The Quantum Communication Complexity of Sampling

Andris Ambainis
University of California, Berkeley
Berkeley, CA 94720-1776
ambainis@cs.berkeley.edu

Leonard J. Schulman
Georgia Institute of Technology
Atlanta, GA 30332-0280
schulman@cc.gatech.edu

Amnon Ta-Shma
International Computer Science Institute
1947 Center St., Berkeley, CA 94704
amnon@icsi.berkeley.edu

Umesh Vazirani
University of California, Berkeley
Berkeley, CA 94720-1776
vazirani@cs.berkeley.edu *

Avi Wigderson
The Hebrew University
Jerusalem 91904, Israel
avi@cs.huji.ac.il †

Abstract

Sampling is an important primitive in probabilistic and quantum algorithms. In the spirit of communication complexity, given a function $f : X \times Y \rightarrow \{0, 1\}$ and a probability distribution $\mathcal{D}$ over $X \times Y$, we define the sampling complexity of $(f, \mathcal{D})$ as the minimum number of bits Alice and Bob must communicate for Alice to pick $x \in X$ and Bob to pick $y \in Y$ as well as a value z s.t. the resulting distribution of (x, y, z) is close to the distribution $(\mathcal{D}, f(\mathcal{D}))$.

In this paper we initiate the study of sampling complexity, in both the classical and quantum model. We give several variants of the definition. We completely characterize some of these tasks, and give upper and lower bounds on others. In particular, this allows us to establish an exponential gap between quantum and classical sampling complexity, for the set disjointness function. This is the first exponential gap for any task where the classical probabilistic algorithm is allowed to err.

*This research was supported by NSF Grant CCR-9800024, and a JSEP grant.

†This research was supported by grant number 69/96 of the Israel Science Foundation, founded by the Israel Academy for Sciences and Humanities.

1. Introduction

A central question in quantum information theory is the amount of information that can be encoded into n-qubits. There are different ways to formulate this question and, surprisingly, they yield completely different answers. The most natural variant of this question is the following: Alice gets a string $x = x_1, \ldots, x_m$ which is drawn according to some distribution X. The rules of the game are that Alice can encodes x into a superposition ψ over n-qubits ($n \leq m$) in an arbitrary way. She then sends ψ to Bob who can measure ψ in an arbitrary way to get a classical random variable Y. The question is what is the maximal amount of mutual information between X and Y ? More than two decades ago Holevo [10] proved that the answer is n. That is, although $2^n - 1$ complex numbers are necessary to specify the state of n quantum bits, only n bits of information can be retrieved from the superposition, and communicating qubits is not more useful than just communicating classical bits.

Yet, there is something in quantum bits that is more powerful than classical ones. The first demonstration of that was in [4] where it was shown that if the two parties share predefined entangled qubits (that are absolutely independent of the message) then Alice can communicate $2n$ classical bits to Bob using only n communication qubits.

Another example was recently supplied by [2] where Alice's task is to encode m classical bits into n qubits

($m > n$) such that Bob can choose to read any *one* of the m encoded bits of his choice (thereby possibly destroying the information about the remaining $m-1$ bits). On the positive side they showed a scheme for beating Holevo's bound by a constant factor, but on the negative side they showed that m can be no larger than $n \log n$.

A rich hunting ground for such examples is the communication complexity model [18, 17]. [5] showed that if Alice and Bob get two subsets of $[1..n]$ and they wish to compute whether their sets intersect or not, then even though any classical probabilistic protocol must exchange a linear number of bits, the task can be carried out with only square root as many quantum bits. The result is based on Grover's quantum search algorithm [9]. This quadratic gap provided the first asymptotic separation in power between classical and quantum communication.

[5] also gave another communication task based on the Deutsch-Jozsa problem [7], where the number of classical bits required to compute the function *with zero error* is exponentially larger than the corresponding number of quantum bits. However, there is a probabilistic protocol with a small error probability where the number of bits exchanged is as small as the number exchanged by the quantum protocol.

In this paper we give the first example of a communication task which can be carried out by transferring exponentially fewer quantum bits than classical bits, even when error is allowed. The task is the following: Alice has a cardinality k subset $S \subseteq \{1, \dots, n\}$, and Bob must pick a uniformly random cardinality k subset $T \subseteq \{1, \dots, n\}$ disjoint from S. We consider the case $k = \Theta(\sqrt{n})$, and give a quantum protocol in which Alice sends $O(\log n \cdot \log 1/\epsilon)$ quantum bits to Bob, enabling him to sample from a distribution which is ϵ close (in total variation distance) to the desired uniform distribution on subsets disjoint from S. We also show that any purely classical protocol for this task must involve the exchange of $\Omega(\sqrt{n})$ bits between Alice and Bob even for constant ϵ.

In the quantum model there are two natural notions of sampling: one is sampling a given classical distribution (which we just call sampling), the other is constructing a given quantum pure state (which we call q–sampling). We give a tight characterization of the complexity of q–sampling which we believe is important by itself. We also give a tight lower bound on the quantum complexity of zero-error sampling. Finally we show how a lower bound on q–sampling translates into a lower bound for quantum communication complexity and how an upper bound on q–sampling translates into an upper bound on sampling using a quantum protocol.

One interesting aspect of our tight bounds is that the communication complexity of zero error sampling (or q–sampling) of a function f is in the order of the logarithm of the rank of the matrix describing f ($M_f[x, y] = f(x, y)$). This is the first example of a communication task for which the famous log-rank conjecture holds, and reduces this conjecture to a polynomial relation between quantum sampling and classical computation.

1.1. The Communication Complexity Model

The two party communication complexity model, as introduced by [18], consists of two players that have private inputs and wish to compute a known function that depends on both inputs. The players follow a predefined protocol, and exchange communication bits until they are ready to make a decision.

In the quantum communication complexity model [17] Alice and Bob hold qubits. When the game starts Alice holds x and Bob holds y, and so the initial superposition is simply $|x, y\rangle$. The players take turns. Suppose it is Alice's turn to play. 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 the newly received qubits. Each player can also (partially) measure his/her qubits. By the end of the protocol the two players have to decide on a value. If during the protocol the two players are in the system ϕ, then ϕ_{Alice} denotes the state of the subsystem of Alice's qubits, and ϕ_{Bob} is the state of the subsystem of Bob's qubits.

The complexity of a classical (quantum) protocol is the number of bits (qubits) exchanged between the two players. We say a (quantum) protocol *computes* $f : X \times Y \mapsto \{0, 1\}$ with $\epsilon \geq 0$ error, if for any input x, y the probability that the two players compute $f(x, y)$ is at least $1 - \epsilon$. We denote by $R_\epsilon(f)$ ($Q_\epsilon(f)$) the complexity of the best (quantum) protocol that computes f with at most ϵ error. The deterministic complexity $D(f)$ is simply $R_0(f)$.

1.2. Sampling complexity

Next we define the sampling complexity. In the previous definitions the two players had to *compute* the right answer for a given input (x, y). A sampling protocol, however, starts with no input to the two players. Instead, by the end of the protocol Alice holds some $x \in X$, Bob holds some $y \in Y$ and they also hold some "answer" $z \in \{0, 1\}$. We say the protocol induces a distribution $\mathcal{P}$ on (x, y, z), where $\mathcal{P}(x, y, z)$ is the probability that x and y are sampled along with the answer z.

Definition 1.1 *(Sampling) Let $f : X \times Y \mapsto \{0,1\}$, and let $\mathcal{D}$ be any distribution on $X \times Y$. We say the protocol samples f according to $\mathcal{D}$ with ϵ error if the distribution the protocol induces on $\{(x, y, z)\}$ is ϵ close, in the total variation distance, to the distribution $(\mathcal{D}, f(\mathcal{D}))$ obtained by first picking (x, y) according to $\mathcal{D}$ and then evaluating $f(x, y)$. We denote by $\overset{\circ}{R}_\epsilon (f, \mathcal{D})$ ($\overset{\circ}{Q}_\epsilon (f, \mathcal{D})$) the number of communication bits (qubits) needed for a randomized (quantum) protocol P to sample f according to $\mathcal{D}$ with ϵ error. When $\mathcal{D}$ is the uniform distribution we sometimes omit it.*

If $\mathcal{D}$ is a distribution over $\{(x, y)|f(x, y) = 0\}$ we say f samples the 0 values of f, and we denote the classical (quantum) complexity by $\overset{\circ}{R}_\epsilon (f_0, \mathcal{D})$ ($\overset{\circ}{Q}_\epsilon (f_0, \mathcal{D})$).

1.3. q–sampling

In the quantum model it makes sense not only to sample the right classical distribution, but also to approximate the right quantum super-position. For example, we can ask how many communication qubits are needed for two players to generate (or approximate) the super-position $\psi = \Sigma_{x,y}(-1)^{\Sigma_i x_i y_i}|x, y\rangle$. We need to specify what is a good approximation of a super-position and a natural choice is the so called "fidelity" measure: ϕ approximates ψ to within ϵ if $|\langle\phi|\psi\rangle| \geq 1 - \epsilon$.

Definition 1.2 *(q–sampling) We say a quantum protocol q–samples a super-position $\psi = \Sigma_{x\in X,y\in Y} a_{x,y}|x, y\rangle$ to within ϵ error if it starts with no inputs to the two players, and by the end of the protocol the two players compute a superposition ϕ where ϕ_{Alice} has support in X, ϕ_{Bob} has support in Y and $|\langle\phi|\psi\rangle| \geq 1 - \epsilon$.*

Notice that w.l.o.g. we can assume that P does not do any measurement during its execution. The reason is that the protocol ends up in a superposition (a "pure" state rather than a mixed state) which is impossible if even a single non-trivial measurement occurred.

We also note that Alice can always change, at no cost, the phase of the resulting super-position and this does not change the quality of the approximation. Hence, we can assume w.l.o.g that $\langle\phi|\psi\rangle$ is a positive real.

Given a super-position $\phi = \Sigma_{x,y} a_{x,y}|x, y\rangle$ we identify ϕ with the matrix $M_\phi = (a_{x,y})$. Clearly, $\langle\phi|\psi\rangle$ is $\langle M_\phi|M_\psi\rangle$ when we treat the matrices M_ϕ, M_ψ as vectors (of length $|X| \cdot |Y|$). Given a matrix $M = M_{i,j}$ the l_2 norm of M is $||M||_2^2 = \Sigma_{i,j}|M_{i,j}|^2 = Trace(MM^\dagger)$. We note that $||M_\phi - M_\psi||_2^2 = \langle M_\phi - M_\psi|M_\phi - M_\psi\rangle$ and therefore whenever $\langle\phi|\psi\rangle$ is real we have: $||M_\phi - M_\psi||_2^2 = \langle M_\phi|M_\phi\rangle + \langle M_\psi|M_\psi\rangle - 2\langle M_\psi|M_\phi\rangle = 2 - 2\langle\phi|\psi\rangle$. Thus, the following is an equivalent definition for q–sampling:

Definition 1.3 *(q–sampling) We say a quantum protocol q–samples a super-position $\psi = \Sigma_{x\in X,y\in Y} a_{x,y}|x, y\rangle$ to within ϵ if it starts with no inputs to the two players, and by the end of the protocol the two players compute a superposition ϕ where ϕ_{Alice} has support in X, ϕ_{Bob} has support in Y and $||M_\phi - M_\psi||_2^2 \leq 2\epsilon$.*

1.4. q–Sampling a function

Definition 1.4 *Let $f : X \times Y \mapsto \{0,1\}$ be any boolean function, and $\mu : X \times Y \mapsto C$ an l_2 distribution (i.e.,$\Sigma_{x,y}|\mu_{x,y}|^2 = 1$). We say a quantum protocol q–samples f according to the distribution μ with ϵ error if it q–samples the super-position $\Sigma_{x,y}\mu_{x,y}(-1)^{f(x,y)}|x, y\rangle$ to within ϵ error. We denote the number of communication qubits needed for this by $\overset{\bullet}{Q}_\epsilon (f, \mu)$.*

We say a function $g : X \times Y \mapsto M$ is a "product" function, if $g(x, y) = g_1(x)g_2(y)$ for some functions g_1 and g_2. The next lemma shows that at least for product distributions sampling is easier than q–sampling:

Lemma 1.1 *Suppose $f : X \times Y \mapsto \{0,1\}$, and μ is an l_2 distribution. Let $\mathcal{D}$ be the classical distribution $\mathcal{D} : X \times Y \mapsto [0,1]$ obtained by $\mathcal{D}(x, y) = |\mu_{x,y}|^2$. Then $\overset{\circ}{Q}_{2\sqrt{\epsilon}} (f, \mathcal{D}) \leq \overset{\bullet}{Q}_\epsilon (f, \mu) + O(1)$.*

We stress that we do not know whether the other direction also holds and it is possible that sampling is much easier than q–sampling.

Clearly, for any f, $\overset{\circ}{R}_\epsilon (f) \leq R_\epsilon(f)$. A trick from [6] also shows that for any f and l_2 product distribution μ, $\overset{\bullet}{Q}_\epsilon (f, \mu) \leq 2Q_\epsilon(f)$. For completeness we give the proof in the appendix.

1.5 A Tight Bound on q–Sampling

The next theorem completely characterizes the complexity of q–sampling. Given an $N \times N$ matrix M, $MM^\dagger$ is a non-negative matrix and hence has a complete set of non-negative eigenvalues $\lambda_1 \geq \ldots \lambda_N \geq 0$. The i'th singular value, $\sigma_i(M)$, is $\sqrt{\lambda_i}$. We use a lower bound technique from [11] (where it is attributed to Yao) and prove:

Theorem 1 *For any pure state ψ and $0 \le \epsilon \le \frac{1}{2}$, let K_ϵ be the first integer K s.t. $\Sigma_{i=1}^{K} \sigma_i^2(M_\psi) \ge 1-\epsilon$. We have $\lceil \log K_{2\epsilon} \rceil \le \overset{\bullet}{Q}_\epsilon (\psi) \le \lceil \log K_\epsilon \rceil$.*

We note that K_ϵ can also be expressed as $MIN_{A:||M_\psi - A||_2^2 \le \epsilon} Rank(A)$. The proof is easy and follows Corollary 2.2 and the proof technique in our upper bound. When we apply Theorem 1 together with Lemma 1.1 we see that:

Theorem 2 *(upper bound for sampling) suppose $f : X \times Y \mapsto \{0,1\}$, $\epsilon \ge 0$, and μ is an l_2 product distribution. Let $\mathcal{D}$ be the classical distribution associated with μ. Define $M_{f,\mu}[x,y] = \mu_{x,y}(-1)^{(x,y)}$. Let K_ϵ be the minimum K s.t. $\Sigma_{i \le K} \sigma_i(M_{f,\mu})^2 \ge 1-\epsilon$. Then:*

$$\overset{\circ}{Q}_\epsilon(f,\mathcal{D}) \le \lceil \log(K_{\epsilon^2/4}) \rceil + O(1)$$

Example 1 *Let $f : \{0,1\}^n \times \{0,1\}^n \mapsto \{0,1\}$ be the inner-product function, $f(x,y) = \langle x|y \rangle$ and μ the uniform l_2 distribution, $\mu(x,y) = \frac{1}{N}$ (where $N = 2^n$). Because of the orthogonality relations of the inner-product function the matrix $M_{f,\mu} M_{f,\mu}^\dagger$ is just $\frac{1}{N} I_{N \times N}$. Therefore $\sigma_i^2(M_{f,\mu}) = \frac{1}{N}$ for any $i = 1, \ldots, N$. In particular $K_\epsilon = (1-\epsilon)N$. Thus we see that*

$$\overset{\bullet}{Q}_\epsilon (f) \quad \approx \quad \lceil \log((1-2\epsilon)N) \rceil = \lceil n - \log \frac{1}{1-2\epsilon} \rceil$$

Because μ is a product distribution we can also get a lower bound on $Q_\epsilon(f)$:

$$Q_\epsilon(f) \quad \ge \quad \frac{1}{2}(n - \log \frac{1}{1-2\epsilon})$$

Good lower bounds for the computational complexity of the inner-product function are already known. In fact [6] proved a stronger version with better constants that works even when the two parties share entangled qubits. Yet our lower bound applies also for the q–sampling case. □

1.6. Sampling the Disjointness Function

The $DISJ_k(x,y)$ function gets as input two subsets $S, T \subseteq \{1,\ldots,n\}$ each of cardinality k, and outputs 1 iff $S \cap T = \emptyset$. We apply the above theorem to the $DISJ_k$ under the uniform distribution. To analyze the spectrum of M_{DISJ_k} we use a result by Lovasz [12]. We get:

Theorem 3 *For $k = \Theta(\sqrt{n})$, $\overset{\circ}{Q}_\epsilon (DISJ_k) = O(\log(n)\log(\frac{1}{\epsilon}))$. The result is true even when Alice has an input S, and Bob wants to sample a set T disjoint from S.*

In contrast we prove that classically sampling $DISJ_k$ is hard.

Theorem 4 *let $k = \sqrt{n}$. There is a constant $\epsilon > 0$ s.t. $\overset{\circ}{R}_\epsilon (DISJ_k) = \Omega(\sqrt{n})$.*

1.7. A Tight Characterization of Zero-Error Sampling

The lower bound of Theorem 1 does not cover the sampling case, and it is still possible that sampling is much easier than q–sampling. For the special case of *zero-error* sampling we supply a lower bound even for the easier task of sampling, using a similar method to that used in Theorem 1.

Theorem 5 *For any function f and any distribution $\mathcal{D}$, $\overset{\circ}{Q}_0 (f,\mathcal{D}) \ge \frac{\log(Rank(M_{f,\mathcal{D}}))}{2} - 1$*

Notice that when $\mu(x,y) = 1/N$ (where $|X| = |Y| = N$) then $\mathcal{D}(x,y) = 1/N^2$ and so $M_{f,\mathcal{D}}$ and $M_{f,\mu}$ differ only by a constant factor and have the same rank. By Theorem 1 $\overset{\circ}{Q}_0 (f) \le \lceil \log Rank(M_f M_f^\dagger) \rceil + O(1) = \lceil \log Rank(M_f) \rceil + O(1)$. Together we get a tight characterization for zero-error sampling.

Corollary 1.2 *For any $f : X \times Y \mapsto \{0,1\}$, $\overset{\circ}{Q}_0 (f) = \Theta(\log Rank(M_f))$.*

1.8. The Log-Rank Conjecture

The "Log Rank Conjecture" in communication complexity asks whether always $D(f) \le poly(\log(Rank(M_f)))$. Raz and Spieker [16] were the first to show a super-linear gap, and the biggest gap known as of today, due to Nisan and Wigderson [14], exhibits an f with $D(f) \ge \log(Rank(M_f))^{1.6\ldots}$ (see [15]). It is quite possible, for example, that $D(f) \le \log(Rank(M_f))^2$.

Corollary 1.2 sheds some light on the nature of the Log Rank Conjecture. We can replace the expression $\log(Rank(M_f))$ with $\overset{\circ}{Q}_0 (f)$. We now see that the conjecture has two conceptual steps: one is a transition from $\overset{\circ}{Q}_0 (f)$ to $\overset{\circ}{R}_0 (f)$, the other is a transition from $\overset{\circ}{R}_0 (f)$ to $D(f)$. Up to now we have no example of a transition from a sampling algorithm to a computational one. Thus, it makes sense to present the "Weak Log Rank Conjecture":

Conjecture 1.1 *(the Weak Log Rank Conjecture) $\overset{\circ}{R}_0 (f) \le poly(\overset{\circ}{Q}_0 (f))$*

We do not know whether to believe the above conjecture or not. However, it is implied by the Log Rank Conjecture, so it might be a good place to start.

1.9. Sampling vs. Computing

We conclude with some remarks about sampling vs. computing. We notice two things in the sampling model that are different from the computational model. First, it seems that in the sampling case, rounds do not help (as can be seen for the q–sampling in Theorem 1, zero error sampling in Corollary 1.2 and the classical case in Lemma 6.1). This is in sharp contrast to the computational scenario where an extra round can supply an exponential speedup (for details see [15]). We also note that with public random coins, sampling (both quantum and classical) is easy. Thus, public coins are much more powerful than private coins in the sampling model. On the other hand, public coins are almost equivalent to private ones in the computational model ([13, 15]).

As a result any lower bound technique that proves a lower bound on sampling breaks down when the two parties are allowed to use entanglement, and any lower bound technique that is powerful enough to prove a lower bound even in the presence of entanglement (as the lower bound of [6] for the Inner Product function) does not extend to the sampling case.

2. Preliminaries

A classical distribution over X is a function $\mathcal{D} : X \mapsto [0,1]$ s.t. $\Sigma_{x \in X}\mathcal{D}(x) = 1$. Given two distributions $\mathcal{D}_1, \mathcal{D}_2$ over X, the variational distance between them is $|\mathcal{D}_1 - \mathcal{D}_2|_1 \stackrel{\text{def}}{=} \Sigma_x |\mathcal{D}_1(x) - \mathcal{D}_2(x)|$.

The following lemma asserts that if two superpositions are close to each other (in the sense that $|\langle \phi_1 | \phi_2 \rangle| \geq 1 - \epsilon$) then no measurement can effectively distinguish between them. For a superposition ϕ and a complete measurement $\mathcal{O}$ over it, let us denote by $\phi^{\mathcal{O}}$ the classical distribution (over all possible results) obtained by applying the measurement $\mathcal{O}$ over ϕ. A special case of the work done by [1] gives:

Lemma 2.1 *[1] Suppose ϕ_1, ϕ_2 are two superpositions s.t. $|\langle \phi_1 | \phi_2 \rangle| \geq 1 - \epsilon$, then for any complete measurement $\mathcal{O}$ $|\phi_1^{\mathcal{O}} - \phi_1^{\mathcal{O}}|_1 \leq 2\sqrt{\epsilon}$.*

2.1 The Hoffman Wielandt Theorem

Definition 2.1 *Given a matrix $A = (a_{i,j})$, $|A|_\infty \stackrel{\text{def}}{=} \max_{i,j} |a_{i,j}|$ and $||A||_2 \stackrel{\text{def}}{=} (\Sigma_{i,j} |a_{i,j}^2|)^{1/2}$. Note that $||A||_2^2 = Trace(AA^\dagger)$.*

Theorem 6 *([8],Theorem 8.6.4) Let A and B be $N \times N$ matrices. Then, $\Sigma_{i=1}^N [\sigma_i(A) - \sigma_i(B)]^2 \leq ||B - A||_2^2$.*

A direct corollary is:

Corollary 2.2 *Let A and B be two $N \times N$ matrices. Suppose that $Rank(B) = r$. Then $\Sigma_{i=r+1}^n \sigma_i^2(A) \leq ||B - A||_2^2$.*

2.2. The Singular Value Decomposition

Any normal matrix N can be diagonalized by an appropriate unitary basis change, that is there is some unitary transformation U s.t. $UNU^\dagger$ is diagonal with the eigenvalues $\lambda_1, \ldots, \lambda_N$ on the diagonal. A useful generalization of this is the singular value decomposition:

Theorem 7 *(the singular value decomposition) ([8], Section 2.5.6) For any matrix M there are unitary transformations U_1, U_2 s.t. $U_1 M U_2$ is diagonal with $\sigma_1(M), \ldots, \sigma_N(M)$ on the diagonal.*

3. A Tight Bound on q–Sampling

3.1. An Upper Bound

Suppose Alice and Bob are in a super-position $\phi = \Sigma_{x,y} M_{x,y} |x, y\rangle$ represented by the matrix $M = M_\phi$ (i.e., $M[x,y] = M_{x,y}$). Let us check how does the matrix representation changes as Alice applies a transformation T on her qubits. The resulting super-position is

$$\begin{aligned}(T \otimes I)\phi &= \Sigma_{x,y} M_{x,y} |Tx, y\rangle \\ &= \Sigma_{x,y} M_{x,y} \Sigma_z T_{z,x} |z, y\rangle \\ &= \Sigma_{z,y} (\Sigma_x T_{z,x} M_{x,y}) |z, y\rangle \\ &= \Sigma_{z,y} (TM)_{z,y} |z, y\rangle\end{aligned}$$

and so is the resulting super-position is represented by TM. Similarly if *Bob* applies a transformation T on M the resulting super-position is represented by MT^t.

We now give a general algorithm for one message sampling. Suppose the parties want to sample a superposition ψ represented by $M = M_\psi$. By the singular decomposition theorem (Theorem 7) there are unitary transformations U_1, U_2 s.t. $U_1 M U_2$ is the diagonal matrix D with $\sigma_1(M), \ldots, \sigma_N(M)$ on the diagonal. Let $\Lambda = \{w_i | i = 1, ..., K\}$ the set of the first $K = K_\epsilon$ ("heavy") eigenvectors. Let Π be the projection operator onto Λ, i.e., $\Pi[x,y]$ is 1 if $x = y$ and $1 \leq x \leq K$ and zero otherwise. The protocol is the following:

- Alice prepares the super-position $D\Pi$ (which is simply the super-position $c\Sigma_{i=1}^{K}\sigma_i(M)|i,i\rangle$ where $c = 1/\Sigma_{i=1}^{K}\sigma_i^2(M)$. Notice that $1 \le c \le \frac{1}{1-\epsilon}$.) and sends the Y qubits to Bob.
- Alice applies the transformation U_1 on her qubits and Bob applies the transformation U_2^t on his qubits.

Say the resulting super-position is ϕ and its matrix is M_ϕ. We know that $M_\phi = cU_1 D\Pi U_2$. We have:

$$\begin{aligned}
||M_\phi - M_\psi||_2^2 &= ||cU_1 D\Pi U_2 - U_1 D U_2||_2^2 \\
&= ||U_1(cD\Pi - D)U_2||_2^2 \\
&= ||cD\Pi - D||_2^2 \\
&= \Sigma_{i>K}\sigma_i^2(M) + (c-1)^2\Sigma_{i=1}^{K}\sigma_i^2(M) \\
&\le \epsilon + \frac{(c-1)^2}{c} \le \epsilon + 2\epsilon^2 \le 2\epsilon
\end{aligned}$$

The third equality is due to the fact that for every matrix X and unitary matrix U, $||UX||_2 = \langle UX|UX\rangle = \langle X|X\rangle = ||X||_2^2$. The last inequality is since $c - 1 \le c\epsilon$ and for $\epsilon > \frac{1}{2}$ $c \le 2$.

And so we see that ϕ ϵ–approximates ψ (we remind the reader that by using an appropriate phase change, which is for free, the two parties can compute a super-position that ϕ' with $\langle\phi'|\psi\rangle \ge 1-\epsilon$) and the protocol uses only $\lceil \log K_\epsilon \rceil$ qubits, hence the upper bound in Theorem 1 follows.

3.2. A Lower Bound

The lower bound idea is an extension of an idea from Kremer's thesis [11] where it is attributed to Yao. We first show that the outcome of any quantum protocol that uses only l communication qubits can be described as a linear combination of up to 2^l product superpositions (we give a precise statement soon). We use this to show that a quantum sampling protocol is actually a low rank approximation of M_ψ. We then use the Hoffman Wielandt inequality to derive a lower bound on l.

Claim 3.1 *[11] Suppose P is a quantum protocol that uses l communication qubits, starts with no input and computes the superposition ϕ. Further assume that the last qubit communicated is w_l. Then $\phi = \Sigma_{w\in\{0,1\}^l}\ |U(w), w_l, V(w)\rangle$, where U and V depend only on w.*

Proof: [of Claim 3.1]: The proof is by induction on l. The case $l = 0$ is immediate. Suppose it is true for l, let us prove for $l+1$. Assume after l steps the two parties are in the superposition $\Sigma_{w\in\{0,1\}^l}\ |U(w), w_l, V(w)\rangle$ and w.l.o.g. it is now Alice's turn to play. Alice first does some unitary transformation on her qubits, which results in $\Sigma_{w\in\{0,1\}^l}\ |U'(w_1,\ldots,w_l), V(w_1,\ldots,w_l)\rangle$. Then she sends the qubit z to Bob. For every $w_1,\ldots,w_l$ we can represent $|U'(w_1,\ldots,w_l)\rangle$ as a superposition of the possible values of z which completes the induction. $\square$

Now suppose P q–samples ψ (represented by M_ψ) with ϵ error and l communication qubits. Let us denote by $\phi = \Sigma_{x,y}a_{x,y}|x,y\rangle$ the final superposition that the two parties compute (which is, again, represented by M_ϕ). Since ϕ ϵ approximates ψ we know that $||M_\phi - M_\psi||_2^2 \le 2\epsilon$.

By Lemma 7.1 we know that we can represent ϕ as $\phi = \Sigma_{w\in\{0,1\}^l}|U(w), V(w)\rangle$. Because ϕ_{Alice} has support in X, and ϕ_{Bob} in Y, this is actually $\phi = \Sigma_{w\in\{0,1\}^l}\Sigma_{x,y}U_x(w)\cdot V_y(w)|x,y\rangle$ where $U_x(w)$ and $V_y(w)$ are complex numbers. Thus $M_\phi[x,y] = \Sigma_{w\in\{0,1\}^l}U_x(w)V_y(w)$. Let us define a $|X| \times 2^l$ matrix A by $A[x,w] = U_x(w)$, and a $2^l \times |Y|$ matrix $B[w,y] = V_y(w)$. We see that $M_\phi = A\cdot B$ (the $\cdot$ operation is matrix multiplication). In particular $Rank(M_\phi) \le Rank(A) \le 2^l$.

Therefore, M_ϕ is a low-degree matrix approximating (in the l_2 norm) a matrix M_ψ. By Corollary 2.2 we have that $\Sigma_{i=1}^{2^l}\sigma_i^2(M_\psi) \le ||M_\psi - M_\phi||_2^2 \le 2\epsilon$. Therefore, $K_{2\epsilon} \le 2^l$, which proves the lower bound in Theorem 1.

4. An Upper Bound for Sampling

We now prove Lemma 1.1 which shows how to sample a function f under a symmetric distribution once you know how to q–sample it.

Proof: [of Lemma 1.1] Suppose the approximation protocol computes ϕ s.t. $|\langle\phi|\psi\rangle| \ge 1-\epsilon$ where ψ is the ideal super-position $\psi = \Sigma_{x,y}\mu_{x,y}(-1)^{f(x,y)}|x,y\rangle$. We give a sampling protocol:

1. Alice computes the superposition $|00\rangle + |11\rangle$ in qubits z_1, z_2. She sends the second qubit z_2 to Bob.

2. If they both have a $|0\rangle$ (i.e. $z_1 = z_2 = 0$) they compute in the qubits X, Y the superposition $\Sigma_{x,y}\mu_{x,y}|x,y\rangle$ (this can be done at no cost as μ is a product distribution) and if they have a 1 they compute ϕ (using $\lceil\log(K_\epsilon)\rceil$ qubits).

3. Now Bob returns the qubit z_2 to Alice. Alice does a unitary transformation over z_1, z_2 that sends $|00\rangle$ to $\frac{1}{\sqrt{2}}(|00\rangle + |01\rangle)$ and $|11\rangle$ to $\frac{1}{\sqrt{2}}(|00\rangle - |01\rangle)$.

4. Finally, both players measure all their qubits.

Now, suppose for the moment that the protocol was run with $\phi = \psi$. In that case after step (2) the two players are in the superposition

$$\Sigma_{x,y}\mu_{x,y}[|00,x,y\rangle + (-1)^{f(x,y)}|11,x,y\rangle]$$

It can then be easily verified that after step (3) the resulting superposition is:

$$\Sigma_{x,y}\mu_{x,y}|0,f(x,y),x,y\rangle$$

and thus when Alice and Bob measure their qubits they actually sample f according to $\mathcal{D}$ with no error.

Now, in the actual protocol the two players compute ϕ which is not quite ψ but close to it, namely, $|\langle\phi|\psi\rangle| \geq 1-\epsilon$. By Lemma 2.1 we know that the resulting distribution is $2\sqrt{\epsilon}$ close (in the l_1 norm) to the right one, and the theorem follows. $\square$

5. The $DISJ_k$ function

We now wish to study the spectrum of $M = M_{f,\mu}$ when $f = DISJ_k$ and μ is the l_2 uniform distribution ($\mu_{x,y} = 1/N$). In particular we are interested in the size of K_ϵ for $\epsilon > 0$. We note that $M[x,y]$ depends only on the intersection size of x and y. It is not too hard to see that all matrices that are indexed by k-subsets and depend only on the intersection size commute. In particular they share the same eigenspaces. Lovasz [12] analyzed the spectrum of these matrices. Here we give a slightly different description of the eigenspaces of M:

Lemma 5.1 *([12], a different presentation) M has $k+1$ eigenspaces $E_0,\ldots,E_k$. E_0 is of dimension 1 and contains the all $1'$s vector. E_i has dimension $\binom{n}{i}-\binom{n}{i-1}$. The typical eigenvector in E_i is indexed by $x_1,x_2,\ldots,x_{2i-1},x_{2i} \in \{1,\ldots,n\}$. The corresponding eigenvector e (unnormalized) is given by: $e_S = 0$ if there is an index $j : |S \cap \{x_{2j-1},x_{2j}\}| \neq 1$, otherwise $e_S = \Pi_j(-1)^{|S\cap x_{2j}\}|}$. The corresponding eigenvalues are $\lambda_0 = \frac{2\binom{n-k}{k}-\binom{n}{k}}{\binom{n}{k}}$, and $\lambda_i = \frac{2\binom{n-k-i}{k-i}}{\binom{n}{k}}$ for $i > 0$.*

The eigenvalues in the spectrum of M decay rapidly. If we denote $q_i = \Sigma_{w_i\in E_i}|\lambda_i|^2$ (so $\Sigma_{i=0}^k q_i = 1$) then:

Claim: For $k = \Theta(\sqrt{n})$, $\frac{q_{i+1}}{q_i} = O(\frac{1}{i+1})$.

Proof: To calculate q_{i+1}/q_i we first bound λ_{i+1}/λ_i. We get that $\frac{-\lambda_{i+1}}{\lambda_i} = \frac{k-i}{n-k-i} \leq \frac{2k}{n}$. The number of eigenvalues is $\binom{n}{i}-\binom{n}{i-1}$ for E_i and $\binom{n}{i+1}-\binom{n}{i}$ for E_{i+1}, and $\frac{\binom{n}{i+1}-\binom{n}{i}}{\binom{n}{i}-\binom{n}{i-1}} \leq \frac{2n}{i+1}$. Hence

$$\frac{q_{i+1}}{q_i} = \frac{(\binom{n}{i+1}-\binom{n}{i})\lambda_{i+1}^2}{(\binom{n}{i}-\binom{n}{i-1})\lambda_i^2} \leq \frac{2n}{i+1}\cdot\frac{4k^2}{n^2} = \Omega(\frac{1}{i+1}).$$

$\square$

Therefore $q_t \leq \frac{c^t}{t!}$. Now we are set to prove:

Lemma 5.2 $\log K_\epsilon \leq O(\log(n)\frac{\log 1/\epsilon}{\log\log 1/\epsilon})$.

Proof: We set $t = O(\frac{\log 1/\epsilon}{\log\log 1/\epsilon})$ and take $\Lambda = E_0 \cup E_1 \cup \ldots \cup E_t$. We have:

$$\begin{aligned}\Sigma_{i\in\Lambda}|\lambda_i|^2 &= 1-\Sigma_{i\notin\Lambda}|\lambda_i|^2 = 1-\Sigma_{i=t+1}^k q_i \\ &\geq 1-\Sigma_{i=t+1}^k\frac{c^i}{i!} \geq 1-O(\frac{c^t}{t!}) \geq 1-\epsilon\end{aligned}$$

Hence $K_\epsilon \leq |E_0\cup\ldots\cup E_t| \leq t\cdot\binom{n}{t} \leq n^{t+1}$, and $\log K_\epsilon \leq (t+1)\log(n)$ is as required. $\square$

By Theorem 2 this implies that $\overset{\circ}{Q}_\epsilon\ (DISJ_k) \leq O(t\log(n)) = O(\log(n)\log 1/\epsilon)$ which gives the first part of Theorem 3. In the next two subsections we prove the second part of the theorem.

5.1. Sampling 0 values

Let us denote by M_{f_0} the normalized matrix $M_{f_0}[x,y] = \frac{1}{\sqrt{\Delta_0 N}}\begin{cases}1 & f(x,y)=0\\ 0 & f(x,y)=1\end{cases}$, where Δ_0 is the number of values y in a row x s.t. $f(x,y) = 0$ (Δ_0 does not depend on x). since $k = \Theta(\sqrt{N})$, $\Delta_0 \geq \frac{N}{c}$ for some constant c. M_{f_0} is symmetric and has full spectrum $\zeta_1,\ldots,\zeta_N$, $|\zeta_1|^2 \geq \ldots|\zeta_N|^2$. We say K_ϵ^0 is the first K s.t. $\Sigma_{i\leq K}|\zeta_i|^2 \geq 1-\epsilon$. We want to show that $K_\epsilon^0 \leq K_{\epsilon/c}+1$.

One way to show that is to compute the eigenvalues of M_{f_0}. However, there is an easier way. M_f and M_{f_0} share the same eigenspaces (as they commute) and can easily be expressed in terms of each other, therefore we can use the analysis we have for M_f (We remind the reader that M_f is the normalized matrix $M_f[x,y] = \frac{1}{N}(-1^{f(x,y)})$).

Let us denote by $B = \sqrt{N\Delta_0}M_{f_0}$, so B is a $0,1$ matrix. It can be easily verified that $NM_f = B-(J-B) = 2B-J$, where J is the all 1 matrix. Hence, $M_{f_0} = \frac{N}{2\sqrt{N\Delta_0}}M_f + dJ$, for some value d. In particular for any eigenvector $w_i \neq (1,\ldots,1)$, $Jw_i = 0$ and $\zeta_i = \frac{1}{2}\sqrt{\frac{N}{\Delta_0}}\lambda_i$. Thus, $|\zeta_i| = \frac{1}{2}\sqrt{\frac{N}{\Delta_0}}|\lambda_i| \leq \sqrt{c}|\lambda_i|$.

Therefore, suppose $\Sigma_{i\in S}|\lambda_i|^2 \geq 1-\epsilon/c$. Denote $S' = S\cup\{(1,\ldots,1)\}$. Clearly, $\Sigma_{i\notin S'}|\zeta_i|^2 \leq \Sigma_{i\notin S'}c|\lambda_i|^2 \leq \epsilon$. Hence $K_\epsilon^0 \leq K_\epsilon/c+1$.

5.2. Sampling for a given input x

Let g be any function on two k–subsets x and y that depends only on the intersection size of x and

y (e.g., $g(x,y)$ is a if x and y intersect and b otherwise). Let M_g demote the normalized matrix of $(g(x,y))$. We denote by $M_{g,z}$ the normalized matrix $M_g[x,y] = \Delta \begin{cases} g(x,y) & \text{if } x = z \\ 0 & \text{otherwise} \end{cases}$.

The algorithm for approximating M_g (the upper bound in Theorem 1) translates to the following. First, Alice represents M_g in the eigenvector basis W, i.e. she computes $D = WM_gW^\dagger$. Then Alice projects D on the first K_ϵ eigenvectors Λ, i.e. Alice computes $D\Pi$, and she sends the Y qubits to Bob. Finally, they apply W^{-1} and W on their qubits to get the final result $\overline{M_g} = M_gW^\dagger\Pi W$. We showed that $||M_g - \overline{M_g}||_2^2 \leq \epsilon$. We want to show that when we apply the same algorithm over $M_{g,x}$ we get an approximation $\overline{M_{g,x}}$ s.t. $||M_{g,x} - \overline{M_{g,x}}||_2^2 \leq \epsilon$.

We notice that $||M_g - \overline{M_g}||_2^2 = ||\Sigma_x \frac{1}{\sqrt{|X|}}(M_{g,x} - \overline{M_{g,x}}||_2^2 = \frac{1}{|X|}\Sigma_x||M_{g,x} - \overline{M_{g,x}}||_2^2$. The last equality is true because the non-zero entries of the matrices $M_{g,x} - \overline{M_{g,x}})$ are disjoint for different x's.

The basic observation is that if Λ is a union of eigenspaces then $||M_{g,x} - \overline{M_{g,x}}||_2^2$ does not depend on x. Once we know that we get that for any x $||M_g - \overline{M_g}||_2^2 = ||M_{g,x} - \overline{M_{g,x}}||_2^2$ and we are done.

To see the basic observation note that $||M_{g,x} - \overline{M_{g,x}}||_2^2 = 2 - 2 < M_{g,x}|\overline{M_{g,x}} >$. thus, this value depends only on the projection length of $M_{g,x}$ on Λ. However, we show that because of the symmetries of the eigenspaces, the projection of $M_{g,x}$ on each eigenspace, does not depend on x. A full proof will appear in the final version of the paper.

6. A Lower Bound on Classical Sampling

We prove a lower bound for the problem of sampling a pair of disjoint uniformly random subsets of cardinality k with error probability at most ϵ. In general, a sampling protocol for this kind of sampling problem may be described as follows: the game starts with no inputs to Alice and Bob, and after the exchange of a number of messages between the two, Alice picks a random sample from the set X and Bob picks a random sample from the set Y. However, we claim that classical sampling protocols can always be made one message at no cost:

Lemma 6.1 *Given any sampling protocol P with k communication bits and ϵ error, there is an optimal one message sampling protocol that samples from the desired distribution with the same complexity.*

Proof: The protocol goes as follows:

- Alice simulates the protocol P, playing the role of both players. She then announces the resulting sequence of messages M to Bob.
- Alice and Bob pick inputs S and T according to the respective conditional distributions for the protocol P given the messages M.

The crucial observation is that conditioned on the sequence of messages exchanged, the distribution from which Alice and Bob sample is a product distribution. □

In the case of the disjoint subsets sampling problem, X and Y are the collection of all cardinality k subsets of $\{1,\ldots,n\}$. Let $R = U \times V$ be a rectangle, with $U \subseteq X$ and $V \subseteq Y$. To prove the lower bound on the sampling complexity of disjoint cardinality k subsets with error at most ϵ, we shall make use of the following property:

Lemma 6.2 *[3] There exist constant $\epsilon > 0$ and $\delta = 2^{-\Omega(\sqrt{n})}$ such that for any rectangle $R = U \times V$ with $\frac{|R|}{|X||Y|} \geq \delta$, at least ϵ fraction of the pairs of subsets in R intersect (are not disjoint).*

We are now ready to prove the lower bound:

Theorem 8 *There is a constant $\epsilon > 0$ such that any classical protocol to sample from the uniform distribution on disjoint subsets of $\{1,\ldots,n\}$ of cardinality $\Theta(\sqrt{n})$ with error at most ϵ must require the exchange of $\Omega(\sqrt{n})$ bits.*

Proof: Lemma 6.1 implies that the final distribution that Alice and Bob sample from is a linear combination of L *product* distributions D_M, where L is the size of the message space from which Alice chooses her message to Bob (i.e. $logL$ is the number of bits transmitted during the protocol).

Say that a distribution D on rectangle R is *almost uniform* if for any pair of elements $u, v \in R$, the ratio of their probabilities $D(u)/D(v) \leq 4$. We begin by showing that any product distribution can be very closely approximated by a linear combination of a small number of almost uniform distributions on rectangles:

Claim: Let D be a product distribution on $X \times Y$. Then there are rectangles $R_1,\ldots,R_{9n^2}$ such that D is within (total variation distance) 2^{-2n+1} of a linear combination of almost uniform distributions on $R_1,\ldots,R_{9n^2}$.

Proof: We partition X to sets $X_0,\ldots,X_{3n-1}$ and X_{Bad} where $X_i = \{x|\frac{1}{2^{i+1}} \leq D(x) \leq \frac{1}{2^i}\}$ and X_{Bad} is all other strings. We similarly partition Y. We define the distribution $D_{i,j}$ to be the distribution D induces on the rectangle $X_i \times Y_j$ $(0 \leq i,j \leq 3n-1)$. It is clear that $D_{i,j}$ is almost uniform. Let us denote by $\overline{D}$

the appropriate linear combination of the distributions $D_{i,j}$, $\overline{D} = \Sigma_{i,j} p_{i,j} D_{i,j}$ (where $p_{i,j}$ is the weight of the rectangle $X_i \times Y_j$ under D). It is clear that $\overline{D}(a,b) = D(a,b)$ for any (a,b) that belongs to some rectangle $X_i \times X_j$. Thus, $|\overline{D} - D|_1$ is bounded by the total weight (under D) of entries in $X_{Bad} \times Y$ and $X \times Y_{Bad}$, and so is bounded by $2 \cdot 2^n \cdot 2^{-3n} = 2^{-2n+1}$. The lemma follows. □

The final distribution that Alice and Bob sample is a linear combination of L product distributions, where L, the number of distinct messages, is at most 2^n. Therefore by the claim above, the final distribution is 2^{-n+1} close to a linear combination of $9n^2L$ almost uniform distributions on rectangles.

Since $k = \Theta(n)$, the probability p that x and y are disjoint is a constant. By Lemma 6.2 any "*large*" rectangle R with $\frac{|R|}{|X||Y|} \geq \delta$ ($\delta = 2^{-\Omega(\sqrt{n})}$) must have at least ϵ fraction of pairs which are not disjoint sets. Therefore any almost uniform distribution on such a large rectangle must have at least $\epsilon/4$ fraction of its probability on pairs of intersecting sets. This means that there is not too much weight on large rectangles, or, more precisely, if the error probability of the protocol is less than $p\epsilon/8$ than the total weight put on large rectangles is at most $p/2$.

This implies that at least $\frac{p}{2} - \epsilon = \Omega(1)$ fraction of the sets are disjoint and covered by small rectangles. This means that the total area covered by small rectangles is at least $\Omega(|X||Y|)$. But each small rectangle has area at most $\delta|X||Y|$ so the number of rectangles needed is at least $\Omega(\frac{1}{\delta}) = 2^{\Omega(\sqrt{n})}$. Therefore if we want the error to be smaller than $\epsilon' = p\epsilon/8$, we must have $9n^2L \geq 2^{\Omega(\sqrt{n})}$. Therefore the number of bits exchanged, $\log L = \Omega(\sqrt{n})$.

Intuitively, the proof shows that large rectangles introduce large error, while small rectangles provide a very slow progress. □

7. Zero Error Sampling

Finally we show a lower bound on zero error sampling. The lower bound idea is similar to that in Theorem 1 and is an adaptation of ideas from Kremer's thesis [11] where it is attributed to Yao.

Given a protocol P for sampling f we define the $|X| \times |Y|$ matrix M_P^0 by letting $M_P^0[x,y]$ be the probability that P samples (x,y) with the answer 0. We similarly define M_P^1. We let $M_P = M_P^0 - M_P^1$. Note that M_P does not necessarily correspond any more to the probability the protocol answers with a yes or no to an instance (x,y).

Lemma 7.1 *[11] Suppose P uses only l communication qubits. Then $Rank(M_P^0), Rank(M_P^1) \leq 2^{2l}$.*

Proof: Let P be a quantum protocol for sampling f using l qubits. Suppose by the end of the protocol the superposition is ϕ, and w_l, the last qubit communicated, contains the answer (0 or 1). Define $Y_0 = \{w \in \{0,1\}^l | w_l = 0\}$ and $\phi_{x,y}^0 = \Sigma_{w \in Y_0} |x, U_x(w), w_l, y, V_y(w)\rangle$. Then we have:

$$M_P^0[x,y] = \Sigma_{w,z \in Y_0} \langle U_x(w)|U_x(z)\rangle \langle V_y(w)|V_y(z)\rangle$$

If we define a matrix A of dimension $|X| \times |Y_0|^2$ by $A[x,(w,z)] = \langle U_x(w)|U_x(z)\rangle$ and a matrix B of dimension $|Y_0|^2 \times |Y|$ by $B[(w,z),y] = \langle V_y(w)|V_y(z)\rangle$ then we see that $M_P^0[x,y] = (AB)[x,y]$. That is, $M_P^0 = AB$. In particular $Rank(M_P^0) = rank(AB) \leq rank(A) \leq |Y_0|^2 \leq 2^{2l}$. Clearly, a similar argument shows that $Rank(M_P^1) \leq 2^{2l}$. □

We remind the reader that for $f : X \times Y \mapsto \{0,1\}$ the matrix $M_{f,\mathcal{D}}$ has dimensions $|X| \times |Y|$ and is defined by $M_{f,\mathcal{D}}[x,y] = (-1)^{f(x,y)} \mathcal{D}_{x,y}$ ($M_{f,\mathcal{D}}$ is not normalized, i.e., $||M_{f,\mathcal{D}}||_2$ is not necessarily 1). We notice that if P samples f with zero error using l qubits, than $M_P = M_{f,\mathcal{D}}$. Moreover $Rank(M_{f,\mathcal{D}}) = Rank(M_P) \leq Rank(M_P^0) + Rank(M_P^1) \leq 2^{2l} + 2^{2l} = 2^{2l+1}$. In particular $2l + 1 \geq \log(Rank(M_{f,\mathcal{D}}))$. Hence $\overset{\circ}{Q}_0 (f, \mathcal{D}) \geq \frac{\log(Rank(M_{f,\mathcal{D}}))}{2} - 1$ and Theorem 5 follows.

8. Acknowledgments

The authors would like to thank Dorit Aharonov, Ike Chuang, Michael Nielsen, Ran Raz and Steven Rudich for very helpful discussions. We thank Michael Nielsen for showing us how to extend the upper bound of Theorem 1 to the case where M_ψ is not normal.

References

[1] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *STOC*, 1998.

[2] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Are quantum automata always space efficient? manuscript.

[3] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *27th Annual Symposium on Foundations of Computer Science*, pages 337–347, Los Angeles, Ca., USA, Oct. 1986. IEEE Computer Society Press.

[4] C. Bennett and S. Wiesner. *Phys. Rev. Lett.*, 69(2881), 1992.

[5] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *ACM Symposium on Theory of Computing (STOC)*, May 1998.

[6] 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 (Springer-Verlag)*, 1998.

[7] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London Ser. A*, A439:553–558, 1992.

[8] G. H. Golub and C. F. V. Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, USA, third edition, 1996.

[9] L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 212–219, Philadelphia, Pennsylvania, 22–24 May 1996.

[10] A. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. In *Problemy Peredachi Informatsii*, volume 9 (3), pages 3–11, 1973. English translation Problems of Information Transmission, vol 9, 1973, pp. 177-183.

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

[12] L. Lovasz. On the shannon capacity of a graph. *IEEE Transactions on Information theory*, IT-25 (1979).

[13] I. Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, July 1991.

[14] Nisan and Wigderson. On rank vs. communication complexity. *Combinatorica*, 15:557–566, 1995.

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

[16] Raz and Spieker. On the "log rank"-conjecture in communication complexity. *Combinatorica*, 15, 1995.

[17] A. Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual Symposium on Foundations of Computer Science*, pages 352–361, 1993.

[18] A. C. Yao. Some complexity questions related to distributive computing (preliminary report). In *Conference Record of the Eleventh Annual ACM Symposium on Theory of Computing*, pages 209–213, Atlanta, Georgia, 30 Apr.–2 May 1979.

A. $\dot{Q}_\epsilon\ (f,\mu) \le 2Q_\epsilon(f)$

Suppose we can compute f, and we want to q–sample it according to a product distribution μ. Since μ is product we can enter the superposition $\Sigma_{x,y}\mu_{x,y}|x,y\rangle$. Then we can compute f. However, this does not give a q–sampling protocol, because we might use some auxiliary qubits for the computation and thus have garbage entangled with the result. The following proof is basically from [6] and shows how to clean this garbage. The proof is given here for completeness.

Proof: The two players get into the superposition $\Sigma_{x,y}\mu_{x,y}|x,y\rangle$. Since μ is a product distribution this is done at no cost. We run a three step protocol:

Computing f :
We run the protocol for computing f. We use the safe storage principle and each time the protocol wants to measure a qubit we simply copy it to a new qubit that is left untouched. This results in $\phi_1 = \Sigma_{x,y}\mu_{x,y}\ |x,y\rangle(\alpha^0_{x,y}|f(x,y),g^0_{x,y}\rangle + \alpha^1_{x,y}|1-f(x,y),g^1_{x,y}\rangle)$ where $g^0_{x,y}$ (and $g^1_{x,y}$) is the correlated garbage that is produced during the computation, i.e., the right answer $f(x,y)$ is computed with amplitude $\alpha^0_{x,y}$, and is accompanied by $g^0_{x,y}$ in the garbage qubits. The important thing we have to bear in mind is that since the protocol has only ϵ error on average we know that $\Sigma_{x,y}\ |\mu_{x,y}|^2\ |\alpha^0_{x,y}|^2 \ge 1-\epsilon$.

Lifting the result : Next, we lift the result $f(x,y)$ to the amplitude, I.e., the player with the result qubit R changes the amplitude by $(-1)^R$. The resulting superposition is $\phi_2 = \Sigma_{x,y}\mu_{x,y}(-1)^{f(x,y}|x,y\rangle(\alpha^0_{x,y}|f(x,y),g^0_{x,y}\rangle - \alpha^1_{x,y}|1-f(x,y),g^1_{x,y}\rangle)$.

Notice the sign change in the garbage belonging to the wrong answer. We don't like this sign change and we notice that this sign change is immaterial. Namely, if we define $\psi_2 = \Sigma_{x,y}\mu_{x,y}(-1)^{f(x,y}|x,y\rangle(\alpha^0_{x,y}|f(x,y),g^0_{x,y}\rangle + \alpha^1_{x,y}|1-f(x,y),g^1_{x,y}\rangle)$ then $|\langle\phi_2|\psi_2\rangle| \ge \Sigma_{x,y}|\mu_{x,y}|^2\ |\alpha^0_{x,y}|^2 \ge 1-\epsilon$

Reversing the computation : Finally, we would like to get rid of the garbage, so we run the reverse of step 1, the computation step. This can be done by reversing the protocol used in step 1, and thus it at most doubles the number of communication qubits transfered.

We notice that because of the sign change in ϕ_2 the resulting superposition is ugly and depends on the actual computation. However, had the reversing step been applied to ψ_2 we would have received the ideal superposition $\psi = \Sigma_{x,y}\mu_{x,y}(-1)^{f(x,y)}|x,y\rangle$. Now $|\langle\phi_2|\psi_2\rangle| \ge 1-\epsilon$ and step 3 is just a unitary transformation. We conclude that $|\langle\phi_3|\psi\rangle| \ge 1-\epsilon$.

□