communication complexity of the set disjointness function.

There are two aspects of our proof that require further comment.

Local transition: Recall the argument used above to eliminate Alice's first message. We know that when Bob's input is 0, the state of his qubits after receiving the first message is roughly the same whether Alice's input is 0 or 1. However, this does not immediately imply that the error probability of the modified protocol is not changed by much. The final answer is not just a function of Bob's state but the combined state of Alice and Bob. In particular, even though Bob's state is similar after the first round for the two inputs of Alice, his work qubits might be entangled with Alice's qubits differently in the two cases. This problem arises often in round elimination arguments and by now standard solutions exist for it by considering the *fidelity* between quantum states. This allows Alice to perform a *local transition* [KNTZ01a] on her work qubits, in order to restore them to the correct state should she discover later that her actual input is 1 (recall that in the modified protocol, Alice prepares her first message assuming that her input is always 0). We use a stronger local transition lemma (Lemma 5.3) than the one in [KNTZ01a]. The stronger lemma is crucial for getting an $\Omega(n/k^2)$ lower bound in Result 1.2; the local transition lemma of [KNTZ01a] gives an $\Omega(n/k^4)$ lower bound.

A paradox?: In our notion of information loss of quantum protocols it is important that the parties start in a *pure* global state. In fact, this notion is unsuited for classical private coin randomised communication complexity. Consider the following classical private coin protocol for computing the AND of two bits (a, b). Alice sends Bob a random bit r, retaining a copy of r if and only if $a = 1$. Bob sends Alice $r \oplus b$; if $a = 1$, Alice can recover b using the copy of r she has and determine $a \wedge b$. Now clearly, when Bob's input is 0 he has no information about Alice's input at the end of the first round; also when Alice's input is 0 she has no information about Bob's input at the end of the second round because she does not retain a copy of r in this case. So, according to our definition this protocol has zero information loss for both the distributions considered above. Yet, the protocol computes the AND of two bits correctly! Interestingly, no such quantum protocol starting with a pure global state is possible.

Below we define some important information theoretic concepts needed in our proof.

6.1.1 Information cost

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

Definition 6.1 (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 minimum information cost of a k-round δ-error protocol for f under μ. $\mathrm{IC}^{\mathrm{sim}}_{\delta}(f)$ denotes the minimum information cost of a private coin simultaneous δ-error protocol for f under the uniform probability distribution on the inputs.*

Remark: $\mathrm{IC}^{\mathrm{sim}}_{\delta}(f)$ as defined above coincides with the definition of information cost of a simultaneous message protocol in Chakrabarti et al. [CSWY01].

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 $\delta > 0$,

$$IC^k_{\mu^m,\delta}(f^m) \geq m \cdot IC^k_{\mu,\delta}(f).$$

Unfortunately for non-product distributions μ, no such nice sub-additivity theorem is known. To get over this shortcoming, 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 6.2 (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 minimum conditional information cost of a k-round δ-error protocol for f under (μ, κ).

The following facts follow easily from the results in Bar-Yossef et al. [BJKS02] and Fact 2.1.

Fact 6.1 *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 6.2 *With the notation and assumptions of Fact 6.1,*

$$C^k_{\mu,\delta}(f) \geq IC^k_{\mu,\delta}(f \mid \kappa).$$

The following lemma relates the t-party k-round δ-error communication complexity of set disjointness to the conditional information loss of t-party k-round δ-error *AND* function.

6.1.2 Conditional information loss

We now define the *conditional information loss* of a t-party quantum communication protocol with prior entanglement. For technical reasons, we need to work with a *conditional* version of information loss instead of the unconditional version described in the introduction. A similar *conditional* version of information cost is used in [BJKS02] to prove their lower bounds. But first, we need a couple of preliminary definitions.

Definition 6.3 (Embedding) *For $\mathbf{x} \in \mathcal{X}^n$, $j \in [n]$, and $x \in \mathcal{X}$, let $\mathsf{embed}(\mathbf{x}, j, x)$ be the element of $\mathcal{X}^n$ obtained by replacing $\mathbf{x}[j]$ by x, that is, $\mathsf{embed}(\mathbf{x}, j, x)[\ell] \stackrel{\Delta}{=} \mathbf{x}[\ell]$ for $\ell \neq j$, and $\mathsf{embed}(\mathbf{x}, j, x)[j] \stackrel{\Delta}{=} x$.*

Definition 6.4 (Collapsing) *Suppose $F : \mathcal{X}^n \to \mathcal{Z}$. We say that $\mathbf{x} \in \mathcal{X}^n$ collapses F to the function $h : \mathcal{X} \to \mathcal{Z}$ if for all $u \in \mathcal{X}$, $j \in [n]$, $F(\mathsf{embed}(\mathbf{x}, j, u)) = h(u)$. We say that a random variable $\mathbf{X}$ taking values in $\mathcal{X}^n$ collapses F to h if it collapses F to h with probability* 1.

Let D, $X_1, \ldots, X_t$ be classical random variables taking values in some finite sets $\mathcal{D}$, $\mathcal{X}_1, \ldots, \mathcal{X}_t$ respectively. Let $X \triangleq (X_1, \ldots, X_t)$. We say that D *partitions* X, if for all possible values d that D can take, $X_1, \ldots, X_t$ are independent conditioned on the event $D = d$. The random variable $(X, D)^n$ is obtained by taking n independent copies of (X, D). Thus, X^n takes values in $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_t)^n$ which we identify with $\mathcal{X}_1^n \times \cdots \times \mathcal{X}_t^n$ in the natural way. Suppose D partitions X, and $(\mathbf{X}, \mathbf{D}) \triangleq (X, D)^n$; then it is easy to verify that $\mathbf{D}$ partitions $\mathbf{X}$. Let Π be a t-party k-round δ-error quantum protocol for computing $F : \mathcal{X}_1 \times \cdots \times \mathcal{X}_t \to \mathcal{Z}$. Suppose $X_1, \ldots, X_t$ are the random variables corresponding to the inputs of $\mathcal{P}_1, \ldots, \mathcal{P}_t$. Let $\mathcal{P}^j$ denote the active player in round j. Let X^j denote the input random variable of $\mathcal{P}^j$. $\hat{P}^j$ denote the qubits all players except $\mathcal{P}^j$ just after round j is complete. Let $k(j)$ denote the number of rounds of Π in which the player $\mathcal{P}^j$ is active.

Definition 6.5 (Conditional information loss) *In the notation above, the conditional information loss of* Π *under* (X, D) *is defined by*

$$\mathsf{IL}(\Pi \mid (X, D)) \triangleq \sum_{j=1}^{k} \frac{k}{k(j)} \cdot I((X^j : \hat{P}^j) \mid D).$$

The t-party k-round δ-error conditional information loss of F under (X, D), denoted by $\mathsf{IL}_\delta^{t,k}(F \mid (X, D))$, is the infimum $\mathsf{IL}(\Pi \mid (X, D))$ taken over all t-party k-round δ-error quantum protocols with prior entanglement Π for F. [Note that δ upper bounds the error of Π for all *inputs in $\mathcal{X}_1 \times \cdots \times \mathcal{X}_t$. In particular, this error bound applies even to inputs not in the support of X.]*

The following lemma is the first part of our proof as mentioned earlier.

Lemma 6.1 *Let $F : \mathcal{X}_1^n \times \cdots \times \mathcal{X}_t^n \to \mathcal{Z}$. Let $X_1, \ldots, X_t$ be classical random variables taking values in $\mathcal{X}_1, \ldots, \mathcal{X}_t$ respectively. Define*

$$X \triangleq (X_1, \ldots, X_t).$$

Suppose X is partitioned by a classical random variable D taking values in some set $\mathcal{D}$. Let

$$(\mathbf{X}, \mathbf{D}) \triangleq (X, D)^n.$$

Suppose $\mathbf{X}$ collapses F to the function

$$h : \mathcal{X}_1 \times \cdots \times \mathcal{X}_t \to \mathcal{Z}.$$

Then,

$$\mathsf{IL}_\delta^{t,k}(h \mid (X, D)) \leq \frac{2k}{n} \cdot Q_\delta^{t,k}(F).$$

Proof: Suppose Π is a t-party k-round δ-error quantum protocol with prior entanglement for F with communication cost

$$c \stackrel{\Delta}{=} Q_\delta^{t,k}(F).$$

Our goal is to show that there is a t-party k-round δ-error quantum protocol with prior entanglement for h having information loss at most $\frac{2kc}{n}$ under (X, D). While analysing Π, we will need to maintain that the global state of $\mathcal{P}_1, \ldots, \mathcal{P}_t$ is pure at all times. However, we will run Π on random inputs drawn from certain product probability distributions. In such a situation, we will adopt the following convention. We will assume that in addition to the usual input registers IN_i, $\mathcal{P}_i$ has another set of registers $\widetilde{\mathsf{IN}}_i$. When we require that $\mathcal{P}_i$'s inputs be some random variable $\mathbf{X}_i$, we in fact, start with the following state in the registers $\mathsf{IN}_i\widetilde{\mathsf{IN}}_i$:

$$\sum_{\mathbf{x} \in \mathcal{X}_i^n} \sqrt{p_\mathbf{x}} |\mathbf{x}\rangle |\mathbf{x}\rangle,$$

where

$$p_\mathbf{x} \stackrel{\Delta}{=} \Pr[\mathbf{X}_i = \mathbf{x}].$$

Then, we run the protocol Π as before with input registers IN_i. During this execution no quantum gates are applied to registers $\widetilde{\mathsf{IN}}_i$, they are not sent as messages and they are never measured. From now on the classical random variable $\mathbf{X}_i$ denotes the state of the registers IN_i, which stays unchanged throughout the protocol Π because Π is safe. In this revised protocol Π', $\widetilde{\mathsf{IN}}_i$ is included amongst the qubits of $\mathcal{P}_i$. Π' has the same communication cost as Π. Π' is a δ-error protocol for F with communication cost c. Consider the execution of Π' on input $\mathbf{X} \stackrel{\Delta}{=} (\mathbf{X}_1, \ldots, \mathbf{X}_t)$ conditioned on $\mathbf{D} = \mathbf{d}$; note that under this condition $\mathbf{X}_1, \ldots, \mathbf{X}_t$ are independent random variables. Let $c(i)$ denote the total number of qubits sent by the party $\mathcal{P}^i$ in protocol Π' (which is the same as the total number of qubits sent by $\mathcal{P}^i$ in protocol Π). Then we have, for all $1 \leq i \leq k$,

$$\sum_{j=1}^{n} I((\mathbf{X}^i[j] : \hat{P}^i) \mid \mathbf{D} = \mathbf{d}) \leq I((\mathbf{X}^i : \hat{P}^i) \mid \mathbf{D} = \mathbf{d}) \leq 2c(i).$$

The first inequality above follows from Fact 2.4 because by our definition of $(\mathbf{X}, \mathbf{D})$,

$$(\mathbf{X}_A[j] : 1 \leq j \leq n)$$

are independent random variables when conditioned on $\mathbf{D} = \mathbf{d}$; the second inequality follows from Fact 2.7.

Averaging over the possible values of $\mathbf{D}$, we obtain:

$$\forall i, 1 \leq i \leq k, \sum_{j=1}^{n} I((\mathbf{X}^i[j] : \hat{P}^i) \mid \mathbf{D}) \leq 2c(i).$$

Summing these inequalities with weight $k/k(i)$ over all rounds i, we obtain

$$\sum_{j=1}^{n} \sum_{i=1}^{k} \frac{k}{k(i)} \cdot I((\mathbf{X}^i[j] : \hat{P}^i) \mid \mathbf{D}) \leq 2ck,$$

which implies:

$$\exists j, 1 \le j \le n, \sum_{i=1}^{k} \frac{k}{k(i)} \cdot I((\mathbf{X}^i[j] : \hat{P}^i) \mid \mathbf{D}) \le \frac{2ck}{n}. \tag{6.1}$$

Fix a value of j so that the last inequality holds. For $\mathbf{d} \in \mathcal{D}^n$, let

$$I(\mathbf{d}) \triangleq \sum_{i=1}^{k} \frac{k}{k(i)} \cdot I((\mathbf{X}^i[j] : \hat{P}^i) \mid \mathbf{D} = \mathbf{d}) \tag{6.2}$$

Then from (6.1),

$$\mathop{\mathrm{E}}_{\mathbf{D}}[I(\mathbf{D})] \le \frac{2ck}{n}.$$

We will now obtain a protocol for h by 'embedding' its input as the jth input of Π'. Using a straightforward averaging argument we first fix a value $\hat{\mathbf{d}} \in \mathcal{D}^n$ so that

$$\begin{aligned} &\sum_{d \in \mathcal{D}} \Pr[D = d] I(\mathsf{embed}(\hat{\mathbf{d}}, j, d)) \\ =\ & \mathrm{E}_D[I(\mathsf{embed}(\hat{\mathbf{d}}, j, D)) \le \tfrac{2ck}{n}. \end{aligned} \tag{6.3}$$

Consider the following quantum protocol with prior entanglement Π_h for computing $h(u_1, \ldots, u_t)$. On input $u_i \in \mathcal{X}_i$, $\mathcal{P}_i$ prepares her input registers as follows. In the registers $(\mathsf{IN}_i[\ell], \widetilde{\mathsf{IN}}_i[\ell] : \ell \ne j)$ $\mathcal{P}_i$ places the superposition

$$\sum_{\mathbf{x} \in \mathcal{X}_i^{n-1}} \sqrt{p_{\mathbf{x}}} |\mathbf{x}\rangle |\mathbf{x}\rangle,$$

where

$$p_{\mathbf{x}} \triangleq \Pr[(\mathbf{X}_i[\ell] : \ell \ne j) = \mathbf{x} \mid \mathbf{D} = \hat{\mathbf{d}}];$$

register $\mathsf{IN}_i[j]$ is set to $|u_i\rangle$. Then, $P_1, \ldots, P_t$ run the protocol Π'. Note that the registers $\widetilde{\mathsf{IN}}_i[j], 1 \le i \le t$ do not exist in Π_h.

We need to verify that protocol Π_h has two properties. First, that it is a δ-error protocol for h. For this we note that in Π_h, at all times, the state of the registers that were present in the original protocol Π (that is all registers except $\widetilde{\mathsf{IN}}_i$) is identical to their state when Π is run with input $\mathsf{embed}(\mathbf{X}, j, (u_1, \ldots, u_t))$ conditioned on the event $\mathbf{D} = \hat{\mathbf{d}}$. Since $\mathbf{X}$ collapses F to h, we conclude that Π_h computes $h(u_1, \ldots, u_t)$ with probability at least $1 - \delta$.

Second, we need to verify that

$$\mathsf{IL}(\Pi_h \mid (X, D)) \le \frac{2ck}{n}.$$

We expand the left hand side of (6.3) using definition (6.2) of $I(\mathbf{d})$ and show that each term in it is at least the corresponding term in $\mathsf{IL}(\Pi \mid (X, D))$. For example, consider the term

$$I((X^i : \mathcal{P}^i) \mid D = d), \quad 1 \le i \le k$$

in the definition of $\mathsf{IL}(\Pi_h \mid (X, D))$. Note that the state of $(X^i, \mathcal{P}^i)$ in Π_h on input X conditioned on $D = d$ is identical to the state of $(\mathbf{X}^i[j], \mathcal{P}^i)$ in Π' with registers $\widetilde{\mathsf{IN}}_\ell[j]$, ℓ ranging over all parties except $\mathcal{P}^i$ traced out, when Π' is run on input $\mathbf{X}$ conditioned on

$$\mathbf{D} = \mathsf{embed}(\hat{\mathbf{d}}, j, d).$$

It follows from the monotonicity of mutual information that

$$I((X^i : \mathcal{P}^i) \mid D = d) \leq I((\mathbf{X}^i[j] : \mathcal{P}^i) \mid \mathbf{D} = \mathsf{embed}(\hat{\mathbf{d}}, j, d)).$$

We can thus conclude that

$$\mathsf{IL}(\Pi_h \mid (X, D)) \leq \frac{2ck}{n}.$$

■

Let D be a random variable taking values in $\{1, \ldots, t\}$, with $\Pr[D = d] \stackrel{\Delta}{=} k(d)/k$. Let

$$\mathcal{X}_1 = \cdots = \mathcal{X}_t \stackrel{\Delta}{=} \{0, 1\}.$$

Let X_i be a random variable taking values in $\mathcal{X}_i$ and $X \stackrel{\Delta}{=} (X_1, \ldots, X_t)$.
When $D = d$,

$$\Pr[X_d = 0] = \Pr[X_d = 1] = 1/2$$

and

$$\Pr[X_i = 0] = 1, i \neq d.$$

It is clear that D partitions X. Note that X^n collapses DISJ to AND (here DISJ denotes the promise t-party set disjointness problem and AND denotes the AND function on t bits).

We now show a lower bound for the conditional information loss of AND under (X, D) which is the second part of the proof.

Lemma 6.2 *Let (X, D) be as above. Let $0 \leq \epsilon \leq 1/2$. Then,*

$$\mathsf{IL}_\epsilon^{t,k}(\text{AND} \mid (X, D)) \geq \frac{(1 - 2\epsilon)^2}{8k}.$$

Proof: Let $\theta > 0$. Let Π be a t-party k-round ϵ-error quantum protocol with prior entanglement for AND with

$$\eta \stackrel{\Delta}{=} \mathsf{IL}(\Pi \mid (X, D)) \leq \mathsf{IL}_\epsilon^{t,k}(\text{AND} \mid (X, D)) + \theta.$$

Consider the situation in Π just after the ith round of communication. For any $\mathbf{x} \in \{0, 1\}^t$, let $|\phi_{\mathbf{x}}^i\rangle$ be the global state vector of the qubits of $P_1, \ldots, P_t$ at this point in time, when protocol Π is started with input $X = \mathbf{x}$. Define

$$s(i) \stackrel{\Delta}{=} I((X^i : \hat{P}^i) \mid D = \mathcal{P}^i).$$

Then,

$$s(i) = \frac{k}{k(i)} \cdot I((X^i : \hat{P}^i) \mid D).$$

Hence,

$$\eta = \sum_{i=1}^{k} s(i).$$

Let $\mathbf{e}_i \in \{0,1\}^t$ denote the vector which has an 1 in the ith coordinate and 0 everywhere else. Let $\mathbf{0}, \mathbf{1} \in \{0,1\}^t$ denote the all-zeroes and all-ones vectors respectively. To keep our notation concise, for state vectors $|\phi\rangle$ and $|\psi\rangle$ we write $\| |\phi\rangle - |\psi\rangle \|_t$ instead of $\| |\phi\rangle\langle\phi| - |\psi\rangle\langle\psi| \|_t$. By Lemma 5.3, there is a 'correction' unitary transformation V^i acting on the qubits in the possession of $\mathcal{P}^i$ just after round i such that

$$\left\| V^i |\phi^i_{\mathbf{0}}\rangle - |\phi^i_{\mathbf{e}_{\mathcal{P}^i}}\rangle \right\|_t \leq \sqrt{8s(i)}. \tag{6.4}$$

For any $1 \leq j \leq t$, let W^i_j denote the 'correction' unitary transformation of party $\mathcal{P}_j$ in the last round at or before round i when $\mathcal{P}_j$ was active. Then,

$$W^i_{\mathcal{P}^i} = V^i.$$

For any $j \neq j'$ W^i_j and $W^i_{j'}$ act on disjoint sets of qubits. Without loss of generality, $\mathcal{P}^i = \mathcal{P}_1$ and $\mathcal{P}^{i+1} = \mathcal{P}_2$. Define

$$\delta_i \triangleq \left\| W^i_1 W^i_3 \cdots W^i_t |\phi^i_{\mathbf{e}_2}\rangle - |\phi^i_{\mathbf{1}}\rangle \right\|_t .$$

Let U^i denote the unitary transformation of protocol Π that $\mathcal{P}_1$ applies to the qubits in her possession just after round $i-1$ in order to prepare the messages of round i. Let i' denote the last round before i when $\mathcal{P}_2$ was active. Let $U^{i',i}$ denote the product of the unitary transformations applied by the parties in protocol Π after round i' is complete and till the end of round i. Then,

$$|\phi^i_{\mathbf{x}}\rangle = U^i |\phi^{i-1}_{\mathbf{x}}\rangle$$

and

$$|\phi^i_{\mathbf{x}}\rangle = U^{i',i} |\phi^{i'}_{\mathbf{x}}\rangle.$$

For $j \neq 1$, U^i and W^i_j act on disjoint sets of qubits and

$$W^i_j = W^{i-1}_j.$$

Also, W^i_2 and $U^{i',i}$ act on disjoint sets of qubits. Using the unitary invariance and triangle inequality of the trace distance, the fact that unitary transformations on disjoint sets of qubits commute, and (6.4),

$$\begin{aligned} \delta_i &\triangleq \left\| W^i_1 W^i_3 \cdots W^i_t |\phi^i_{\mathbf{e}_2}\rangle - |\phi^i_{\mathbf{1}}\rangle \right\|_t \\ &\leq \left\| \ W^i_1 W^i_3 \cdots W^i_t U^{i',i} |\phi^{i'}_{\mathbf{e}_2}\rangle - W^i_1 W^i_3 \cdots W^i_t U^{i',i} W^i_2 |\phi^{i'}_{\mathbf{0}}\rangle \ \right\|_t + \end{aligned}$$

$$
\begin{aligned}
& \left\| \; W_1^i W_3^i \cdots W_t^i U^{i',i} W_2^i |\phi_{\mathbf{0}}^{i'}\rangle - U^i W_2^i \cdots W_t^i |\phi_{\mathbf{e}_1}^{i-1}\rangle \; \right\|_t + \\
& \left\| U^i W_2^i \cdots W_t^i |\phi_{\mathbf{e}_1}^{i-1}\rangle - U^i |\phi_{\mathbf{1}}^{i-1}\rangle \right\|_t \\
= \; & \left\| |\phi_{\mathbf{e}_2}^{i'}\rangle - W_2^i |\phi_{\mathbf{0}}^{i'}\rangle \right\|_t + \left\| W_1^i U^{i',i} |\phi_{\mathbf{0}}^{i'}\rangle - U^i |\phi_{\mathbf{e}_1}^{i-1}\rangle \right\|_t \\
& + \left\| W_2^i \cdots W_t^i |\phi_{\mathbf{e}_1}^{i-1}\rangle - |\phi_{\mathbf{1}}^{i-1}\rangle \right\|_t \\
= \; & \left\| |\phi_{\mathbf{e}_2}^{i'}\rangle - W_2^i |\phi_{\mathbf{0}}^{i'}\rangle \right\|_t + \left\| W_1^i |\phi_{\mathbf{0}}^{i}\rangle - |\phi_{\mathbf{e}_1}^{i}\rangle \right\|_t + \left\| W_2^{i-1} \cdots W_t^{i-1} |\phi_{\mathbf{e}_1}^{i-1}\rangle - |\phi_{\mathbf{1}}^{i-1}\rangle \right\|_t \\
\leq \; & \sqrt{8s(i')} + \sqrt{8s(i)} + \delta_{i-1}.
\end{aligned}
$$

It is easy to check that $\delta_0 = 0$. Hence,

$$\delta_k \leq \sum_{i=1}^{k} 2\sqrt{8s(i)}.$$

Using concavity of the square root function, we get that

$$\delta_k \leq \sqrt{32\eta k}.$$

Using Fact 2.8, the fact that a local unitary transformation does not affect the density matrix of the remote system and monotonicity of trace distance, we get that a correct k-round ϵ-error protocol for AND must have

$$\delta_k \geq 2 - 4\epsilon.$$

Hence,

$$\eta \geq \frac{(1-2\epsilon)^2}{8k}$$

implying that

$$\mathsf{IL}_{\epsilon}^{t,k}(\text{AND} \mid (X, D)) \geq \frac{(1-2\epsilon)^2}{8k} - \theta$$

for any $\theta > 0$. This completes the proof of the lemma. ■

Remark: In fact, there are bounded error two-party k-round quantum protocols for the AND of two bits with conditional information loss $O(\log k/k)$. Such protocols can be obtained from the protocols of [BCW98, HdW02, AA03] for set disjointness on a universe of size $O(k^2)$ by setting the first coordinate of Alice and Bob to the two input bits and setting the rest of the coordinates to 0. Another such protocol with one qubit messages can be obtained by adapting the 'reflections in a plane' visualisation (see e.g. [NC00]) of Grover's algorithm on a universe of size $O(k^2)$.

The following is now immediate from Lemma 6.1 and Lemma 6.2.

Theorem 6.1 *Any t-party k-round bounded error quantum protocol for the set disjointness problem needs to have communication cost at least $\Omega(n/k^2)$.*

Chapter 7

Conclusions

In this thesis, we studied problems in communication complexity both in the classical and the quantum models of computation. In the classical setting we studied the direct sum problem. In the quantum setting we studied the index function problem, the set disjointness problem and the pointer chasing problem. We have primarily used information theory to derive our lower bounds. In this process we have also developed a new information theoretic tool which we call the substate theorem. In this chapter we discuss our results briefly and also mention some of the questions that arise naturally as an extension of our study.

7.1 The direct sum problem

In this work, we prove lower bounds for the direct sum problem for protocols with more than one round of communication. We prove the following:

Result 7.1 *Let m, k be positive integers, and $\epsilon, \delta > 0$. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ be a function. Then,*

$$R^k_\delta(f^m) \geq m \cdot \left(\frac{\epsilon^2}{2k} \cdot C^k_{[],\delta+2\epsilon}(f) - 2 \right).$$

The proof of this result works via a message compression result which is as follows:

Result 7.2 *Let X and M be random variables (with some joint distribution), where X is uniformly distributed over $\{0,1\}^n$ and their mutual information $I(X : M) \leq a$. Let $[m]$ be the range of M. Let $S^x_y, x, y \in \{0,1\}^n$ be randomised predicates from $[m]$ to $[0,1]$. Then, there exists a random variable M' (correlated with X) such that*

(a) M' takes values in a set of size $n \cdot 2^{O(a/\epsilon)}$;

(b) There exists $A \subseteq \{0,1\}^n$ of size at least $\frac{2}{3} \cdot 2^n$ such that for all $x \in A$ and $y \in \{0,1\}^n$,

$$|\Pr[S^x_y(M') \mid X = x] - \Pr[S^x_y(M) \mid X = x]| \leq \epsilon.$$

In other words, the above result states that if Alice's message contains only a bits of information about her input, she can compress it to $O(a + \log n)$ bits without changing the error probability of the protocol significantly. A similar message compression argument holds for Bob too. This gives us an alternative proof of the main result of Chakrabarti et al. [CSWY01], with better dependence on the parameters.

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.

A quantum analogue? One might ask if a similar compression of messages is possible in the quantum setting. 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

$$(x, y) \in A \times \{0,1\}^n, |\mathrm{Tr}\ M_y^x \rho_x - \mathrm{Tr}\ M_y^x \rho'_x| \leq \epsilon.$$

Unfortunately the answer is in the negative. We prove the following strong negative result about compressibility of quantum information:

Result 7.3 *For sufficiently large constant a, there exist density matrices ρ'_x, $x \in \{0,1\}^n$ such that there is no subspace of dimension less than $2^{n^{1/5}}$ that contains the supports of most of the ρ'_x.*

This strong negative result seems to suggest that new techniques (not based on information cost) may be required to tackle the direct sum problem for quantum communication.

7.1.1 Open problems

Ideally for the direct sum problem, one would like to prove a result which is independent of the number of rounds in the communication protocol. One possible approach for this is to improve the existing sampling argument so that the error does not grow with the rounds. This will make the final result independent of the number of rounds.

Also in our result we are showing a lower bound for the randomised communication complexity of the m-fold function f^m in terms of the distributional complexity of the function f. Can be improved to obtain a lower bound for the randomised communication complexity of f^m in terms of the ranomized communication complexity of f?

In the wake of the negative result about quantum compressibility, one needs to come up with a different approach to prove some direct sum result in this case.

7.2 Substate theorem

We prove a fundamental theorem about relative entropy of quantum states, which roughly states that if the relative entropy, of two quantum states ρ and σ is at most c, then $\rho/2^{O(c)}$ 'sits inside' σ. We have made crucial use of the substate theorem to arrive at our results for the index function problem and the pointer chasing problem. More formally the result is as follows:

Result 7.4 *Consider two finite dimensional Hilbert spaces* $\mathbf{H}$ *and* $\mathcal{K}$*, where* $\dim(\mathcal{K}) \geq \dim(\mathbf{H})$. *Let* $\mathbb{C}^2$ *denote the two dimensional complex Hilbert space. Let* ρ, σ *be density matrices in* $\mathbf{H}$. *Let* $r > 1$ *be any real number. Let* $|\psi\rangle$ *be a purification of* ρ *in* $\mathbf{H} \otimes \mathcal{K}$. *Then there exist pure states* $|\phi\rangle, |\theta\rangle \in \mathbf{H} \otimes \mathcal{K}$ *(depending on r) and* $|\zeta\rangle \in \mathbf{H} \otimes \mathcal{K} \otimes \mathbb{C}^2$ *such that* $|\zeta\rangle$ *is a purification of* σ *and*

$$\left\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \right\|_t \leq 2/\sqrt{r},$$

where

$$|\zeta\rangle \triangleq \sqrt{\frac{r-1}{r2^{rk}}}\,|\phi\rangle|0\rangle + \sqrt{1 - \frac{r-1}{r2^{rk}}}\,|\theta\rangle|1\rangle \quad \text{and} \quad k \triangleq 8S(\rho\|\sigma) + 14.$$

7.2.1 Open problems

One of the important theorems used in proving the substate theorem is the observational divergence lifting theorem. This relates the observational divergence of two states to the observational divergence of the parent states of which these are sub states. A similar lifting result, known as the Local transition theorem, is known for another metric of distance between two states, namely the trace distance. It will be interesting to see if some lifting result can be shown for relative entropy between two states.

In the substate theorem, the dependence on r (the trace distance term) is different in the classical version and the quantum version. It will be interesting to see if the dependence in the quantum version can be made similar to the one in the classical case or whether the dependence is tight. In case the dependence has to be inverse square root in the quantum case then it will suggest an inherent difference between classical information and quantum information.

7.3 The index function problem

We prove the following about the index function problem:

Result 7.5 *(informal statement) If there is a protocol for the index function problem where B leaks only b bits of information about his index i, then A must send* $\Omega(n/2^{O(b)})$ *bits.*

Result 7.6 *(informal statement) If there is a protocol for the index function problem where B leaks only b bits of information about his input i, then A must leak* $\Omega(n/2^{O(b)})$ *bits of information about her input x. (Note that this implies Result 7.5.)*

Corollary (informal statement) For the index function problem, one of the players must leak $\Omega(\log n)$ bits of information about his input.

General result and other problems: The index function problem is just one of several problems where a statement like the above Corollary can be proved using our technique. In fact, it follows easily that if the communication matrix of the function has VC-dimension at least k, then one of the players must leak at least $\Omega(\log k)$ bits of information about his input. In particular, this implies an $\Omega(\log n)$ loss in privacy for the set disjointness and inner product modulo 2 problems.

7.3.1 Open problems

Our technique gives information loss tradeoff for Alice and Bob for some problems where VC-dimension is large. It will be interesting to obtain new techniques to prove information loss for problems where VC-dimension is not large. Also it will be interesting to find other applications of the substate theorem.

7.4 The pointer chasing problem

In the two-party quantum communication complexity model we consider two versions of the pointer chasing problem, namely the full pointer version and the bit version, and give lower bounds on the amount of communication required. We show the following for the full version.

Result 7.7 *For any constant k, the bounded error quantum communication complexity of the pointer jumping problem P_k (full pointer version) is $\Omega(n \log^{(k)} n)$.*

The information theoretic tool that we developed namely the substate theorem plays a very crucial role in this proof. For the bit version of the problem we show the following:

Result 7.8 *In any bounded error quantum protocol for the pointer chasing problem P_k^{bit}, Alice and Bob must exchange $\Omega(\frac{n}{k^2})$ qubits.*

7.4.1 Open problems

For the full pointer version of the pointer chasing problem, we have shown tight lower bound matching the upper bound well. Whereas in the bit version of the problem, there is still a gap between the upper bound $(O(\frac{n}{k}))$ and the lower bound $(\Omega(\frac{n}{k^2}))$. It will be interesting to fill this gap. Also our lower bounds hold for protocols which do not use any prior entanglement Can they be extended to hold for protocols with prior entanglement as well?

7.5 The set disjointness problem

We show the following for the communication complexity of set disjointness-like functions.

Result 7.9 *The t-party k-round bounded error quantum communication complexity of a set disjointness-like function F is $\Omega(s_m(f_F)/k^2)$.*

In fact, Result 7.9 follows from the following result via easy reductions.

Result 1.2' *The t-party k-round bounded error quantum communication complexity of the promise set disjointness problem is $\Omega(n/k^2)$. This lower bound also holds for Nisan's approximate set disjointness problem [Nis02].*

In particular this implies that for two-party quantum protocols with an unbounded number of rounds, we get a lower bound of $\Omega(n^{1/3})$ for the set disjointness problem.

In a related work (not included in this thesis), we get the following lower bound for the $\mathcal{L}_\infty$ promise problem.

Result 7.10 *The two-party k-round quantum communication complexity of the $\mathcal{L}_\infty$ promise problem is $\Omega(n/(k^3 m^{k+1}))$.*

7.5.1 Open problems

For the two-party quantum communication complexity of set disjointness, the upper and lower bound have been made quite close. There is a small gap remaining still and we believe that the lower bound can be improved to match the upper bound. Our proof of the lower bound, essentially has two parts. We believe that the first part can be improved where we relate the communication complexity of set disjointness with the information loss of the AND function. If the lower bound is improved to match the upper bound then it will give another proof of the $\Omega(\sqrt{n})$ lower bound of the general set disjointness problem, independent of the number of rounds.

Another interesting problem that can be considered is the asymmetric version of the set disjointness problem in the context of quantum communication complexity. Let us say that Bob is allowed to communicate only k qubits to Alice. How many qubits then Alice must communicate for them to solve the set disjointness problem? Our result implies that Alice must communicate at least $\Omega(\frac{n}{k^2} - k)$ qubits. We believe that this can be improved quite a bit possibly to something like $\Omega(n - \text{poly}(k))$.

Bibliography

[AA03] S. Aaronson and A. Ambainis. Quantum search of spatial regions. Manuscript at quant-ph/0303041, 2003.

[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.

[AMS99] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

[ANTV99] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 376–383, 1999.

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

[BBC$^+$93] C. Bennett, G. Brassard, C. Crepeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70:1895–1899, 1993.

[BCKO93] R. Bar-Yehuda, B. Chor, E. Kushilevitz, and A. Orlitsky. Privacy, additional information, and communication. *IEEE Transactions on Information Theory*, 39(6):1930–1943, 1993.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 63–68, 1998. Also quant-ph/9702040.

[BdW01] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th IEEE Conference on Computational Complexity*, pages 120–130, 2001.

[BFS86] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of the 30th ACM Symposium on Theory of Computing*, pages 337–347, 1986.

[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*, 2002.

[CKGS98] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan. Private information retrieval. *Journal of the Association for Computing Machinery*, 45(6):965–981, 1998.

[CKS03] A. Chakrabarti, S. Khot, and X. Sun. Near-optimal lower bounds on the multiparty communication complexity of set disjointness. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, 2003.

[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*, 2001.

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

[CvDNT98] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. In *Proceedings of the 1st NASA International Conference on Quantum Computing and Quantum Communications*, Lecture Notes in Computer Science, vol. 1509, pages 61–74. Springer-Verlag, 1998. Also quant-ph/9708019.

[DGS87] P. Duris, Z. Galil, and G. Schnitger. Lower bounds on communication complexity. *Information and Computation*, 73:1–22, 1987.

[DHR78] D. Dacunha-Castelle, H. Heyer, and B. Roynette. *Ecole d'Eté de Probabilités de Saint-Flour VII.* Lecture Notes in Mathematics, vol. 678. Springer-Verlag, 1978.

[DJS98] C. Damm, S. Jukna, and J. Sgall. Some bounds on multiparty communication complexity of pointer jumping. *Computational Complexity*, 7:109–127, 1998.

[FC95] C. Fuchs and C. Caves. Mathematical techniques for quantum communication theory. *Open Systems and Information Dynamics*, 3(3):345–356, 1995. Also quant-ph/9604001.

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

[FKS02] J. Feigenbaum, S. Kannan, and M. Strauss. An approximate l^1-difference algorithm for massive data streams. *SIAM Journal of Computing*, 32:131–151, 2002.

[GGI$^+$02] A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Fast small space algorithms for approximate histogram maintenance. In *Proceedings of the 34th Annual ACM Symposium Theory of Computing*, pages 389–398, 2002.

[GMMO00] S. Guha, N. Mishra, Rajeev Motwani, and L. O'Callaghan. Clustering data streams. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 359–366, 2000.

[HdW02] P. Hoyer and R. de Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *Symposium on Theoretical Aspects of Computer Science*, pages 299–310, 2002. Also quant-ph/0109068.

[Ind00] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computations. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 189–197, 2000.

[JKS03] T.S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 673–682, 2003.

[Joz94] R. Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

[JRS02a] 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.

[JRS02b] R. Jain, J. Radhakrishnan, and P. Sen. The quantum communication complexity of the pointer chasing problem: the bit version. In *Proceedings of the 22nd Foundations of Software Technology and Theoretical Computer Science Conference*, pages 218–229, 2002.

[KKN92] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. In *Structures in Complexity Theory '92*, pages 262–274, 1992.

[KKN95] M. Karchmer, E. Kushilevitz, and N. Nisan. Fractional covers and communication complexity. *SIAM Journal on Discrete Mathematics*, 8(1):76–92, 1995.

[Kla00] H. Klauck. On quantum and probabilistic communication: Las Vegas and one-way protocols. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 644–651, 2000.

[Kla01] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 288–297, 2001. Also at quant-ph/0106160.

[Kla02] H. Klauck. On quantum and approximate privacy. In *Proceedings of the 19th Annual Symposium on Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, vol. 2285, pages 335–346. Springer-Verlag, 2002. Also quant-ph/0110038.

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

[KNTZ01a] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 124–133, 2001.

[KNTZ01b] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 124–133, 2001.

[Kre95] I. Kremer. Quantum communication. Master's thesis, Hebrew University, 1995.

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

[Lin91] J Lin. Divergence measures based on Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991.

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

[MNSW98a] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37–49, 1998.

[MNSW98b] P. B. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57(1):37–49, 1998.

[Nay99] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40rd Annual IEEE Symposium on Foundations of Computer Science*, pages 369–376, 1999.

[NC00] M. Nielsen and I. Chuang. *Quantum computation and quantum information.* Cambridge University Press, 2000.

[Nis02] N. Nisan. The communication complexity of approximate set packing and covering. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science, vol. 2380, pages 868–875, 2002.

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

[OR94] M. Osborne and A. Rubinstein. *A course in game theory.* MIT Press, 1994.

[PRV01a] S. Ponzio, J. Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62(2):323–355, 2001.

[PRV01b] S. Ponzio, J. Radhakrishnan, and S. Venkatesh. The communication complexity of pointer chasing. *Journal of Computer and System Sciences*, 62(2):323–355, 2001.

[PS84] C. Papadimitriou and M. Sipser. Communication complexity. *Journal of Computer and System Sciences*, 28:260–269, 1984.

[Raz92] A.A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[Raz02] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya Math*, 6, 2002. In Russian. To appear. English version at quant-ph/0204025.

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

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

[Sen03] P. Sen. Lower bounds for predecessor searching in the cell probe model. In *Proceedings of the 18th Annual IEEE Conference on Computational Complexity*, 2003.

[SS02] M. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 360–369, 2002.

[SV98] P. Sen and S. Venkatesh. Lower bounds in quantum cell probe model. In *Proceedings of the 33rd Annual ACM Symposium on Theory of Computing*, pages 20–30, 1998.

[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.

[Yao93] A. C-C. Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual IEEE Symposium on Foundations of Computer Science*, pages 352–361, 1993.