A Hypercontractive Inequality for Matrix-Valued Functions with Applications to Quantum Computing and LDCs

Avraham Ben-Aroya* Oded Regev† Ronald de Wolf‡

Abstract

The Bonami-Beckner hypercontractive inequality is a powerful tool in Fourier analysis of real-valued functions on the Boolean cube. In this paper we present a version of this inequality for matrix-valued *functions on the Boolean cube. Its proof is based on a powerful inequality by Ball, Carlen, and Lieb. We also present a number of applications. First, we analyze maps that encode n classical bits into m qubits, in such a way that each set of k bits can be recovered with some probability by an appropriate measurement on the quantum encoding; we show that if $m < 0.7n$, then the success probability is exponentially small in k. This result may be viewed as a direct product version of Nayak's quantum random access code bound. It in turn implies strong direct product theorems for the one-way quantum communication complexity of Disjointness and other problems. Second, we prove that error-correcting codes that are locally decodable with 2 queries require length exponential in the length of the encoded string. This gives what is arguably the first "non-quantum" proof of a result originally derived by Kerenidis and de Wolf using quantum information theory.*

1 Introduction

Fourier analysis of real-valued functions on the Boolean cube is widely used in the theory of computing. Applications include analyzing the influence of variables on Boolean functions [28], probabilistically-checkable proofs and associated hardness of approximation [21], analysis of threshold phenomena [29], noise stability [39, 44], voting schemes [46], learning under the uniform distribution [37, 38, 25, 40], communication complexity [47, 32, 17], etc.

One of the main technical tools in this area is a hypercontractive inequality that is sometimes called the *Bonami-Beckner inequality* [10, 5], though its history would also justify other names (see Lecture 16 of [45] for some background and history). For a fixed $\rho \in [0, 1]$, consider the linear operator T_ρ on the space of all functions $f : \{0, 1\}^n \to \mathbb{R}$ defined by $(T_\rho(f))(x) = \mathbb{E}_y[f(y)]$, where the expectation is taken over y obtained from x by negating each bit independently with probability $(1 - \rho)/2$. In other words, the value of $T_\rho(f)$ at a point x is obtained by averaging the values of f over a certain neighborhood of x. One important property of T_ρ for $\rho < 1$ is that it has a "smoothing" effect: any "high peaks" present in f are smoothed out in $T_\rho(f)$. The hypercontractive inequality formalizes this intuition. To state it precisely, define the p-norm of a function f by $\|f\|_p = (\frac{1}{2^n}\sum_x |f(x)|^p)^{1/p}$. It is not difficult to prove that the norm is nondecreasing with p. Also, the higher p is, the more sensitive the norm becomes to peaks in the function f. The hypercontractive inequality says that for certain $q > p$, the q-norm of $T_\rho(f)$ is upper bounded by the p-norm of f. This exactly captures the intuition that $T_\rho(f)$ is a smoothed version of f: even though we are considering a higher norm, the norm does not increase. More precisely, the inequality says that as long as $1 \leq p \leq q$ and $\rho \leq \sqrt{(p-1)/(q-1)}$, we have

$$\|T_\rho(f)\|_q \leq \|f\|_p. \tag{1}$$

The most interesting case for us is when $q = 2$, since in this case one can view the inequality as a statement about the Fourier coefficients of f, as we describe next. Let us first recall some basic definitions from Fourier analysis. For every $S \subseteq [n]$ (which by some abuse of notation we will also view as an n-bit string) and $x \in \{0, 1\}^n$, define $\chi_S(x) = (-1)^{x \cdot S}$ to be the parity of the bits of x indexed by S. The *Fourier transform* of a function $f : \{0, 1\}^n \to \mathbb{R}$ is the function $\widehat{f} : \{0, 1\}^n \to \mathbb{R}$ defined by

$$\widehat{f}(S) = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x).$$

[*] School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities, by the Israel Science Foundation, and by the European Commission under the Integrated Project QAP funded by the IST directorate as Contract Number 015848.

[†] School of Computer Science, Tel-Aviv University, Tel-Aviv 69978, Israel. Supported by the Binational Science Foundation, by the Israel Science Foundation, and by the EC under QAP.

[‡] Centrum voor Wiskunde en Informatica (CWI), Amsterdam, The Netherlands. Supported by a Veni grant from the Netherlands Organization for Scientific Research (NWO), and by the EC under QAP.

These $\widehat{f}(S)$ are the *Fourier coefficients* of f; the value $\widehat{f}(S)$ may be viewed as measuring the correlation between f and the parity function χ_S. Since the functions χ_S form an orthonormal basis of the space of all functions from $\{0,1\}^n$ to $\mathbb{R}$, we can express f in terms of its Fourier coefficients as

$$f = \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S. \tag{2}$$

Using the same reasoning we obtain Parseval's identity, $\|f\|_2 = \left(\sum_{S\subseteq[n]} \widehat{f}(S)^2\right)^{1/2}$. The operator T_ρ has a particularly elegant description in terms of the Fourier coefficients. It simply multiplies each Fourier coefficient $\widehat{f}(S)$ by a factor of $\rho^{|S|}$: $T_\rho(f) = \sum_{S\subseteq[n]} \rho^{|S|}\widehat{f}(S)\chi_S$. The higher $|S|$ is, the stronger the Fourier coefficient $\widehat{f}(S)$ is "attenuated" by T_ρ. Using Parseval's identity, we can now write the hypercontractive inequality (1) for the case $q = 2$ as follows. For every $p \in [1,2]$,

$$\sum_{S\subseteq[n]} (p-1)^{|S|}\widehat{f}(S)^2 \le \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} |f(x)|^p\right)^{2/p} \tag{3}$$

This gives an upper bound on a weighted sum of the squared Fourier coefficients of f, where each coefficient is attenuated by a factor $(p-1)^{|S|}$. We are interested in generalizing this hypercontractive inequality to *matrix-valued* functions. Let $\mathcal{M}$ be the space of $d \times d$ complex matrices and suppose we have a function $f : \{0,1\}^n \to \mathcal{M}$. For example, a natural scenario where this arises is in quantum information theory, if we assign to every $x \in \{0,1\}^n$ some m-qubit *density matrix* $f(x)$ (so $d = 2^m$). We define the Fourier transform $\widehat{f}$ of a matrix-valued function f exactly as before:

$$\widehat{f}(S) = \frac{1}{2^n}\sum_{x\in\{0,1\}^n} f(x)\chi_S(x).$$

The Fourier coefficients $\widehat{f}(S)$ are now also $d \times d$ matrices. Equivalently, we can apply the standard Fourier transform to each i, j-entry separately: $\widehat{f}(S)_{ij} = \widehat{f(\cdot)_{ij}}(S)$. This extension of the Fourier transform to matrix-valued functions is quite natural, and has also been used in, e.g., [42, 16]. Our main tool, proved in Section 3, is an extension of the hypercontractive inequality to matrix-valued functions. For $M \in \mathcal{M}$ with singular values $\sigma_1, \ldots, \sigma_d$, we define its (normalized Schatten) p-norm as $\|M\|_p = (\frac{1}{d}\sum_{i=1}^d \sigma_i^p)^{1/p}$.

Theorem 1 *For every $f : \{0,1\}^n \to \mathcal{M}$ and $1 \le p \le 2$,*

$$\sum_{S\subseteq[n]} (p-1)^{|S|}\left\|\widehat{f}(S)\right\|_p^2 \le \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|f(x)\|_p^p\right)^{2/p}.$$

This is the analogue of Eq. (3) for matrix-valued functions, with p-norms replacing absolute values. The case $n = 1$ can be seen as a geometrical statement that extends the familiar parallelogram law in Euclidean geometry and is closely related to the notion of uniform convexity. This case was first proven for certain values of p by Tomczak-Jaegermann [50] and then in full generality by Ball, Carlen, and Lieb [3]. Among its applications are the work of Carlen and Lieb on fermion fields [14], and the more recent work of Lee and Naor on metric embeddings [36].

To the best of our knowledge, the general case $n \ge 1$ has not appeared before.[1] Its proof is not difficult, and follows by induction on n, similar to the proof of the usual hypercontractive inequality.[2] Although one might justly regard Theorem 1 as a "standard" corollary of the result by Ball, Carlen, and Lieb, such "tensorized inequalities" tend to be extremely useful (see, e.g., [9, 19]) and we believe that the matrix-valued hypercontractive inequality will have more applications in the future.

1.1 k-out-of-n random access codes

Our main application of Theorem 1 is for the following information-theoretic problem. Suppose we want to encode n-bit string x into m bits or qubits, such that for any set $S \subseteq [n]$ of k indices, the k-bit substring x_S can be recovered with probability at least p by an appropriate measurement on the encoding. We are allowed to use probabilistic encodings here, so the encoding need not be a function mapping x to a fixed classical string or a fixed quantum pure state. We will call such encodings *k-out-of-n random access codes*, since they allow us to access any set of k out of n bits. As far as we know, for $k > 1$ neither the classical nor the quantum case has been studied before. Here we focus on the quantum case, because our lower bounds for quantum encodings of course also apply to classical encodings.

We are interested in the tradeoff between the length m of the quantum random access code, and the success probability p. Clearly, if $m \ge n$ then we can just use the identity encoding to obtain $p = 1$. If $m < n$ then by Holevo's theorem [23] our encoding will be "lossy", hence $p < 1$. The case $k = 1$ was first studied by Ambainis et al. [2], who showed that if p is bounded away from 1/2, then $m = \Omega(n/\log n)$. Nayak [41] subsequently strengthened this to $m \ge (1 - H(p))n$, where $H(\cdot)$ is the binary entropy function. This bound is optimal up to an additive

[1] A different generalization of Bonami-Beckner was given by Borell [11]. His generalization, however, is an easy corollary of Bonami-Beckner itself and is hence relatively weak (although it does apply to any Banach space, not just to the space of matrices with the Schatten p-norm).

[2] We remark that Carlen and Lieb's proof in [14] also uses induction and has some superficial resemblance to the proof given here. Their induction, however, is on the *dimension* of the matrices (or more precisely, the number of fermions), and moreover leads to an entirely different inequality.

$\log n$ term both for classical and quantum encodings. The intuition of Nayak's proof is that, for average i, the encoding only contains $m/n < 1$ bits of information about the bit x_i, which limits our ability to predict x_i given the encoding.

Now suppose that $k > 1$, and m is much smaller than n. Clearly, for predicting one specific bit x_i, with i uniformly chosen, Nayak's result applies, and we will have a success probability that is bounded away from 1. But intuitively this should apply to each of the k bits that we need to predict. Moreover, these k success probabilities should not be very correlated, so we expect an overall success probability that is exponentially small in k. Nayak's proof does not generalize to the case $k \gg 1$ (or at least, we do not know how to do it). The reason it fails is the following. Suppose we probabilistically encode $x \in \{0,1\}^n$ as follows: with probability 1/4 our encoding is x itself, and with probability 3/4 our encoding is the empty string. Then the average length of the output (and hence the entropy or amount of information in the encoding) is only $n/4$ bits, or 1/4 bit for an average x_i. Yet from this encoding one can predict *all of* x with success probability $1/4$! Hence, if we want to prove our intuition, we should make use of the fact that the encoding is always confined to a 2^m-dimensional space (a property which the above example lacks). The new hypercontractive inequality offers an alternative approach—in fact the only alternative approach to entropy-based methods that we are aware of in quantum information. Applying the inequality to the matrix-valued function that gives the encoding implies $p \leq 2^{-\Omega(k)}$ if $m \ll n$. More precisely:

Theorem 2 *For any $\eta > 2\ln 2$ there exists a constant C_η such that if n/k is large enough then for any k-out-of-n quantum random access code on m qubits, the success probability satisfies $p \leq C_\eta \left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\eta m}{n}}\right)^k$.*

In particular, the success probability is exponentially small in k if $m/n < 1/(2\ln 2) \approx 0.721$. Notice that for very small m/n the bound on p gets close to 2^{-k}, which is what one gets by guessing the k-bit answer randomly. We also obtain bounds if k is close to n, but these are a bit harder to state. We believe that the theorem can be extended to the case that $m/n > 1/(2\ln 2)$, although proving this would probably require a strengthening of the inequality by Ball, Carlen, and Lieb. Luckily, in all our applications we are free to choose a small enough m. Finally, we note that in contrast to Nayak's approach, our proof does not use the strong subadditivity of von Neumann entropy.

The König-Renner result. Independently but subsequent to our work (which first appeared on the arxiv preprint server in May 2007), König and Renner [34] used sophisticated quantum information theoretic arguments to show a result with a similar flavor to ours. Each of the results is tuned for different scenarios. In particular, the results are incomparable, and our applications to direct product theorems do not follow from their result, nor do their applications follow from our result. We briefly describe the distinction.

Let $X = X_1, \ldots, X_n$ be random variables, not necessarily uniformly distributed or even independent. Suppose that each $X_i \in \{0,1\}^b$. Suppose further that the "smooth min-entropy of X relative to a quantum state ρ" is at least h (see [34] for the precise technical definitions). If we randomly pick r distinct indices $i_1, \ldots, i_r$, then intuitively the smooth min-entropy of $X' = X_{i_1}, \ldots, X_{i_r}$ relative to ρ should not be much smaller than hr/n. König and Renner show that if b is larger than n/r then this is indeed the case, except with probability exponentially small in r. Note that they are picking b-bit blocks $X_{i_1}, \ldots, X_{i_r}$ instead of individual bits, but this can also be viewed as picking (not quite uniformly) $k = rb$ bits from a string of nb bits.

On the one hand, the constants in their bounds are essentially optimal, while ours are a factor $2\ln 2$ off from what we expect they should be. Also, while they need few assumptions on the random variables $X_1, \ldots, X_n$ and on the quantum encoding, we assume the variables are uniformly distributed bits, and our encoding is confined to a 2^m-dimensional space. We can in fact slightly relax both the assumption on the input and the encoding, but do not discuss these relaxations since they are of less interest to us. Finally, their result still works if the indices $i_1, \ldots, i_r$ are not sampled uniformly, but are sampled in some randomness-efficient way. This allows them to obtain efficient key-agreement schemes in a cryptographic model where the adversary can only store a bounded number of quantum bits.

On the other hand, our result works even if only a small number of bits is sampled, while theirs kicks in when the number of bits being sampled ($k = rb$) is at least the square-root of the total number of bits nb. This is not very explicit in their paper, but can be seen by observing that the parameter $\kappa = n/(rb)$ on page 8 and in Corollary 6.19 needs to be at most a constant (whence the assumption that b is larger than n/r). So the total number of bits is $nb = O(rb^2) = O(r^2b^2) = O(k^2)$. Since we are interested in small as well as large k, this limitation of their approach is significant. A final distinction is in the length of the proof. While the information-theoretic intuition in their paper is clear and well-explained, the technical details result in a proof which is significantly longer than ours.

1.2 Direct product theorems

Our result for k-out-of-n random access codes has the flavor of a direct product theorem: the success probability of performing a certain task on k instances (i.e., k distinct indices) goes down exponentially with k. In Section 5, we use this to prove a new strong direct product theorem for one-way communication complexity.

Consider the 2-party Disjointness function: Alice receives input $x \in \{0,1\}^n$, Bob receives input $y \in \{0,1\}^n$, and they want to determine whether the sets represented by their inputs are disjoint, i.e. whether $x_i y_i = 0$ for all $i \in [n]$. They want to do this while communicating as few qubits as possible (allowing some small error probability, say 1/3). We can either consider one-way protocols, where Alice sends one message to Bob who then computes the output; or two-way protocols, which are interactive. The quantum communication complexity of Disjointness is fairly well understood: it is $\Theta(n)$ qubits for one-way protocols [13], and $\Theta(\sqrt{n})$ qubits for two-way protocols [12, 24, 1, 48].

Now consider the case of k independent instances: Alice receives inputs $x_1, \ldots, x_k$ (each of n bits), Bob receives $y_1, \ldots, y_k$, and their goal is to compute all k bits $\mathrm{DISJ}_n(x_1, y_1), \ldots, \mathrm{DISJ}_n(x_k, y_k)$. Klauck et al. [33] proved an optimal direct product theorem for *two-way* quantum communication: every protocol that communicates fewer than $\alpha k\sqrt{n}$ qubits (for some small constant $\alpha > 0$) will have a success probability that is exponentially small in k. Surprisingly, prior to our work no strong direct product theorem was known for the usually simpler case of *one-way* communication—not even for *classical* one-way communication.[3] In Section 5 we derive such a theorem from our k-out-of-n random access code lower bound: if $\eta > 2\ln 2$, then every one-way quantum protocol that sends fewer than kn/η qubits will have success probability at most $2^{-\Omega(k)}$.

These results can straightforwardly be generalized to a bound for all functions in terms of their *VC-dimension*. If f has VC-dimension d, then any one-way quantum protocol for computing k independent copies of f that sends kd/η qubits, has success probability $2^{-\Omega(k)}$. For simplicity, Section 5 only presents the case of Disjointness. By the work of Beame et al. [4], direct product theorems imply lower bounds on *3-party* protocols where the first party sends only one message. We elaborate on this in the full version [7].

1.3 Locally decodable codes

A locally decodable error-correcting code (LDC) $C : \{0,1\}^n \to \{0,1\}^N$ encodes n bits into N bits, in such a way that each encoded bit can be recovered from a noisy codeword by a randomized decoder that queries only a small number q of bit-positions in that codeword. Such codes have applications in a variety of different complexity-theoretic and cryptographic settings; see for instance Trevisan's survey and the references therein [51]. The main theoretical issue in LDCs is the tradeoff between q and N. The best known constructions of LDCs with constant q have a length N that is sub-exponential in n but still superpolynomial [15, 6, 54]. On the other hand, the only superpolynomial *lower* bound known for general LDCs is the tight bound $N = 2^{\Omega(n)}$ for $q = 2$ due to Kerenidis and de Wolf [31] (generalizing an earlier exponential lower bound for *linear* codes by [18]). Rather surprisingly, the proof of [31] relied heavily on techniques from quantum information theory: despite being a result purely about classical codes and classical decoders, the quantum perspective was crucial for their proof. In particular, they show that the two queries of a classical decoder can be replaced by one quantum query, then they turn this quantum query into a random access code for the encoded string x, and finally invoke Nayak's lower bound for quantum random access codes.

In Section 6 we reprove an exponential lower bound on N for the case $q = 2$ without invoking any quantum information theory: we just use classical reductions, matrix analysis, and the hypercontractive inequality for matrix-valued functions. Hence it is a classical (non-quantum) proof as asked for by Trevisan [51, Open question 3 in Section 3.6].[4] It should be noted that this new proof is still quite close in spirit (though not terminology) to the quantum proof of [31]. This is not too surprising given the fact that the proof of [31] uses Nayak's lower bound on random access codes, generalizations of which follow from the hypercontractive inequality. We discuss the similarities and differences between the two proofs in Section 6.

We feel the merit of this new approach is not so much in giving a partly new proof of the known lower bound on 2-query LDCs, but in its potential application to codes with more than 2 queries. Recently Yekhanin [54] constructed 3-query LDCs with $N = 2^{O(n^{1/32582657})}$ (and $N = 2^{n^{O(1/\log\log n)}}$ for infinitely many n if there exist infinitely many Mersenne primes). For $q = 3$, the best known lower bounds on N are slightly less than n^2 [30, 31, 53]. Despite considerable effort, this gap still looms large. Our hope is that our approach can be generalized to 3 or more queries. Specifically, what we would need is a generalization of tensors of rank 2 (i.e., matrices) to tensors of rankr q; an appropriate tensor norm; and a generalization of the hypercontractive inequality from matrix-valued to tensor-valued functions. Some preliminary progress towards this goal was obtained in [22].

2 Preliminaries

Norms: The p-norm of a d-dimensional vector v is $\|v\|_p = \left(\frac{1}{d}\sum_{i=1}^d |v_i|^p\right)^{1/p}$. We extend this to matrices by defining the (normalized Schatten) p-norm of a matrix

[3] Recently and independently of our work, Jain et al. [26] did manage to prove such a direct product theorem for classical one-way communication, based on information-theoretic techniques.

[4] Alex Samorodnitsky has been developing a classical proof along similar lines in the past two years. However, as he told us at the time of writing [49], his proof is still incomplete.

$A \in \mathbb{C}^{d\times d}$ as $\|A\|_p = \left(\frac{1}{d}\mathrm{Tr}|A|^p\right)^{1/p}$. This is equivalent to the p-norm of the vector of singular values of A. For diagonal matrices this definition coincides with the one for vectors. For convenience we defined all norms to be under the normalized counting measure, even though for matrices this is nonstandard. The advantage of the normalized norm is that it is nondecreasing with p. We also define the *trace norm* $\|A\|_{\mathrm{tr}}$ of a matrix A as the sum of its singular values, hence we have $\|A\|_{\mathrm{tr}} = d\|A\|_1$ for any $d \times d$ matrix A.

Quantum states: An m-qubit *pure state* is a superposition $|\phi\rangle = \sum_{z\in\{0,1\}^m} \alpha_z|z\rangle$ over all classical m-bit states. The α_z's are complex numbers called *amplitudes*, and $\sum_z |\alpha_z|^2 = 1$. Hence a pure state $|\phi\rangle$ is a unit vector in $\mathbb{C}^{2^m}$. Its complex conjugate (a row vector with entries conjugated) is denoted $\langle\phi|$. The inner product between $|\phi\rangle = \sum_z \alpha_z|z\rangle$ and $|\psi\rangle = \sum_z \beta_z|z\rangle$ is the dot product $\langle\phi|\cdot|\psi\rangle = \langle\phi|\psi\rangle = \sum_z \alpha_z^*\beta_z$. An m-qubit *mixed state* (or *density matrix*) $\rho = \sum_i p_i|\phi_i\rangle\langle\phi_i|$ corresponds to a probability distribution over m-qubit pure states, where $|\phi_i\rangle$ is given with probability p_i. The eigenvalues $\lambda_1,\dots,\lambda_d$ of ρ are non-negative reals that sum to 1, so they form a probability distribution. If ρ is pure then one eigenvalue is 1 while all others are 0. Hence for any $p \geq 1$, the maximal p-norm is achieved by pure states:

$$\|\rho\|_p^p = \frac{1}{d}\sum_{i=1}^d \lambda_i^p \leq \frac{1}{d}\sum_{i=1}^d \lambda_i = \frac{1}{d}. \tag{4}$$

A k-outcome *positive operator-valued measurement* (POVM) is given by k positive semidefinite operators $E_1,\dots,E_k$ with the property that $\sum_{i=1}^k E_i = I$. When this POVM is applied to a mixed state ρ, the probability of the ith outcome is given by the trace $\mathrm{Tr}(E_i\rho)$. The following well known fact gives the close relationship between trace distance and distinguishability of density matrices:

Fact 3 *The best possible measurement to distinguish two density matrices ρ_0 and ρ_1 has bias $\frac{1}{2}\|\rho_0-\rho_1\|_{\mathrm{tr}}$.*

Here "bias" is defined as twice the success probability, minus 1. We refer to Nielsen and Chuang [43] for more details.

3 The hypercontractive inequality for matrix-valued functions

Here we prove Theorem 1, based on the following powerful inequality by Ball et al. [3] (they give an equivalent statement for the usual unnormalized Schatten p-norm).

Lemma 4 *([3, Theorem 1]) For any matrices A, B and any $1 \leq p \leq 2$, it holds that*

$$\left\|\frac{A+B}{2}\right\|_p^2 + (p-1)\left\|\frac{A-B}{2}\right\|_p^2 \leq \left(\frac{\|A\|_p^p + \|B\|_p^p}{2}\right)^{2/p}.$$

Theorem 1 For any $f : \{0,1\}^n \to \mathcal{M}$ and $1 \leq p \leq 2$,

$$\sum_{S\subseteq[n]} (p-1)^{|S|}\left\|\widehat{f}(S)\right\|_p^2 \leq \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|f(x)\|_p^p\right)^{2/p}.$$

Proof: By induction. The case $n = 1$ follows from Lemma 4 by setting $A = f(0)$ and $B = f(1)$, and noting that $(A+B)/2$ and $(A-B)/2$ are exactly the Fourier coefficients $\widehat{f}(0)$ and $\widehat{f}(1)$.

We now assume the lemma holds for n and prove it for $n+1$. Let $f : \{0,1\}^{n+1} \to \mathcal{M}$ be some matrix-valued function. For $i \in \{0,1\}$, let $g_i = f|_{x_{n+1}=i}$ be the function obtained by fixing the last input bit of f to i. We apply the induction hypothesis on g_0 and g_1 to obtain

$$\sum_{S\subseteq[n]} (p-1)^{|S|}\|\widehat{g_0}(S)\|_p^2 \leq \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|g_0(x)\|_p^p\right)^{2/p}$$

$$\sum_{S\subseteq[n]} (p-1)^{|S|}\|\widehat{g_1}(S)\|_p^2 \leq \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n} \|g_1(x)\|_p^p\right)^{2/p}.$$

Raise each of these two equations to the $p/2$th power, average them, and take the $p/2$th root. We get

$$\left(\frac{1}{2}\sum_{i\in\{0,1\}}\left(\sum_{S\subseteq[n]} (p-1)^{|S|}\|\widehat{g_i}(S)\|_p^2\right)^{p/2}\right)^{2/p} \tag{5}$$

$$\leq \left(\frac{1}{2^{n+1}}\sum_{x\in\{0,1\}^n}\left(\|g_0(x)\|_p^p + \|g_1(x)\|_p^p\right)\right)^{2/p}$$

$$= \left(\frac{1}{2^{n+1}}\sum_{x\in\{0,1\}^{n+1}} \|f(x)\|_p^p\right)^{2/p}.$$

The right-hand side is the expression we wish to lower bound. To bound the left-hand side, we need:

Lemma 5 (Minkowski's inequality, [20, Theorem 26]) *For any $r_1 \times r_2$ matrix whose rows are given by $u_1,\dots,u_{r_1}$ and whose columns are given by $v_1,\dots,v_{r_2}$, and any $1 \leq q_1 < q_2 \leq \infty$,*

$$\left\|\left(\|v_1\|_{q_2},\dots,\|v_{r_2}\|_{q_2}\right)\right\|_{q_1} \geq \left\|\left(\|u_1\|_{q_1},\dots,\|u_{r_1}\|_{q_1}\right)\right\|_{q_2},$$

i.e., the value obtained by taking the q_2-norm of each column and then taking the q_1-norm of the results, is at least that obtained by first taking the q_1-norm of each row and then taking the q_2-norm of the results.

Consider now the $2^n \times 2$ matrix with entries given by

$$c_{S,i} = 2^{n/2}\left\|(p-1)^{|S|/2}\widehat{g_i}(S)\right\|_p$$

where $i \in \{0,1\}$ and $S \subseteq [n]$. The left-hand side of (5) is

$$\left(\frac{1}{2}\sum_{i\in\{0,1\}}\left(\frac{1}{2^n}\sum_{S\subseteq[n]} c_{S,i}^2\right)^{p/2}\right)^{2/p}$$

$$\geq \frac{1}{2^n}\sum_{S\subseteq[n]}\left(\frac{1}{2}\sum_{i\in\{0,1\}} c_{S,i}^p\right)^{2/p}$$

$$= \sum_{S\subseteq[n]}(p-1)^{|S|}\left(\frac{\|\widehat{g_0}(S)\|_p^p + \|\widehat{g_1}(S)\|_p^p}{2}\right)^{2/p}$$

where the inequality follows from squaring Lemma 5 with $q_1 = p$, $q_2 = 2$. We now apply Lemma 4 to deduce that the above is lower bounded by

$$\sum_{S\subseteq[n]}(p-1)^{|S|}\left(\left\|\frac{\widehat{g_0}(S)+\widehat{g_1}(S)}{2}\right\|_p^2 + \right.$$

$$\left.(p-1)\left\|\frac{\widehat{g_0}(S)-\widehat{g_1}(S)}{2}\right\|_p^2\right) = \sum_{S\subseteq[n+1]}(p-1)^{|S|}\left\|\widehat{f}(S)\right\|_p^2$$

where we used $\widehat{f}(S) = \frac{1}{2}(\widehat{g_0}(S)+\widehat{g_1}(S))$ and $\widehat{f}(S \cup \{n+1\}) = \frac{1}{2}(\widehat{g_0}(S)-\widehat{g_1}(S))$ for any $S \subseteq [n]$. ∎

4 k-out-of-n quantum random access codes

In this section we prove Theorem 2. Recall that a k-out-of-n random access code allows us to encode n bits into m qubits, such that we can recover any k-bit substring with probability at least p. We now define this notion formally. In fact, we consider a somewhat weaker notion where we only measure the success probability for a random k subset, and a random input $x \in \{0,1\}^n$. Since we only prove impossibility results, this clearly makes our results stronger.

Definition 1 *A k-out-of-n quantum random access code on m qubits with success probability p (for short (k,n,m,p)-QRAC), is a map $f : \{0,1\}^n \to \mathbb{C}^{2^m\times 2^m}$ that assigns an m-qubit density matrix $f(x)$ to every $x \in \{0,1\}^n$, and a quantum measurement $\{M_{S,z}\}_{z\in\{0,1\}^k}$ to every set $S \in \binom{[n]}{k}$, with the property that $\mathbb{E}_{x,S}[\mathrm{Tr}(M_{S,x_S}\cdot f(x))] \geq p$, where the expectation is taken over a uniform choice of $x \in \{0,1\}^n$ and $S \in \binom{[n]}{k}$, and x_S denotes the k-bit substring of x specified by S.*

To prove Theorem 2, we introduce another notion of QRAC, called *XOR-QRAC*. Here, the goal is to predict the XOR of the k bits indexed by S (as opposed to guessing all the bits in S). Since one can always predict a bit with probability $\frac{1}{2}$, it is convenient to define the *bias* of the prediction as $\varepsilon = 2p-1$ where p is the probability of a correct prediction. Hence a bias of 1 means that the prediction is always correct, whereas a bias of -1 means that it is always wrong. The advantage of dealing with an XOR-QRAC is that it is easy to express the best achievable prediction bias without any need to introduce measurements. Namely, if $f : \{0,1\}^n \to \mathbb{C}^{2^m\times 2^m}$ is the encoding function, then the best achievable bias in predicting the XOR of the bits in S (over a random $\{0,1\}^n$) is exactly half the trace distance between the average of $f(x)$ over all x with the XOR of the bits in S being 0 and the average of $f(x)$ over all x with the XOR of the bits in S being 1. Using our notation for Fourier coefficients, this can be written simply as $\|\widehat{f}(S)\|_{\mathrm{tr}}$.

Definition 2 *A k-out-of-n XOR quantum random access code on m qubits with bias ε (for short (k,n,m,ε)-XOR-QRAC), is a map $f : \{0,1\}^n \to \mathbb{C}^{2^m\times 2^m}$ that assigns an m-qubit density matrix $f(x)$ to every $x \in \{0,1\}^n$ and has the property that $\mathbb{E}_{S\sim\binom{[n]}{k}}\left[\|\widehat{f}(S)\|_{\mathrm{tr}}\right] \geq \varepsilon$.*

Our hypercontractive inequality allows us to easily derive the following key lemma:

Lemma 6 *Let $f : \{0,1\}^n \to \mathbb{C}^{2^m\times 2^m}$ be any mapping from n-bit strings to m-qubit density matrices. Then for any $0 \leq \delta \leq 1$, we have $\sum_{S\subseteq[n]} \delta^{|S|}\|\widehat{f}(S)\|_{\mathrm{tr}}^2 \leq 2^{2\delta m}$.*

Proof: On one hand, by Theorem 1 and Eq. (4) we have

$$\sum_{S\subseteq[n]}(p-1)^{|S|}\|\widehat{f}(S)\|_p^2 \leq \left(\frac{1}{2^n}\sum_{x\in\{0,1\}^n}\|f(x)\|_p^p\right)^{2/p}$$

$$\leq \left(\frac{1}{2^n}\cdot 2^n \cdot \frac{1}{2^m}\right)^{2/p} = 2^{-2m/p}.$$

On the other hand, by norm monotonicity we have

$$\sum_{S\subseteq[n]}(p-1)^{|S|}\|\widehat{f}(S)\|_p^2 \;\geq\; \sum_{S\subseteq[n]}(p-1)^{|S|}\|\widehat{f}(S)\|_1^2$$

$$= 2^{-2m}\sum_{S\subseteq[n]}(p-1)^{|S|}\|\widehat{f}(S)\|_{\mathrm{tr}}^2$$

Choosing $p = 1+\delta$ and rearranging gives the lemma. ∎

The following is our main theorem regarding XOR-QRAC. In particular, if $k = o(n)$ and $m/n < 1/(2\ln 2) \approx 0.721$, then the bias is exponentially small in k.

Theorem 7 *For any (k,n,m,ε)-XOR-QRAC we have the following bound on the bias: $\varepsilon \leq \left(\frac{(2e\ln 2)m}{k}\right)^{k/2}\binom{n}{k}^{-1/2}$. Hence for any $\eta > 2\ln 2$ there is a constant C_η such that if n/k is large enough then for any (k,n,m,ε)-XOR-QRAC: $\varepsilon \leq C_\eta\left(\frac{\eta m}{n}\right)^{k/2}$.*

Proof: Apply Lemma 6 with $\delta = \frac{k}{(2\ln 2)m}$ and only take the sum on S with $|S| = k$. This gives

$$\mathbb{E}_{S\sim\binom{[n]}{k}} \left\|\widehat{f}(S)\right\|_{\mathrm{tr}}^2 \leq \frac{2^{2\delta m}}{\delta^k}\binom{n}{k}^{-1} = \left(\frac{(2e\ln 2)m}{k}\right)^k \binom{n}{k}^{-1}$$

The first bound on ε now follows by convexity (Jensen's inequality). To derive the second bound, approximate $\binom{n}{k}$ using Stirling's approximation $n! = \Theta(\sqrt{n}(n/e)^n)$:

$$\frac{n!}{k!(n-k)!} = \Theta\left(\sqrt{\frac{n}{k(n-k)}}\left(\frac{n}{k}\right)^k \left(1+\frac{k}{n-k}\right)^{n-k}\right)$$

Now use the fact that for large enough n/k we have $(1 + k/(n-k))^{(n-k)/k} > (2e\ln 2)/\eta$, and absorb the factor $\sqrt{n/k(n-k)} \geq \sqrt{1/k}$ in this approximation. ■

We now derive Theorem 2 from Theorem 7.

Proof of Theorem 2: Consider a (k, n, m, p)-QRAC, given by encoding function f and measurements $\{M_{T,z}\}_{z\in\{0,1\}^k}$ for all $T \in \binom{[n]}{k}$. Define $p_T(w) = \mathbb{E}_x\left[\Pr[z \oplus x_T = w]\right]$ as the distribution on the "error vector" $w \in \{0,1\}^k$ of the measurement outcome $z \in \{0,1\}^k$ when applying $\{M_{T,z}\}$. By definition, we have that $p \leq \mathbb{E}_T[p_T(0^k)]$.

Now suppose we want to predict the parity of the bits of some set S of size at most k. We can do this as follows: uniformly pick a $T \in \binom{[n]}{k}$ that contains S, measure $f(x)$ with $\{M_{T,z}\}$, and output the parity of the bits corresponding to S in the measurement outcome z. Our output is correct if and only if the bits corresponding to S in the error vector w have even parity. Hence the bias β_S of our output is

$$\mathbb{E}_{T:T\supseteq S}\left[\sum_{w\in\{0,1\}^k} p_T(w)\chi_S(w)\right] = 2^k \mathbb{E}_{T:T\supseteq S}\left[\widehat{p_T}(S)\right]$$

(We slightly abuse notation here by viewing S both as a subset of T and as a subset of $[k]$ obtained by identifying T with $[k]$.) Notice that β_S can be upper bounded by the best-achievable bias $\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}$.

Consider the distribution $\mathcal{S}$ on sets S defined as follows: first pick j from the binomial distribution $B(k, 1/2)$ and then uniformly pick $S \in \binom{[n]}{j}$. Notice that the distribution on pairs (S,T) obtained by first choosing $S \sim \mathcal{S}$ and then choosing a uniform $T \supseteq S$ from $\binom{[n]}{k}$ is identical to the one obtained by first choosing uniformly T from $\binom{[n]}{k}$ and then choosing a uniform $S \subseteq T$. This allows us to show that the average bias β_S over $S \sim \mathcal{S}$ is at least p, as follows:

$$\mathbb{E}_{S\sim\mathcal{S}}\left[\beta_S\right] = 2^k \mathbb{E}_{S\sim\mathcal{S},T\supseteq S}\left[\widehat{p_T}(S)\right] = 2^k \mathbb{E}_{T\sim\binom{[n]}{k},S\subseteq T}\left[\widehat{p_T}(S)\right]$$

$$= \mathbb{E}_{T\sim\binom{[n]}{k}}\left[\sum_{S\subseteq T}\widehat{p_T}(S)\right] = \mathbb{E}_{T\sim\binom{[n]}{k}}\left[p_T(0^k)\right] \geq p,$$

where the last equality follows from Eq. (2). On the other hand, using Theorem 7 we obtain

$$\mathbb{E}_{S\sim\mathcal{S}}\left[\beta_S\right] \leq \mathbb{E}_{S\sim\mathcal{S}}\left\|\widehat{f}(S)\right\|_{\mathrm{tr}} = \frac{1}{2^k}\sum_{j=0}^{k}\binom{k}{j}\mathbb{E}_{S\sim\binom{[n]}{j}}\left\|\widehat{f}(S)\right\|_{\mathrm{tr}}$$

$$\leq \frac{1}{2^k}\sum_{j=0}^{k}\binom{k}{j}C_\eta\left(\frac{\eta m}{n}\right)^{j/2} = C_\eta\left(\frac{1}{2}+\frac{1}{2}\sqrt{\frac{\eta m}{n}}\right)^k$$

where the last equality uses the binomial theorem. Combining the two inequalities completes the proof. ■

5 Direct product theorem for one-way quantum communication complexity

The setting of communication complexity is by now well-known, so we will not give formal definitions of protocols etc., referring to [35, 52] instead. Consider the n-bit Disjointness problem in 2-party communication complexity. Alice receives n-bit string x and Bob receives n-bit string y. They interpret these strings as subsets of $[n]$ and want to decide whether their sets are disjoint. In other words, $\mathrm{DISJ}_n(x,y) = 1$ if and only if $x \cap y = \emptyset$. Let $\mathrm{DISJ}_n^{(k)}$ denote k independent instances of this problem. That is, Alice's input is a k-tuple $x_1, \ldots, x_k$ of n-bit strings, Bob's input is a k-tuple $y_1, \ldots, y_k$, and they should output all k bits: $\mathrm{DISJ}_n^{(k)}(x_1, \ldots, x_k, y_1, \ldots, y_k) = \mathrm{DISJ}_n(x_1, y_1), \ldots, \mathrm{DISJ}_n(x_k, y_k)$. The trivial protocol where Alice sends all her inputs to Bob has success probability 1 and communication complexity kn. We want to show that if the total one-way communication is much smaller than kn qubits, then the success probability is exponentially small in k. We will do that by deriving a random access code from the protocol's message.

Lemma 8 *Let $\ell \leq k$. If there is a c-qubit one-way communication protocol for $\mathrm{DISJ}_n^{(k)}$ with success probability σ, then there is an ℓ-out-of-kn quantum random access code of c qubits with success probability $p \geq \sigma\left(1 - \ell/k\right)^\ell$.*

Proof: Consider the following one-way communication setting: Alice has a kn-bit string x, and Bob has ℓ distinct indices $i_1, \ldots, i_\ell \in [kn]$ chosen uniformly from $\binom{[kn]}{\ell}$ and wants to learn the corresponding bits of x. To do this, Alice sends the c-qubit message corresponding to input x in the $\mathrm{DISJ}_n^{(k)}$ protocol. View x as consisting of k disjoint blocks of n bits each. The probability (over the choice of Bob's input) that $i_1, \ldots, i_\ell \in [kn]$ are in ℓ different blocks is

$$\prod_{i=0}^{\ell-1}\frac{kn - in}{kn - i} \geq \left(\frac{kn - \ell n}{kn}\right)^\ell = \left(1 - \frac{\ell}{k}\right)^\ell.$$

If this is the case, Bob chooses his Disjointness inputs $y_1, \ldots, y_k$ as follows. If index i_j is somewhere in block

$b \in [k]$, then he chooses y_b to be the string having a 1 at the position where i_j is, and 0s elsewhere. Note that the correct output for the b-th instance of Disjointness with inputs x and $y_1, \ldots, y_k$ is exactly $1 - x_{i_j}$. Now Bob completes the protocol and gets a k-bit output for the k-fold Disjointness problem. A correct output tells him the ℓ bits he wants to know (he can just disregard the outcomes of the other $k - \ell$ instances). Overall the success probability is at least $\sigma(1 - \ell/k)^\ell$. Therefore, the random access code that encodes x by Alice's message proves the lemma. ■

Combining the previous lemma with our earlier upper bound on p for ℓ-out-of-kn quantum random access codes (Theorem 2), we obtain the following upper bound on the success probability σ of c-qubit one-way communication protocols for $\mathrm{DISJ}_n^{(k)}$. For every $\eta > 2 \ln 2$ there exists a constant C_η such that:

$$\begin{aligned}\sigma &\leq 2p(1-\ell/k)^{-\ell} \\ &\leq 2C_\eta \left(\left(\frac{1}{2} + \frac{1}{2}\sqrt{\frac{\eta(c + O(k + \log(kn)))}{kn}} \right) \left(\frac{k}{k-\ell} \right) \right)^\ell\end{aligned}$$

Choosing ℓ a sufficiently small constant fraction of k (depending on η), we obtain our direct product theorem:

Theorem 9 *For any $\eta > 2 \ln 2$ the following holds: for any large enough n and any k, every one-way quantum protocol for $\mathrm{DISJ}_n^{(k)}$ that communicates $c \leq kn/\eta$ qubits, has success probability $\sigma \leq 2^{-\Omega(k)}$ (where $\Omega(\cdot)$ depends on η).*

The above strong direct product theorem (SDPT) bounds the success probability for protocols that are required to compute *all* k instances correctly. We call this a *zero-error* SDPT. What if we settle for a weaker notion of "success", namely getting a $(1 - \varepsilon)$-fraction of the k instances right, for some small $\varepsilon > 0$? An *ε-error SDPT* is a theorem to the effect that even in this case the success probability is exponentially small. An ε-error SDPT follows from a zero-error SDPT as follows. Run an ε-error protocol with success probability p ("success" now means getting $1 - \varepsilon$ of the k instances right), guess up to εk positions and change them. With probability at least p, the number of errors of the ε-error protocol is at most εk, and with probability at least $1/\sum_{i=0}^{\varepsilon k} \binom{k}{i}$ we now have corrected all those errors. Since $\sum_{i=0}^{\varepsilon k} \binom{k}{i} \leq 2^{kH(\varepsilon)}$ (see, e.g., [27, Corollary 23.6]), we have a protocol that computes all instances correctly with success probability $\sigma \geq p2^{-kH(\varepsilon)}$. If we have a zero-error SDPT that bounds $\sigma \leq 2^{-\gamma k}$ for some $\gamma > H(\varepsilon)$, then it follows that p must be exponentially small as well: $p \leq 2^{-(\gamma - H(\varepsilon))k}$. Hence Theorem 9 implies:

Theorem 10 *For any $\eta > 2 \ln 2$ there exists an $\varepsilon > 0$ such that the following holds: for every one-way quantum protocol for $\mathrm{DISJ}_n^{(k)}$ that communicates $c \leq kn/\eta$ qubits, its probability to compute at least a $(1 - \varepsilon)$-fraction of the k instances correctly is at most $2^{-\Omega(k)}$.*

6 Lower bounds on locally decodable codes

When analyzing locally decodable codes, it will be convenient to view bits as elements of $\{\pm 1\}$ instead of $\{0, 1\}$. Formally, a locally decodable code is defined as follows.

Definition 3 *$C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a (q, δ, ε)-*locally decodable code *(LDC) if there is a randomized decoding algorithm A such that*

1. *For all $x \in \{\pm 1\}^n$, $i \in [n]$, and $y \in \{\pm 1\}^N$ with Hamming distance $d(C(x), y) \leq \delta N$, we have $\Pr[A^y(i) = x_i] \geq 1/2 + \varepsilon$. Here $A^y(i)$ is the random variable that is A's output given input i and oracle y.*
2. *A makes at most q queries to y, non-adaptively.*

In the full version of this paper [7] we show this implies the following: For each $i \in [n]$, there is a set M_i of at least $\delta \varepsilon N/q^2$ disjoint tuples, each of at most q elements from $[N]$, and a sign $a_{i,Q} \in \{\pm 1\}$ for each $Q \in M_i$, such that

$$\mathbb{E}_x[a_{i,Q} x_i \prod_{j \in Q} C(x)_j] \geq \frac{\varepsilon}{2^q},$$

where the expectation is uniformly over all $x \in \{\pm 1\}^n$. In other words, the parity of each of the tuples in M_i allows us to predict x_i with non-trivial bias (averaged over all x).

Kerenidis and de Wolf [31] used quantum information theory to show the bound $N = 2^{\Omega(\delta \varepsilon^2 n)}$ on the length of 2-query LDCs. Using the hypercontractive inequality we prove a bound with slightly worse dependence on ε and δ (which can probably be improved by more careful analysis).

Theorem 11 *If $C : \{\pm 1\}^n \to \{\pm 1\}^N$ is a $(2, \delta, \varepsilon)$-LDC, then $N = 2^{\Omega(\delta^2 \varepsilon^4 n)}$.*

Proof: Define $f(x)$ as the $N \times N$ matrix whose (i, j)-entry is $C(x)_i C(x)_j$. Since $f(x)$ has rank 1 and its N^2 entries are all $+1$ or -1, its only non-zero singular value is N. Hence $\|f(x)\|_p^p = N^{p-1}$. Consider the $N \times N$ matrices $\widehat{f}(\{i\})$ that are the Fourier transform of f at the singleton sets $\{i\}$:

$$\widehat{f}(\{i\}) = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x) x_i.$$

We want to lower bound $\left\|\widehat{f}(\{i\})\right\|_p$. With the above notation, each set M_i consists of at least $\delta \varepsilon N/4$ disjoint pairs of indices.[5] For simplicity assume $M_i =$

[5] Actually some of the elements of M_i may be singletons. Dealing with this is a technicality that we will ignore here to simplify the presentation.

$\{(1,2),(3,4),(5,6),\ldots\}$. The 2×2 submatrix in the upper left corner of $f(x)$ is

$$\begin{pmatrix} 1 & C(x)_1C(x)_2 \\ C(x)_1C(x)_2 & 1 \end{pmatrix}.$$

Since $(1,2)\in M_i$, $\mathbb{E}_x[C(x)_1C(x)_2x_ia_{i,(1,2)}]\in[\varepsilon/4,1]$. Hence the submatrix in the upper left corner of $\widehat{f}(\{i\})$ is

$$\begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}$$

for some a with $|a|\in[\varepsilon/4,1]$. The same is true for each of the first $\delta\varepsilon N/4$ 2×2 diagonal blocks of $\widehat{f}(\{i\})$ (each such 2×2 block corresponds to a pair in M_i). Let P be the $N\times N$ permutation matrix that swaps rows 1 and 2, swaps rows 3 and 4, etc. Then the first $\delta\varepsilon N/2$ diagonal entries of $F_i = P\widehat{f}(\{i\})$ all have absolute value in $[\varepsilon/4,1]$. We have

$$\begin{aligned}\left\|\widehat{f}(\{i\})\right\|_p &= \|F_i\|_p \geq \|\mathrm{diag}(F_i)\|_p \\ &\geq \left(\frac{1}{N}(\delta\varepsilon N/2)(\varepsilon/4)^p\right)^{1/p} = (\delta\varepsilon/2)^{1/p}\varepsilon/4\end{aligned}$$

where the first inequality follows from [8, Eq. (IV.52) on p. 97]. By Theorem 1, for any $p\in[1,2]$ we have

$$\begin{aligned}n(p-1)(\delta\varepsilon/2)^{2/p}(\varepsilon/4)^2 &\leq \sum_{i=1}^{n}(p-1)\left\|\widehat{f}(\{i\})\right\|_p^2 \\ &\leq \left(\frac{1}{2^n}\sum_x \|f(x)\|_p^p\right)^{2/p} = N^{2(p-1)/p}.\end{aligned}$$

Choosing $p = 1+1/\log N$ implies the result. ■

Let us elaborate on the similarities and differences between this proof and the quantum proof of [31]. On the one hand, the present proof makes no use of quantum information theory. It only uses the well known version of LDCs mentioned after Definition 3, some basic matrix analysis, and our hypercontractive inequality for matrix-valued functions. On the other hand, the proof may still be viewed as a translation of the original quantum proof to a different language. The quantum proof defines, for each x, a $\log(N)$-qubit state $|\phi(x)\rangle$ which is the uniform superposition over the N indices of the codeword $C(x)$. It then proceeds in two steps: (1) by viewing the elements of M_i as 2-dimensional projectors in a quantum measurement of $|\phi(x)\rangle$, we can with good probability recover the parity $C(x)_jC(x)_k$ for a random element (j,k) of the matching M_i. Since that parity has non-trivial correlation with x_i, the states $|\phi(x)\rangle$ form a quantum random access code: they allow us to recover each x_i with decent probability (averaged over all x); (2) the quantum proof then invokes Nayak's linear lower bound on the number of qubits of a random access code to conclude $\log N = \Omega(n)$. The present proof mimics this quantum proof quite closely: the matrix $f(x)$ is, up to normalization, the density matrix corresponding to the state $|\phi(x)\rangle$; the fact that matrix $\widehat{f}(\{i\})$ has fairly high norm corresponds to the fact that the parity produced by the quantum measurement has fairly good correlation with x_i; and finally, our invocation of Theorem 1 replaces (but is not identical to) the linear lower bound on quantum random access codes. We feel that by avoiding any explicit use of quantum information theory, the new proof holds some promise for extension to $q\geq 3$.

Acknowledgments. This work started while the second author was visiting the group in CWI Amsterdam, and he would like to thank them for their hospitality. Part of this work was done while the authors were visiting the Institut Henri Poincaré in Paris, as part of the program "Quantum information, computation and complexity", and we would like to thank the organizers for their efforts. We thank Shiri Artstein, Julia Kempe, Hartmut Klauck, Robert König, Assaf Naor, Ashwin Nayak, Ryan O'Donnell, Renato Renner, Alex Samorodnitsky, Falk Unger, Emanuele Viola, and Avi Wigderson for useful discussions and comments.

References

[1] S. Aaronson and A. Ambainis. Quantum search of spatial regions. In *44th FOCS*, pp. 200–209, 2003.

[2] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Quantum dense coding and a lower bound for 1-way quantum finite automata. In *31st STOC*, pp. 376–383, 1999.

[3] K. Ball, E. Carlen, and E. Lieb. Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones Mathematicae*, 115:463–482, 1994.

[4] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of set disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[5] W. Beckner. Inequalities in Fourier analysis. *Ann. of Mathematics*, 102:159–182, 1975.

[6] A. Beimel, Y. Ishai, E. Kushilevitz, and J. Raymond. Breaking the $O(n^{1/(2k-1)})$ barrier for information-theoretic Private Information Retrieval. In *43rd FOCS*, pp. 261–270, 2002.

[7] A. Ben-Aroya, O. Regev, and R. de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing, 2007. Preprint at http://arxiv.org/abs/0705.3806.

[8] R. Bhatia. *Matrix Analysis*. Number 169 in Graduate Texts in Mathematics. Springer, New York, 1997.

[9] S. Bobkov. An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. *Ann. of Probability*, 25(1):206–214, 1997.

[10] A. Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Ann. de l'Institut Fourier*, 20(2):335–402, 1970.

[11] C. Borell. On the integrability of Banach space valued Walsh polynomials. In *Séminaire de Probabilités, XIII*

(Univ. Strasbourg, 1977/78), volume 721 of *Lecture Notes in Math.*, pp. 1–3. Springer, Berlin, 1979.

[12] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *30th STOC*, pp. 63–68, 1998.

[13] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *16th Conference on Computational Complexity*, pp. 120–130, 2001.

[14] E. A. Carlen and E. H. Lieb. Optimal hypercontractivity for Fermi fields and related noncommutative integration inequalities. *Communications in Mathematical Physics*, 155(1):27–46, 1993.

[15] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan. Private information retrieval. *J. of the ACM*, 45(6):965–981, 1998.

[16] S. Fehr and C. Schaffner. Randomness extraction via delta-biased masking in the presence of a quantum attacker. In *TCC*, pp. 465–481, 2008.

[17] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *39th STOC*, pp. 516–525, 2007.

[18] O. Goldreich, H. Karloff, L. Schulman, and L. Trevisan. Lower bounds for linear locally decodable codes and private information retrieval. *Computational Complexity*, 15(3):263–296, 2006.

[19] L. Gross. Logarithmic Sobolev inequalities. *American J. of Mathematics*, 97(4):1061–1083, 1975.

[20] G. H. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 1988.

[21] J. Håstad. Some optimal inapproximability results. *J. of the ACM*, 48(4):798–859, 2001.

[22] I. Haviv and O. Regev. On tensor norms and locally decodable codes, 2008. In progress.

[23] A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973. English translation in *Problems of Information Transmission*, 9:177–183, 1973.

[24] P. Høyer and R. de Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *19th STACS*, volume 2285 of *LNCS*, pp. 299–310, 2002.

[25] J. Jackson. An efficient membership-query algorithm for learning DNF with respect to the uniform distribution. *J. of Computer and System Sciences*, 55(3):414–440, 1997.

[26] R. Jain, H. Klauck, and A. Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *40th STOC*, pp. 599–608, 2008.

[27] S. Jukna. *Extremal Combinatorics*. Springer, 2001.

[28] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proc. of 29th IEEE FOCS*, pp. 68–80, 1988.

[29] G. Kalai and S. Safra. Threshold phenomena and influence. In A. Percus, G. Istrate, and C. Moore, editors, *Computational Complexity and Statistical Physics*, pp. 25–60, 2006.

[30] J. Katz and L. Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *32nd STOC*, pp. 80–86, 2000.

[31] I. Kerenidis and R. de Wolf. Exponential lower bound for 2-query locally decodable codes via a quantum argument. *J. of Computer and System Sciences*, 69(3):395–420, 2004.

[32] H. Klauck. Lower bounds for quantum communication complexity. In *42nd FOCS*, pp. 288–297, 2001.

[33] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *45th FOCS*, pp. 12–21, 2004.

[34] R. König and R. Renner. Sampling of min-entropy relative to quantum knowledge, 28 Dec 2007. quant-ph/0712.4291.

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

[36] J. R. Lee and A. Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. *Geometric and Functional Analysis*, 14(4):745–747, 2004.

[37] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *J. of the ACM*, 40(3):607–620, 1993.

[38] Y. Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *J. of Computer and System Sciences*, 50(3):543–550, 1995.

[39] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 2008. To appear.

[40] E. Mossel, R. O'Donnell, and R. Servedio. Learning functions of k relevant variables. *J. of Computer and System Sciences*, 69(3):421–434, 2004.

[41] A. Nayak. Optimal lower bounds for quantum automata and random access codes. In *40th FOCS*, pp. 369–376, 1999.

[42] A. Nayak and A. Vishwanath. Quantum walk on the line. quant-ph/0010117, Oct 2000.

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

[44] R. O'Donnell. *Computational applications of noise sensitivity*. PhD thesis, MIT, 2003.

[45] R. O'Donnell. Lecture notes for a course "Analysis of Boolean functions", 2007. Available at http://www.cs.cmu.edu/˜odonnell/boolean-analysis/.

[46] R. O'Donnell. Some topics in analysis of boolean functions. ECCC Report TR08–055, 2008.

[47] R. Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5:205–221, 1995.

[48] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, mathematics*, 67(1):159–176, 2003.

[49] A. Samorodnitsky. Personal communication with O. Regev, March 2008.

[50] N. Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of trace classes $S_p(1 \leq p < \infty)$. *Studia Mathematica*, 50:163–182, 1974.

[51] L. Trevisan. Some applications of coding theory in computational complexity. *Quad. di Matematica*, 13:347–424, 2004.

[52] R. de Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 287(1):337–353, 2002.

[53] D. Woodruff. New lower bounds for general locally decodable codes. ECCC Report TR07–006, 2006.

[54] S. Yekhanin. Towards 3-query locally decodable codes of subexponential length. In *39th STOC*, pp. 266–274, 2007.