THE QUANTUM COMMUNICATION COMPLEXITY OF SAMPLING[†]

ANDRIS AMBAINIS [‡], LEONARD J. SCHULMAN [§], AMNON TA-SHMA [¶], UMESH VAZIRANI [‖], AND AVI WIGDERSON [**]

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 such that 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 variants, 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.

Key words. Communication complexity, Quantum communication complexity, Quantum information theory, Set-disjointness, The Log-Rank conjecture in communication complexity.

AMS subject classifications. 68M10, 68Q10, 68R05.

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 maximal amount of mutual information that can exist between a classical random variable X and a classical probability distribution Y that is obtained from a short quantum encoding of X. More than two decades ago Holevo [9] proved that the mutual information can be at most the number of qubits communicated. 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 a superposition on n quantum bits, 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 by Bennett and Wiesner [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 supplied by Ambainis, Nayak, Ta-Shma and Vazirani [2] and by Nayak [13] 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

[†]An early version of this paper appeared in Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS), 1998, pp. 342-351.

[‡]Institute of Mathematics and Computer Science, University of Latvia, Raina Bulv. 29, Riga, LV-1459, Latvia. E-mail: ambainis@ias.edu.

[§]Caltech, Pasadena CA 91125, USA. E-mail: schulman@caltech.edu

[¶]Tel-Aviv University, Israel 69978. E-mail: amnon@post.tau.ac.il.

[‖]University of California, Berkeley, CA 94720-1776. E-mail: vazirani@cs.berkeley.edu. Supported by NSF Grant CCR-9800024, and a JSEP grant.

[**]The Hebrew University, Jerusalem 91904, Israel, and School of Mathematics, The Institute for Advanced Studies, Princeton, USA. E-mail: avi@cs.huji.ac.il Supported by grant number 69/96 of the Israel Science Foundation, founded by the Israel Academy for Sciences and Humanities.

the positive side they showed a scheme beating Holevo's bound, but on the negative side they showed that n can be no smaller than $\Omega(m)$.

A rich hunting ground for such examples is the communication complexity model [20, 19]. Buhrman, Cleve and Wigderson considered the disjointness function, where Alice and Bob get two subsets x, y of $[1..n]$, and $DISJ(x, y) = 1$ iff x and y are disjoint. It is well known that any classical probabilistic protocol must exchange a linear number of communication bits. On the other hand they show the task can be carried out with only square root as many quantum bits. The result is based on Grover's quantum search algorithm [8]. Recently, Razborov [16] showed an almost matching lower bound on the quantum communication complexity of the problem. This quadratic gap separation provides the first asymptotic separation in power between classical and quantum communication.

Buhrman, Cleve and Wigderson [5] also gave another communication task based on the Deutsch-Jozsa problem [7], where the number of classical bits required to compute a 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. Raz [18] showed such an exponential gap for a *partial* function even in the presence of errors. However, the result applies only for partial functions when the two players are given a promise that their inputs come from a small (in fact, tiny) set of possible inputs.

In this paper we give the first example of a communication task for a *total* function which can be carried out by transferring exponentially fewer quantum bits than classical bits even when error is allowed. We consider the problem $DISJ_k$ that is the disjointness problem on cardinality k subsets $x, y \subseteq [n]$. However, we do not consider the number of communication bits required to *compute* the function, but rather the number of communication bits required to *sample* the function. The task is the following: Alice has a cardinality k subset $S \subseteq \{1, \ldots, n\}$, and Bob must pick a uniformly random cardinality k subset $T \subseteq \{1, \ldots, n\}$ 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.

More generally, given a function $f : X \times Y \to \{0, 1\}$ we consider three communication complexity measures, which we now informally discuss (and formally define in Section 2):

- The usual communication complexity of f, where Alice gets input x, Bob gets input y and we measure the number of communication bits/qubits needed to compute $f(x, y)$. We denote the classical probabilistic communication complexity by $R_\epsilon(f)$, where ϵ is the error probability, and this probability is over the random coins of Alice and Bob. The communication complexity when no error is allowed is denoted by $D(f)$. The quantum communication complexity is denoted by $Q_\epsilon(f)$.
- The communication complexity of generating the superposition of the function. I.e., there is no input to the two parties, and we measure the number of qubits needed to generate the superposition $\sum_{x,y} (-1)^{f(x,y)} |x, y\rangle$, where Alice holds the X register and Bob the Y register. We call this the complexity of *generating* the function, and denote it by $\overset{\bullet}{Q}_\epsilon (f)$.

- The communication complexity of sampling values of f. I.e., Alice and Bob, again, are given no input, and they want to sample $(x, y, z = f(x, y))$, where Alice holds x and Bob holds y. We call this the complexity of sampling the function and denote it by $\overset{\circ}{R}_\epsilon (f)$ in the classical case and $\overset{\circ}{Q}_\epsilon (f)$ in the quantum case.

For formal definitions see Section 2. As expected, sampling is easier than generating, which in turn is easier than solving the problem on a given instance, $\overset{\circ}{Q}_\epsilon (f) \leq \overset{\bullet}{Q}_\epsilon (f) \leq Q_\epsilon(f)$. For the precise statements we prove see Lemmas 5.1 and 5.2.

We show a tight characterization of $\overset{\bullet}{Q}_\epsilon (f)$, the complexity of generating a function. Given f we define the matrix M_f, $M_f[x,y] = (-1)^{f(x,y)}$. We show that $\overset{\bullet}{Q}_\epsilon (f)$ relates to the best low-rank approximation of M_f, namely,

$$\overset{\bullet}{Q}_\epsilon (f) \approx \min_{A:||A-M_f||_2^2 \leq \epsilon} Log(Rank(A))$$

We believe this characterization is important by itself. From that we deduce that

$$\overset{\bullet}{Q}_\epsilon (DISJ_k) = O(\log n \cdot \log 1/\epsilon).$$

We also show, using a combinatorial lemma of Babai, Frankl and Simon [3], that for some constant $\epsilon > 0$,

$$\overset{\circ}{R}_\epsilon (DISJ_k) = \Omega(\sqrt{n})$$

establishing an exponential gap between classical and quantum sampling. Also, as we can efficiently quantum sample (generate) the $DISJ_k$ function, we can also efficiently quantum sample (generate) the $DISJ$ function. Razborov's lower bound [16] then shows an exponential gap between quantum sampling (generating) and normal quantum communication complexity.

We conclude with a remark concerning the Log-Rank conjecture in communication complexity. The conjecture asks whether always $D(f) \leq Poly(Log(Rank(M_f)))$. Raz and Spieker [17] 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 $D(f) \geq Log(Rank(M_f))^{1.6\cdots}$ (see [15], Section 2.5). It is quite possible, for example, that $D(f) \leq Log(Rank(M_f))^2$. The above characterization shows that when $\epsilon = 0$, $\overset{\bullet}{Q}_0 (f) = \Theta(Log(Rank(M_f)))$. In fact, we show that this holds not only for quantum generating f, but also for quantum sampling f, and $\overset{\circ}{Q}_0 (f) = \Theta(Log(Rank(M_f)))$. For the precise statement see Theorem 8.1. This is the first example of a communication task for which the famous log-rank conjecture holds.

Furthermore, we show that zero error classical computing is almost as easy as sampling, or more precisely that $\sqrt{D(f)} \leq \overset{\circ}{R}_0 (f) \leq D(f)$. We thus see that the Log Rank conjecture is equivalent to the conjecture that $\overset{\circ}{R}_0 (f) \leq Poly(\overset{\circ}{Q}_0 (f))$, and can be cast as asking about the relative power of quantum and classical sampling in the no error case.

2. Sampling. The two party communication complexity model, as introduced by Yao [20], 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 [19] 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. ϕ_{Alice} and ϕ_{Bob} are usually mixed states.

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 $R_0(f)$.

2.1. 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 2.1. *A classical distribution over* X *is a function* $\mathcal{D} : X \mapsto [0,1]$ *s.t.* $\sum_{x \in X} \mathcal{D}(x) = 1$. *Given two distributions* $\mathcal{D}_1, \mathcal{D}_2$ *over* X*, the variational distance between them is* $|\mathcal{D}_1 - \mathcal{D}_2|_1 \stackrel{\text{def}}{=} \sum_x |\mathcal{D}_1(x) - \mathcal{D}_2(x)|$.

Definition 2.2. *(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.*

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

Definition 2.3. *(q–generating) We say a quantum protocol q–generates a superposition* $\psi = \sum_{x \in X, y \in Y} a_{x,y} |x, y\rangle$ *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* $X \otimes Ancila_X$*,* ϕ_{Bob} *has support in* $Y \otimes Ancila_Y$ *and* $|\langle \phi | \psi \rangle| \geq 1 - \epsilon$.

Definition 2.4. *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.,*$\sum_{x,y} |\mu_{x,y}|^2 = 1$*). We say a quantum proto-*

col q–generates f according to the distribution μ with ϵ error if it q–generates the superposition $\sum_{x,y} \mu_{x,y}(-1)^{f(x,y)}|x,y\rangle$ to within ϵ error. We denote the number of communication qubits needed for this by $\overset{\bullet}{Q}_\epsilon(f,\mu)$.

3. Preliminaries.

Two superpositions that are close to each other in the fidelity norm (i.e., $|\langle\phi_1|\phi_2\rangle| \geq 1-\epsilon$) can not be effectively distinguished. More precisely, 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. By, e.g., Aharonov, Kitaev and Nisan ([1], Lemma 11):

Fact 3.1. *[1] For any two superpoistions ϕ_1, ϕ_2 and any complete measurement $\mathcal{O}$*

$$|\phi_1^{\mathcal{O}} - \phi_2^{\mathcal{O}}|_1 \leq 2\sqrt{1 - |\langle\phi_1|\phi_2\rangle|^2}.$$

3.1. Some matrix algebra.

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. Singular values and the singular value decomposition theorem, generalize this to arbitrary matrices. Given any possibly non square 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 $\sqrt{\lambda_i}$. The SVD theorem says:

Theorem 3.1. *([10], Lemma 7.3.1) For any matrix M there are unitary transformations U_1, U_2 s.t. U_1MU_2 is diagonal with the singular values $\sigma_1(M), \ldots, \sigma_N(M)$ on the diagonal.*

We use two matrix norms. Given a matrix $A = (a_{i,j})$, the max norm is $|A|_\infty \stackrel{\text{def}}{=} \max_{i,j}|a_{i,j}|$, and the 2-norm is $||A||_2 \stackrel{\text{def}}{=} (\sum_{i,j}|a_{i,j}^2|)^{1/2}$, i.e., $||A||_2^2 = Trace(AA^\dagger)$. The Hoffman Wielandt Theorem states that:

Theorem 3.2. *([10], Corollary 7.3.8) Let A and B be two matrices of same dimensions. Then,*

$$\sum_{i=1}^{N}[\sigma_i(A) - \sigma_i(B)]^2 \leq ||B-A||_2^2.$$

Let B be an arbitrary norm one matrix, $||B||_2^2 = 1$. It follows that $\sum_i \sigma_i^2(B) = Tr(BB^\dagger) = 1$. Let $K_\epsilon(B)$ denote the number of singular values we need to take to collect $1-\epsilon$ weight, i.e., it is the first integer k such that $\sum_{i=1}^k \sigma_i^2(B) \geq 1-\epsilon$.

Claim 3.1. $K_\epsilon(B) = \min_{A:||A-B||_2^2 \leq \epsilon} Rank(A)$.

Proof. Let us denote $K'_\epsilon(B) = \min_{A:||A-B||_2^2 \leq \epsilon} Rank(A)$.

On the one hand, say $K_\epsilon(B) = k$ and $B = U_1DU_2$ where D is a diagonal matrix with the singular values on the diagonal. Let $\bar{D}$ be the matrix containing only the first k singular values, and $A = U_1\bar{D}U_2$. Then A has low rank, and approximates B to within ϵ. It follows that $K'_\epsilon(B) \leq k = K_\epsilon(B)$.

On the other hand, say $K'_\epsilon(B) = k$ and A has rank k and $||A-B||_2^2 \leq \epsilon$. It then follows by the Hoffman Wielandt Theorem that $\sum_{i=1}^N[\sigma_i(A) - \sigma_i(B)]^2 \leq ||B-A||_2^2 \leq \epsilon$. As A has rank k, for at least $N-k$ values i, $\sigma_i(A) = 0$. It then must follows that the squares of the $N-k$ smallest singular values of B must sum up to no more ϵ, i.e., $K_\epsilon(B) \leq K'_\epsilon(B)$. □

4. A Tight Bound on q–Generating.

We completely characterize the complexity of q–generating. With each superposition $\psi = \sum_{x\in X, y\in Y} a_{x,y}|x,y\rangle$ we associate a $|X| \times |Y|$ matrix $M_\psi = (a_{x,y})$. We characterize the complexity of q–generating ψ in terms of the spectrum of M_ψ. We prove:

THEOREM 4.1. *For any pure state ψ and $0 \le \epsilon \le \frac{1}{2}$,*

$$\lceil \log K_{2\epsilon} \rceil \le \dot{Q}_\epsilon\ (\psi) \le \lceil \log K_\epsilon \rceil$$

where $K_\epsilon = MIN_{A:||M_\psi - A||_2^2 \le \epsilon} Rank(A)$. Equivalently, K_ϵ is the first integer K s.t. $\sum_{i=1}^{K} \sigma_i^2(M_\psi) \ge 1-\epsilon$.

4.1. The Upper Bound. Suppose Alice and Bob are in a superposition $\phi = \sum_{x,y} M_{x,y}|x,y\rangle$ represented by the matrix $M = M_\phi$ (i.e., $M[x,y] = M_{x,y}$). Let us check how the matrix representation changes as Alice applies a local unitary transformation T on her qubits. The resulting superposition is

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

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

We now give a general algorithm for one message generating. Suppose the parties want to generate a superposition ψ represented by $M = M_\psi$. By the singular decomposition theorem (Theorem 3.1) there are unitary transformations U_1, U_2 s.t. $U_1^{-1} M U_2^{-1}$ 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 $1 \le x \le K$ and zero otherwise. The protocol is the following:

- Alice prepares the superposition $D\Pi$ (which is simply the superposition $c \cdot \sum_{i=1}^{K} \sigma_i(M)|i,i\rangle$ where $c = \frac{1}{\sqrt{\sum_{i=1}^{K} \sigma_i^2(M)}}$, and notice that $1 \le c \le \frac{1}{\sqrt{1-\epsilon}}$) 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 superposition is ϕ and its matrix is M_ϕ. We know that $M_\phi = cU_1 D\Pi U_2$. We have:

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

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

To finish the proof of the upper bound of Theorem 4.1 we claim:

CLAIM 4.1. $|\langle\phi|\psi\rangle| \ge 1 - \epsilon$

Proof. We treat the matrices M_ϕ, M_ψ as vectors of length $|X| \cdot |Y|$ and notice that $\langle M_\phi|M_\psi\rangle = \langle\phi|\psi\rangle$ by the way the matrices M_ϕ, M_ψ were defined.

Also, since $(U_1^{-1} \otimes U_2^{-1})\psi = \sum_i \sigma_i |i,i\rangle$ and $(U_1^{-1} \otimes U_2^{-1})\phi = c\sum_{i\in\Lambda} \sigma_i |i,i\rangle$, it follows that $\langle\phi|\psi\rangle$ is real. We then see that:

$$\begin{aligned}||M_\phi - M_\psi||_2^2 &= \langle M_\phi - M_\psi | M_\phi - M_\psi\rangle \\ &= \langle M_\phi|M_\phi\rangle + \langle M_\psi|M_\psi\rangle - 2\langle M_\phi|M_\psi\rangle\end{aligned}$$

But $||M_\phi||_2 = ||M_\psi||_2 = 1$ and so

$$||M_\phi - M_\psi||_2^2 = 2(1 - \langle\phi|\psi\rangle)$$

Plugging $||M_\phi - M_\psi||_2^2 \le 2\epsilon$ we get $\langle\phi|\psi\rangle \ge 1 - \epsilon$ as desired. □

4.2. The 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 4.2. *[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 $\phi = \sum_{w\in\{0,1\}^l} |A(w), w_l, B(w)\rangle$, where A and B depend only on w.*

Proof. [of Claim 4.2]: 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 $\sum_{w\in\{0,1\}^l} |A(w), w_l, B(w)\rangle$ 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 $\sum_{w\in\{0,1\}^l} |A'(w_1,\ldots,w_l), B(w_1,\ldots,w_l)\rangle$. Then she sends the qubit z to Bob. For every $w_1,\ldots,w_l$ we can represent $|A'(w_1,\ldots,w_l)\rangle$ as a superposition of the possible values of z which completes the induction. □

Now suppose P q–generates ψ (represented by M_ψ) with ϵ error and l communication qubits. Let us denote by $\phi = \sum_{x,y} a_{x,y}|x,y\rangle$ the final superposition that the two parties compute (which is, again, represented by M_ϕ). By Claim 4.2 we know that we can represent ϕ as $\phi = \sum_{w\in\{0,1\}^l} |A(w), B(w)\rangle$. Because ϕ_{Alice} has support in X, and ϕ_{Bob} in Y, this is actually $\phi = \sum_{w\in\{0,1\}^l} \sum_{x,y} a_x(w)\cdot b_y(w)|x,y\rangle$ where $a_x(w)$ and $b_y(w)$ are complex numbers. Thus

$$M_\phi[x,y] = \sum_{w\in\{0,1\}^l} a_x(w) b_y(w)$$

Let us define a $|X|\times 2^l$ matrix A by $A[x,w] = a_x(w)$, and a $2^l \times |Y|$ matrix $B[w,y] = b_y(w)$. We see that $M_\phi = A\cdot B$, where the $\cdot$ operation is matrix multiplication. In particular

$$Rank(M_\phi) \le Rank(A) \le 2^l$$

Since ϕ ϵ approximates ψ we know that $||M_\phi - M_\psi||_2^2 = \langle M_\phi - M_\psi | M_\phi - M_\psi\rangle = 2(1-\langle\phi|\psi\rangle) \le 2\epsilon$. It follows that $K_{2\epsilon} = \min_{M:||M-M_\psi||_2^2\le 2\epsilon} Rank(M) \le Rank(M_\phi) \le 2^l$ as desired.

5. Relationships Between Sampling and Computing. 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. For product distributions μ we show that sampling is not harder than q–generating, which in turn is not harder than worst-case solving the problem.

5.1. Sampling vs. q–generating.

Lemma 5.1. *Suppose $f : X\times Y \mapsto \{0,1\}$, and μ is an l_2* product *distribution. Let $\mathcal{D} : X\times Y \mapsto [0,1]$ be the classical distribution associated with μ, $\mathcal{D}(x,y) = |\mu_{x,y}|^2$. Then $\overset{\circ}{Q}_{4\sqrt{\epsilon}}(f,\mathcal{D}) \le \overset{\bullet}{Q}_\epsilon(f,\mu) + O(1)$.*

Proof. Suppose the approximation protocol computes ϕ s.t. $|\langle\phi|\psi\rangle| \ge 1-\epsilon$ where ψ is the ideal superposition $\psi = \sum_{x,y}\mu_{x,y}(-1)^{f(x,y)}|x,y\rangle$. 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 $\sum_{x,y}\mu_{x,y}|x,y\rangle$ (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

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

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

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

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 Fact 3.1 we know that the resulting distribution is $2\sqrt{1-|\langle\phi|\psi\rangle|^2} \leq 2\sqrt{1-(1-\epsilon)^2} \leq 2\sqrt{2\epsilon}$ close (in the l_1 norm) to the right one, and the lemma follows. □

5.2. ***q*–generating vs. Computing.** Suppose we can compute f, and we want to q–generate it according to a product distribution μ. Since μ is product we can enter the superposition $\sum_{x,y} \mu_{x,y}|x,y\rangle$. Then we can compute f. However, this does not give a q–generating protocol because we might use some auxiliary qubits for the computation and thus have garbage entangled with the result. The following proof follows ideas from Cleve, van Dam, Nielsen Tapp [6] who showed how to clean such garbage. The proof is given here for completeness.

LEMMA 5.2. *For any function f and any l_2 product distribution μ, $\overset{\bullet}{Q}_{2\epsilon}(f,\mu) \leq 2Q_\epsilon(f)$.*

Proof. Let T be a small error 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. Now, say that $T|x,0,y,0\rangle = |x,y\rangle \otimes [\alpha^0_{x,y}|f(x,y), g^0_{x,y}\rangle + \alpha^1_{x,y}|1-f(x,y), g^1_{x,y}\rangle]$, where $g^0_{x,y}$ (and $g^1_{x,y}$) is the correlated garbage that is produced during the computation and is divided between the two players, 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 two players get into the superposition $\sum_{x,y} \mu_{x,y}|x,y\rangle$. Since μ is a product distribution this is done at no cost. We run a three step protocol:

<u>**Computing f**</u> : This results in

$$\phi_1 = \sum_{x,y} \mu_{x,y}\ |x,y\rangle \otimes [\alpha^0_{x,y}|f(x,y), g^0_{x,y}\rangle + \alpha^1_{x,y}|1-f(x,y), g^1_{x,y}\rangle].$$

As T has only ϵ error on average we know that $\sum_{x,y}\ |\mu_{x,y}|^2\ |\alpha^0_{x,y}|^2 \geq 1-\epsilon$.

<u>**Lifting the result**</u> : 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

$$\phi_2 = \sum_{x,y} \mu_{x,y}(-1)^{f(x,y)}|x,y\rangle \otimes [\alpha^0_{x,y}|f(x,y), g^0_{x,y}\rangle - \alpha^1_{x,y}|1-f(x,y), g^1_{x,y}\rangle].$$

Notice the sign change in the garbage belonging to the wrong answer. We do not like this sign change and we notice that this sign change is immaterial. Namely, if we define

$$\psi_2 = \sum_{x,y} \mu_{x,y}(-1)^{f(x,y)}|x,y\rangle \otimes [\alpha^0_{x,y}|f(x,y), g^0_{x,y}\rangle + \alpha^1_{x,y}|1-f(x,y), g^1_{x,y}\rangle]$$

then $|\langle\phi_2|\psi_2\rangle| \geq \sum_{x,y}|\mu_{x,y}|^2(|\alpha^0_{x,y}|^2 - |\alpha^1_{x,y}|^2)$ which is at least $1-2\epsilon$.

<u>**Reversing the computation**</u> : Finally, we would like to get rid of the garbage, so we reverse T, this at most doubles the number of communication qubits transferred. 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 $\psi =$

$\sum_{x,y} \mu_{x,y}(-1)^{f(x,y)}|x,y\rangle$. Now $|\langle\phi_2|\psi_2\rangle| \geq 1-2\epsilon$ and reversing T is just a unitary transformation. We conclude that $|\langle\phi_3|\psi\rangle| \; \geq \; 1-2\epsilon$.

□

6. The $DISJ_k$ 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 bound the quantum sampling complexity of the function under the l_2 uniform distribution $\mu_{x,y}=1/N$. We prove:

THEOREM 6.1. *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 subset T disjoint from S.*

By Theorem 4.1 we need to analyze the spectrum of $M=M_{DISJ_k,\mu}$. Indeed, notice that $M[x,y]$ depends only on the intersection size of x and y. It is not too difficult 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 and we give a slightly different description of the eigenspaces of M he obtains:

LEMMA 6.2. *([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. Let $q_i=\sum_{w_i\in E_i}|\lambda_i|^2$ so that $\sum_{i=0}^k q_i=1$. Then:

CLAIM 6.1. *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

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

□

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

LEMMA 6.3. $\log K_\epsilon\leq O(\log(n)\frac{\log 1/\epsilon}{\log\log 1/\epsilon})$.

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:

$$\sum_{i\in\Lambda}|\lambda_i|^2=1-\sum_{i\notin\Lambda}|\lambda_i|^2\;=\;1-\sum_{i=t+1}^{k}q_i$$
$$\geq 1-\sum_{i=t+1}^{k}\frac{c^i}{i!}\geq 1-O(\frac{c^t}{t!})\geq 1-\epsilon$$

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

□

By Theorem 4.1 $\overset{\bullet}{Q}_\epsilon(\psi) \le \lceil \log K_\epsilon \rceil \le O(t\log(n)) = O(\log(n)\log 1/\epsilon)$, and a similar upper bound on $\overset{\circ}{Q}_\epsilon(DISJ_k)$ follows from Lemma 5.1. This gives the first part of Theorem 6.1. This, in particular, shows that it is easy for Alice and Bob to sample a uniform pair of subsets x and y, along with the knowledge whether x and y intersect.

In the next two subsections we prove the second part of the theorem. We want to show two things. One is that the result holds even when Alice and Bob want to sample only *disjoint* subsets, and second that the result holds even when Alice is given an input x and Bob is asked to sample a subset y *disjoint* with x.

6.1. Generating Disjoint Subsets. Alice and Bob want to ϵ–approximate a sample of disjoint k–subsets x and y. This amounts to sampling the disjointness function according to the distribution $\mathcal{D}$ that is uniform over all pairs of disjoint subsets (notice that $\mathcal{D}$ is *not* a product distribution). Clearly, it is enough for Alice and Bob to approximate the normalized superposition $\psi = \sum_{x,y:x\cap y=\emptyset} \frac{1}{\sqrt{\Delta_0 N}}|x,y\rangle$, for once they do that they can measure x and y and get the desired sample. The normalizing factor Δ_0 is the number of values y in a row x s.t. $x\cap y=\emptyset$ and does not depend on x.

Denote by M_{f_0} the normalized matrix $M_{f_0}[x,y] = \frac{1}{\sqrt{\Delta_0 N}}\begin{cases} 1 & x\cap y=\emptyset \\ 0 & otherwise\end{cases}$. M_{f_0} is symmetric and has full spectrum $\zeta_1,\dots,\zeta_N$, $|\zeta_1|^2 \ge \dots |\zeta_N|^2$. We say K^0_ϵ is the first K s.t. $\sum_{i\le K}|\zeta_i|^2 \ge 1-\epsilon$. By Theorem 4.1 (which applies to any superposition) Alice and Bob can ϵ approximate ψ using only $O(\log(K^0_\epsilon))$ communication qubits.

Since $k=\Theta(\sqrt{N})$, $\Delta_0 \ge \frac{N}{c}$ for some constant c. It is left to show that $O(\log(K^0_\epsilon)) = O(\log(n)\log(1/\epsilon))$. One way to show this is to compute the eigenvalues of M_{f_0}. However, there is an easier way. We show that $K^0_\epsilon \le K_{\epsilon/c}+1$ and then Lemma 6.3 implies the bound. We are left with:

Claim 6.2. $K^0_\epsilon \le K_{\epsilon/c}+1$.

Proof. (of claim) Denote $M_f[x,y] = \frac{1}{N}(-1^{f(x,y)}))$. M_f and M_{f_0} share the same eigenspaces (as they commute). We now express M_f and M_{f_0} in terms of each other. 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 $w_i \ne (1,\dots,1)$, $Jw_i=0$ and $\zeta_i = \frac{1}{2}\sqrt{\frac{N}{\Delta_0}}\lambda_i$. Thus,

$$|\zeta_i| = \frac{1}{2}\sqrt{\frac{N}{\Delta_0}}|\lambda_i| \le \sqrt{c}|\lambda_i| \quad ,i>1$$

Therefore, suppose $\sum_{i\in S}|\lambda_i|^2 \ge 1-\epsilon/c$. Denote $S' = S\cup\{(1,\dots,1)\}$. Clearly, $\sum_{i\notin S'}|\zeta_i|^2 \le \sum_{i\notin S'} c|\lambda_i|^2 \le \epsilon$. Hence $K^0_\epsilon \le K_{\epsilon/c}+1$. □

6.2. Sampling For a Given Input x. Alice is given an input $z\in X$ and the goal is that Bob samples $y\in Y$ s.t. $z\cap y=\emptyset$. We follow a protocol similar to that in the upper bound of Theorem 4.1. Given an input $z\in X$ and an $\epsilon>0$ define $M=M_{f_0}$ as in the previous subsection. Let W be the eigenvector basis of M (which is symmetric). Let $\Lambda=\Lambda_\epsilon$ be the union of the first eigenspaces E_i (defined in Lemma 6.2) that contain the first K_ϵ heavy eigenvectors of M. Let Π be the projection operator over Λ.

We now describe the protocol. Alice gets into the normalized superposition $v_z = \frac{1}{\sqrt{\Delta_0}} \sum_{y:y\cap z=\emptyset} |y\rangle$. In the eigenvector basis W, $v_z = \sum_i \gamma_i |w_i>$. Alice then projects v_z onto Λ to get $\overline{v_z} = \sum_{i\in\Lambda} \gamma_i |w_i>$ and sends $\overline{v_z}$ to Bob. Bob returns $\overline{v_z}$ to the original basis and measures to get some y. To prove correctness we show:

LEMMA 6.4. $|\langle v_z|\overline{v_z}\rangle| \geq 1-\epsilon$.

Proof. $\langle v_z|\overline{v_z}\rangle = \sum_{i\in\Lambda} |\gamma_i|^2$, i.e., it is the length of the projection of v_z onto Λ. We show that this quantity is the same for all z. If we know that we can define $\psi = \frac{1}{\sqrt{N}} \sum_z |z, v_z\rangle$ and $\overline{\psi} = \frac{1}{\sqrt{N}} \sum_z |z, \overline{v_z}\rangle$ (so ψ and $\overline{\psi}$ are normalized). Then, from the proof of Theorem 4.1 we know that:

$$|\langle\psi|\overline{\psi}\rangle| \geq 1-\epsilon$$

But

$$\langle\psi|\overline{\psi}\rangle = \frac{1}{N}\sum_z \langle v_z|\overline{v_z}\rangle = \langle v_z|\overline{v_z}\rangle$$

which together implies that $\langle v_z|\overline{v_z}\rangle = \langle\psi|\overline{\psi}\rangle \geq 1-\epsilon$ as required. Indeed,

CLAIM 6.3. *For any eigenspace E_j, $|\langle v_z|E_j\rangle|^2$, which is the length of the projection of v_z on E_j, does not depend on z.*

Proof. Let $z_1, z_2 \in X$ be two k-subsets. I.e., $z_1, z_2 \subset [1..n]$ and $|z_1| = |z_2| = k$. There is a permutation $\pi \in S_n$ s.t. $\pi(z_1) = z_2$ where for a set A, $\pi(A) = \{\pi(a)|a \in A\}$.

The operation of the permutation π can be thought of as a unitary transformation permuting the basis vectors $|x\rangle$ for $x \in X$. I.e., given a superposition $\phi = \sum_{i\in X} a_i|i\rangle$, $\pi(\phi)$ is defined to be $\sum_{i\in X} a_i|\pi(i)\rangle$. In particular, for any two superpositions ϕ_1, ϕ_2 $\langle\pi(\phi_1)|\pi(\phi_2)\rangle = \langle\phi_1|\phi_2\rangle$. As a result $\langle v_{z_1}|E_j\rangle = \langle\pi(v_{z_1})|\pi(E_j)\rangle$ where $\pi(E_j) = Span\{\pi(w)|w \in E_j\}$. However, we observe that

$$\begin{aligned}\pi(v_{z_1}) &= \sum_{y:y\cap z_1=\emptyset} |\pi(y)\rangle \\ &= \sum_{w:\pi^{-1}(w)\cap z_1=\emptyset} |w\rangle \\ &= \sum_{w:w\cap\pi(z_1)=\emptyset} |w\rangle \ = \ v_{z_2}\end{aligned}$$

Finally, because of the symmetries of the eigenspaces E_j, $\pi(E_j) = E_j$. The lemma follows. □ □

7. A Lower Bound on Classical Sampling. In contrast we prove that classically sampling $DISJ_k$ is hard. We begin with the observation that classical sampling protocols can always be made to have just one message at no cost. We prove:

LEMMA 7.1. *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 [*].

[*]We assume all messages belonging to the same round have the same length. If this is not the case, Alice has to send a special sign at the end of each message, which may, at most, increase the number of communication bits by a constant factor.

- 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. □

We are now ready to prove:

THEOREM 7.2. *let* $k = \sqrt{n}$. *There is a constant* $\epsilon > 0$ *s.t.* $\overset{\circ}{R}_\epsilon\ (DISJ_k) = \Omega(\sqrt{n})$.

Proof. Let P be the distribution on $X \times Y$ that Alice and Bob sample from (X and Y is the set of all k-sets). By Lemma 7.1 P is a convex combination of L *product* distributions D_M, $P = \sum p_i D_i$, 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). We say a distribution D on rectangle R is *smooth* if for any pair of elements $u, v \in R$, $\frac{D(u)}{D(v)} \leq 4$. We soon show that any product distribution can be very closely approximated by a convex combination of a small number of smooth distributions on rectangles, namely,

CLAIM 7.1. *Let* D *be a product distribution on* $X \times Y$. *Then there are rectangles* $R_1, \ldots, R_{9n^2}$, *and smooth distributions* D_i *on* R_i, *such that* D *is within (total variation distance)* 2^{-2n+1} *of a convex combination of* D_i.

In particular, P is 2^{-n+1} close to a convex combination $\sum_{i=1}^{9n^2 L} p_i P_i$, where P_i is some smooth distribution over some rectangle R_i. Intuitively, the proof shows that large rectangles R_i introduce large error, while small rectangles provide very slow progress. For that we use the following combinatorial lemma of Babai, Frankl and Simon:

LEMMA 7.3. *[3] There exist a constant* $\epsilon_0 > 0$ *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_0 *fraction of the pairs of subsets in* R *intersect (are not disjoint).*

Let us call a rectangle R_i large if $\frac{|R_i|}{|X||Y|} \geq \delta$. By the lemma, any large rectangle must contain at least ϵ_0 fraction of intersecting pairs. Thus, any smooth distribution P_i on a large rectangle R_i must have at least $\frac{\epsilon_0}{4}$ weight on intersecting pairs. Let h denote the total weight of heavy rectangles in the convex combination (i.e., $h = \sum_{i: R_i \text{ is heavy}} p_i$). We see that intersecting pairs get at least weight $\frac{h\epsilon_0}{4}$. We conclude that $\frac{h\epsilon_0}{4} \leq \epsilon$ and $h \leq \frac{4\epsilon}{\epsilon_0} = O(\epsilon)$.

We now concentrate on the non-heavy rectangles P_i. We say we *touch* a pair (x, y) if some non-heavy rectangle R_i contains it. Let I be the set of all disjoint pairs. We see that we must touch at least $(1 - (\epsilon + h))|I|$ pairs in I, or else there are $(\epsilon + h)|I|$ elements that get weight $\epsilon + h$ in the uniform distribution over disjoint pairs, and only weight h in P. As every non-heavy rectangle R_i can touch at most $|R_i| \leq \delta|X| \cdot |Y|$ elements, we must have that $9n^2 L\delta|X| \cdot |Y| \geq (1 - \epsilon - h)|I| \geq (1 - O(\epsilon))|I|$.

For $k = \sqrt{n}$ the number of disjoint pairs is some $c_0|X| \cdot |Y|$, for some constant c_0. Thus, $L \geq \frac{(1-O(\epsilon))c_0}{9n^2} \cdot 2^{\Omega(\sqrt{n})}$. It follows that for some small enough constant $\epsilon > 0$ we must have $L \geq 2^{\Omega(\sqrt{n})}$ as desired.

We are left with the proof of Claim 7.1 which we give now:

Proof. (Of Claim 7.1) 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 $D_{i,j}$, $\overline{D} = \sum_{i,j} p_{i,j} D_{i,j}$ (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. □
□

8. Zero Error Sampling and the Log-Rank conjecture. Theorem 4.1 characterizes the q–generating complexity $\overset{\bullet}{Q}$. However, it is still possible that sampling is much easier (even in the quantum world) than q–generating. 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 4.1.

THEOREM 8.1. *For every function f and any distribution $\mathcal{D}$, $\overset{\circ}{Q}_0\ (f, \mathcal{D}) \geq \frac{\log(Rank(M_{f,\mathcal{D}}))}{2} - 1$*

Proof. 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 8.2. *[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). By Claim 4.2

$$\phi = \sum_{w \in \{0,1\}^l} \sum_{x \in X, y \in Y} |x, U_x(w), w_l, y, V_y(w)\rangle$$

Define $Y_0 = \{w \in \{0,1\}^l | w_l = 0\}$ and $\phi^0_{x,y} = \sum_{w \in Y_0} |x, U_x(w), w_l, y, V_y(w)\rangle$. Then

$$\begin{aligned} M_P^0[x,y] &= \langle \phi^0_{x,y} | \phi^0_{x,y} \rangle \\ &= \sum_{w,z \in Y_0} \langle U_x(w) | U_x(z) \rangle \langle V_y(w) | V_y(z) \rangle \end{aligned}$$

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 $Rank(M_P^0) = rank(AB) \leq rank(A) \leq |Y_0|^2 \leq 2^{2l}$. 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 8.1 follows. □

We see in particular that for the uniform distribution ($\mathcal{D}(x,y) = 1/N^2$ and $\mu(x,y) = 1/N$) $M_{f,\mathcal{D}}$ and $M_{f,\mu}$ differ only by a constant factor and so have the same rank. By Theorem 4.1 $\overset{\circ}{Q}_0\ (f) \leq \lceil \log Rank(M_f M_f^\dagger) \rceil + O(1) = \lceil \log Rank(M_f) \rceil + O(1)$. Theorem 8.1 gives a matching lower bound. Together we get the following tight characterization for zero-error sampling.

COROLLARY 8.3. *For any* $f : X \times Y \mapsto \{0,1\}$, $\overset{\circ}{Q}_0(f) = \Theta(\log Rank(M_f))$.

8.1. Zero Error Classical Computing is almost as easy as Sampling.

THEOREM 8.4. $\sqrt{D(f)} \leq \overset{\circ}{R}_0(f) \leq D(f)$.

Proof. Given the matrix M_f, a monochromatic cover is a set of monochromatic rectangles in M_f that together cover the whole matrix. Denote $C(f)$ the smallest number of monochromatic rectangles needed to cover M_f. Denote $C^D(f)$ the smallest number of disjoint monochromatic rectangles needed to cover M_f. It is well known (see [15], chapter 2) that

$$\sqrt{D(f)} \leq N(f) = log_2 C(f) \leq \log_2 C^D(f) \leq D(f)$$

where $N(f)$ is the non-deterministic communication complexity. We show that

$$\log_2 C(f) \leq \overset{\circ}{R}_0(f) \leq \log_2 C^D(f)$$

and in particular we get that $\sqrt{D(f)} \leq N(f) \leq \overset{\circ}{R}_0(f) \leq D(f)$ as required.

We first show that $\log_2 C(f) \leq \overset{\circ}{R}_0(f)$. Indeed, by Lemma 7.1 there is a one message zero error sampling protocol whose complexity is $k = \overset{\circ}{R}_0(f)$. In the one message protocol a message M is chosen (out of the 2^k possible messages) according to some probability distribution, and given the message M Alice (Bob) chooses a message $x \in X$ ($y \in Y$) according to some probability distribution that depends on M. Let us say that X_M (Y_M) is the set of elements in X that have non-zero probability of being selected by Alice (Bob) given the message M. As the protocol has zero error, the rectangle $X_M \times Y_M$ must be monochromatic. As Alice and Bob sample inputs according to the uniform distribution, every $(x, y) \in X \times Y$ must be covered. Hence the protocol gives rise to a monochromatic cover of M_f with only 2^k rectangles and hence $C(f) \leq 2^k$.

Next we show that $\overset{\circ}{R}_0(f) \leq \log_2 C^D(f)$. Suppose a disjoint monochromatic cover of M_f with 2^k rectangles exist. Say, the cover contains the rectangles $R_1, \ldots, R_{2^k}$ and $R_i = X_i \times Y_i$. We build a sampling protocol. A message $i \in \{1, \ldots, 2^k\}$ is picked with probability proportional to the area of R_i. Given the message i, Alice picks a random element $x \in X_i$, and Bob picks a random element $y \in Y_i$. It is easy to verify that as the cover is disjoint, this results in the uniform distribution over $X \times Y$ along with the value of $f(x, y)$. Hence $\overset{\circ}{R}_0(f) \leq k$. □

9. Acknowledgments. We thank Dorit Aharonov, Ike Chuang, Michael Nielsen and Steven Rudich for very helpful discussions. We thank the anonymous referees for many helpful comments.

REFERENCES

[1] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *ACM Symposium on Theory of Computing (STOC)*, pages 20–30, 1998.

[2] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. In *Journal of the ACM, 49(4)*, pages 496–511, 2002.

[3] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *IEEE Symposium on Foundations of Computer Science*, pages 337–347, 1986.

[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)*, pages 63–68, 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, pages 553–558, 1992.

[8] L. K. Grover. A fast quantum mechanical algorithm for database search. In *ACM Symposium on the Theory of Computing (STOC)*, pages 212–219, 1996.

[9] 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, pages 177–183.

[10] R. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1987.

[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*, 25, 1979.

[13] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 369–376, 1999.

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

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

[16] A. A. Razborov. *Quantum communication complexity of symmetric predicates. Izvestiya Math*, 6, 2002. In Russian. English version at quant-ph/0204025

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

[18] R. Raz. Exponential separation of quantum and classical communication complexity. In *ACM Symposium on Theory of Computing (STOC)*, pages 358–367, 1999.

[19] A. C. Yao. Quantum circuit complexity. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 352–361, 1993.

[20] A. C. Yao. Some complexity questions related to distributive computing. In *ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.