Electronic Colloquium on Computational Complexity, Report No. 78 (2009)

A Probabilistic Inequality with Applications to Threshold Direct-product Theorems

Falk Unger
UC Berkeley, EECS Department
funger@eecs.berkeley.edu

September 16, 2009

Abstract

We prove a simple concentration inequality, which is an extension of the Chernoff bound and Hoeffding's inequality for binary random variables. Instead of assuming independence of the variables we use a slightly weaker condition, namely bounds on the co-moments.

This inequality allows us to simplify and strengthen several known direct-product theorems and establish new threshold direct-product theorems. Threshold direct-product theorems are statements of the following form: If one instance of a problem can be solved with probability at most p, then solving significantly more than a p-fraction among multiple instances has negligible probability.

Using our concentration inequality we show how to obtain threshold (and standard) direct-product theorems from known XOR Lemmas. We give examples of this approach and establish (threshold) direct-product theorems for quantum XOR games, quantum random access codes, 2-party and multi-party communication complexity, and circuits. Similar results can be obtained for other models of computation, e.g. polynomials over $GF(2)$ and query complexity.

We believe that our inequality has applications in other contexts as well.

1 Introduction

Direct-product theorems (DPT) are useful tools in computer science, with a wide range of applications. They are statements of basically this form: If some process (limited by a certain amount of resources) can compute some function $f : \{-1,+1\}^n \to \{-1,+1\}$ on a randomly chosen input with probability p, then the probability that a process (with comparable resources) can simultaneously compute f on k independently chosen inputs becomes exponentially small in k, ideally at most p^k. Closely related are XOR Lemmas, which state that for independently chosen inputs $x_1, \ldots, x_k$ the advantage of computing $\prod_{i=1}^k f(x_i)$ over guessing becomes exponentially small in k, ideally $(2p-1)^k$.

We consider threshold direct-product theorems, sometimes also called concentration bounds. They are statements saying that even solving significantly more than a p-fraction of k simultaneously given instances correctly has exponentially small probability in k. Statements of this kind are for example useful when one wants to distinguish whether some process succeeds with probability s or c, with $s < c$. Of course, challenging the process sequentially on many inputs and then accepting if more than a $(s+c)/2$-fraction of the challenges are correct gives a way to boost one's confidence. However, sequential repetition might be inefficient, simply impossible in the considered model or

ISSN 1433-8092

undesirable for several other reasons [14]. The question is whether this protocol can also be executed in parallel, such that the process gets all k instances at once and has to output all k answers. In this case, knowledge of the other inputs could give the process a non-trivial advantage over sequential repetition, and numerous examples where this is indeed the case are known [34, 37, 33]. On the other hand, for several models (threshold) direct-product theorems hold under parallel repetition and we give several examples, see Section 1.1 and Section 4.

Technique Our main technical tool will be (a slightly stronger version) of the following Lemma. Relative entropy $D(\cdot||\cdot)$ is defined in equation (5).

Lemma 1 (Threshold Lemma)**.** *Let $Y_1, \dots, Y_k \in \{-1, +1\}$ be random variables and $-1 \leq \beta \leq 1$ such that for all $S \subseteq [k]$*

$$\mathrm{E}[\prod_{i \in S} Y_i] \leq \beta^{|S|}. \tag{1}$$

Let λ be some number such that $\beta \leq \lambda \leq 1$. Then

$$\Pr[\sum_{i=1}^{k} Y_i \geq \lambda k] \leq e^{-kD(1/2+\lambda/2||1/2+\beta/2)}. \tag{2}$$

Note that if we have random variables Y_i which are independent, and $+1$ with probability $1/2 + \beta/2$ and -1 with probability $1/2 - \beta/2$, then the assumptions of the Lemma are satisfied and the bound we get in inequality (2) is the same as the Chernoff bound, see Section 2. However, unlike the Chernoff bound which requires the variables Y_i to be independent, we only require that the (co)-moments of the Y_i are bounded.

Once this lemma is established, our results will follow almost immediately. In our applications we consider processes, for example resource-bounded algorithms, which attempt to solve k instances of a problem with binary output simultaneously. We show that with high probability many of them must fail. For every execution of the k processes we can define *syndrome* variables $Y_i \in \{-1, +1\}$ with $Y_i = 1$ if and only if the i-th process succeeds. To show that this distribution indeed satisfies the assumptions of our Threshold Lemma, we use known appropriate XOR Lemmas, which state that for any $S \subseteq [k]$ the probability that an even number of the processes in S fail is bounded above by $1/2 + \beta^{|S|}/2$ for some constant $0 < \beta < 1$, implying condition (1). Our Lemma then immediately establishes a DPT by choosing $\lambda = 1$ and a threshold DPT when choosing $\lambda \in (\beta, 1)$.

1.1 Applications

Typical examples where threshold DPTs are useful are CAPTCHA puzzles, which are widely used on the internet to verify interaction with a human user. Here a user is challenged to recognize some barely readable word, which humans can do reasonably well (say with probability $c \in (0, 1)$) and computers cannot (say with probability at most $s < c$). That parallel repetition is indeed possible was recently shown [20]. We will give an alternative, much simpler proof of this in Section 4.4 by showing that if a circuit can compute one instance of a challenge with probability at most p, then the probability to compute more than a p-fraction of k instances correctly has probability at most $e^{-\Omega(k)}$.

Another example are proof systems in which a verifier wants to decide whether some input x belongs to some language $\mathrm{L} \in \mathrm{NEXP}$ and interacts with two (computationally unbounded but non-communicating and possibly malicious) provers. It was pointed out in [11], using results from [16], that for a particular kind of proof systems it is possible to set up the challenges to the provers in such a way that if $x \in L$ then with probability $\approx 12/16$ he will accept, and if $x \notin L$ then the provers can trick him into accepting x with probability at most $\approx 11/16$. These particular proof systems are based on XOR games, in which a verifier sends challenges to two provers who reply with bits a resp. b and the verifier's answer is only based on the value of $a \oplus b$. Indeed, [31] recently showed that also in this case a threshold DPT holds, i.e., if one game can be won with probability 11/16 then winning significantly more than an 11/16-fraction of multiple games has small probability. Hence the verifier's error probability can made arbitrarily small, and this implies MIP=NEXP.[1]

In Section 4.1 we will prove a threshold DPT for a similar class of protocols, but for the quantum case, in which the provers may also share entanglement. We show that in order to win a large fraction of games, the essentially best strategy is to play all games independently optimally. This is the first threshold DPT for quantum games. An open question is whether $\mathrm{MIP}_* = \mathrm{NEXP}$, i.e., whether NEXP can be characterized in the same way as above, but when the provers can additionally share entanglement. Although there has been some progress [21, 22], the question is still open. It is conceivable that a possible proof starts with some protocol which accepts $x \in L$ with probability $c < 1$ and $x \notin L$ with probability at most $s < c$ and then amplifies the correctness probability as above, using a (still to be proved) threshold DPT for general quantum games. We make a first step by proving the first threshold DPT for quantum games, although only for a particular class, namely quantum XOR games. Unfortunately, this class will probably not be sufficient to prove $\mathrm{MIP}_* = \mathrm{NEXP}$, unless EXP = NEXP [11], so an extension of this work would be required.

We also show a threshold DPT for quantum random access codes (QRACs). Assume you want to encode a string x of n classical bits into $m < n$ qubits, such that *any* $k < n$ bits of the original n bits can be recovered, with reasonable success probability σ. Then in Theorem 2 of [4] it is shown that this is only possible with success probability that is exponentially small in k (for some reasonable choices of k, n, m and some minor technical assumptions). We extend this and show that it is hard to get some large fraction of the k bits right, see Theorem 7 in Section 4.2.[2] We apply this result to derive a (threshold) direct-product theorem for the communication complexity of the disjointness function in the quantum one-way model. This was already shown in [4], however, our threshold result for QRACs simplifies and strengthens their result.

Furthermore, we show (threshold) DPTs for communication complexity (also for multiparty protocols) in Section 4.3. Although in these cases standard DPTs were previously known, our approach improves and simplifies them and also yields threshold DPTs. Further motivating applications of threshold DPTs can be found in [20, 13].

Further applications There are other XOR Lemmas which can be turned into threshold DPTs: an XOR Lemma for classical query complexity (Theorem 4 in [23]), an XOR Lemma for bounded-degree polynomials over GF(2) (Theorem 1.2 in [37]). Furthermore, Theorem 5 in [28] is a generalization of Yao's XOR Lemma to interactive systems to which our technique can also be applied.

Since our Threshold Lemma is actually a statement about probability distributions, it is likely that it has further applications in other contexts. For example using the bound (9) it is possible to give a simple proof of Vazirani's Parity-Lemma [36], which is used in randomness extraction. This

[1] Note that there are constructions with better parameters [2, 32].

[2] Note that using techniques from [4] one could also obtain a similar but weaker result, which does not give optimal parameters.

Parity-lemma has been extended in [29], and our inequality does not simply imply this stronger result. We are aware of at least one other result which immediately follows from our results, namely the XOR Lemma in [9].

1.2 Relation of XOR Lemmas and DPTs and some related work

The close connection between XOR Lemmas and DPTs is well known, and in particular there are several results which derive DPTs from XOR Lemmas [37, 12, 4, 30]. However, those reductions only work for a particular application or they are not as strong as ours. Furthermore, there are no reductions which establish threshold DPTs, which are our main focus.

The first generic approach to turn XOR Lemmas into direct-product theorems appears in Proposition 1.4 in [37]. However, it does not yield optimal bounds and is slightly more complicated.[3] In particular, starting from the assumption that the XOR on k instances is correct with probability at most $1/2 + \beta^k/2$, their approach can give an upper bound of at most $(1/2 + \beta/2)^{k/2}$ (even slightly worse) for the probability to compute all k instances simultaneously correctly, whereas we can show a quadratic improvement to $(1/2 + \beta/2)^k$.

The first example (we are aware of) of optimally turning XOR Lemmas into direct-product theorems appears in the proof of Theorem 5.1 in [30], but only works for $k = 2$. The first example which works for larger k is in [12], whose work we extend. Their whole analysis is geared only towards quantum XOR games. Their argument was later used in the proof of Theorem 2 in [4]. Our approach can be applied to a much wider class of XOR Lemmas. Furthermore, none of these results can be used to derive threshold direct-product theorems, which are our main application.

There are also connections known in the other direction, i.e., from DPTs to XOR Lemmas. For example [15] show how to use the Goldreich-Levin decoding algorithm to obtain XOR Lemmas from DPTs. However, this approach falls short of giving optimal parameters. In particular, it cannot show that if computing k instances simultaneously correctly has probability at most $(1/2 + \beta/2)^k$, then computing the XOR of k instances correctly has probability at most $1/2 + \beta^k/2$, which would be the expected bound, see also [37]. Furthermore, in certain settings, e.g. quantum XOR games, this approach fails completely. Moreover, [37] contains an example where a direct-product theorem is true, but an XOR Lemma is not.

There is also a quantum version of the Goldreich-Levin Theorem [1] which has slightly better parameters, but suffers essentially the same problems.

In the proof of Theorem 10 in [4] there is also a simple argument for turning direct-product theorems into threshold direct-product theorems. However, there is some loss in the parameters and the thresholds they obtain will generally not be optimal.

The fact that reductions from XOR Lemmas to DPTs are often tight whereas the reverse direction is not, suggests that XOR Lemmas are "better". This statement is further supported by a recent result in [35]. They show that standard techniques for proving XOR Lemmas require that the underlying computational model can compute majority, but there are proofs for DPTs which

[3] To be fair, we should point out that their result is stronger is some sense: they do not need to make the assumption that on all subsets of the bits in y the bias is bounded. In terms of XOR Lemmas this means that when deriving a direct-product theorem for k instances, they do not need to assume that for all $k' \leq k$ an XOR Lemma holds. They only need the assumption that for some sufficiently large $k' = \Omega(k)$ it holds that the XOR on k' instances is correct with probability at most $1/2 + 2^{-O(k')}/2$. However, in order to derive a direct-product theorem for all k, they also need to assume that an XOR Lemma holds for all k'. Furthermore, we are not aware of any XOR Lemmas for which our stronger assumption is not valid and therefore believe that this difference in assumptions is insignificant.

do not require that. So, the fact that XOR Lemmas are often "better" comes at the price that in certain models they are harder to obtain and sometimes simply not true [35, 37]. However, we want to stress that this only applies to certain models, in other models (for example communication complexity [26, 37, 34] and quantum XOR games [12]) XOR Lemmas are easier to obtain.

2 Preliminaries

For a function $f : \{-1,+1\}^k \to \mathbb{R}$ its Fourier transform is a function $\widehat{f} : 2^{[k]} \to \mathbb{R}$ defined as

$$\widehat{f}(S) = \sum_{y \in \{-1,+1\}^k} f(y) \prod_{i \in S} y_i \tag{3}$$

for every $S \subseteq [k]$. Plancherel's identity (also called Parseval's identity) states that for $f, g : \{-1,+1\}^k \to \mathbb{R}$

$$\sum_{y \in \{-1,+1\}^k} f(y)g(y) = \frac{1}{2^k} \sum_{S \subseteq [k]} \widehat{f}(S)\widehat{g}(S). \tag{4}$$

For independently and identically distributed binary random variables $Y_1, \dots, Y_k \in \{-1,+1\}$, where $\forall i : \mathrm{E}[Y_i] = \beta$, the Chernoff bound [8] says that for any $\lambda \geq \beta$

$$\Pr[\sum_i Y_i \geq \lambda k] \leq e^{-kD(1/2+\lambda/2||1/2+\beta/2)},$$

where the binary relative entropy is defined as

$$D(\lambda||p) = \lambda \ln \frac{\lambda}{p} + (1-\lambda) \ln \frac{1-\lambda}{1-p} \tag{5}$$

(and $0 \ln 0 = 0$). Note that $\forall 0 < p < 1, 0 \leq \lambda \leq 1 : D(\lambda||p) \geq 0$ and $D(p||p) = 0$.

More generally, if $Y_1, \dots, Y_k \in \{-1,+1\}$ are independent but not necessarily identically distributed and $\forall i : \mathrm{E}[Y_i] \leq \beta_i$ then Hoeffding's inequality [17] states that

$$\Pr[\sum_{i=1}^{k} Y_i \geq \lambda k] \leq e^{-k\left(\lambda - \sum_i \beta_i/k\right)^2/2}.$$

3 Main technical result

The following Lemma extends our earlier version (Lemma 1) by adding parameters E, C, which make the statement somewhat ugly, but which are crucial to capture XOR Lemmas that do not behave nicely and contain certain error terms, e.g. Yao's XOR Lemma, see Section 4.4. Furthermore, we allow the variables to have different biases and the quantity to be bounded can be a general linear function of the Y_i. Further, we give tight bounds for the probability that $\sum_i Y_i$ obtains its maximal value k.

Lemma 2 (Threshold Lemma)**.** *Let $Y_1, \dots, Y_k \in \{-1,+1\}$ be random variables, $-1 \leq \beta_1, \dots, \beta_k \leq 1$, $c_1, \dots, c_k > 0$ and $E, C > 0$ such that for all $S \subseteq [k]$*

$$\mathrm{E}[\prod_{i \in S} Y_i] \leq C \prod_{i \in S} \beta_i + E. \tag{6}$$

Then for all λ with $\sum_i c_i\beta_i \le \lambda k < \sum_i c_i$ it holds

$$\Pr[\sum_{i=1}^{k} c_iY_i \ge \lambda k] \le Ce^{-\left(\lambda k-\sum_i c_i\beta_i\right)^2/2\sum_i c_i^2} + Ee^{(\lambda k-\sum_i \beta_i)(\sum_i c_i-\lambda k)/\sum_i c_i^2}. \tag{7}$$

Moreover, if $\forall i : \beta_i = \beta, c_i = 1$ and $\beta \le \lambda < 1$ then the following stronger bound holds

$$\Pr[\sum_{i=1}^{k} Y_i \ge \lambda k] \le Ce^{-kD(1/2+\lambda/2||1/2+\beta/2)} + E\left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)}\right)^{k(1/2-\lambda/2)}. \tag{8}$$

For $\lambda = 1$ (but the β_i can be different) we can bound

$$\Pr[\sum_{i=1}^{k} Y_i = k] \ \le C\prod_{i=1}^{k}\frac{1+\beta_i}{2} + E. \tag{9}$$

In most cases, the lemma will be used with $C = 1, \forall_i c_i = 1$ and $E = 0$ (In fact, the lemma is optimized for $E = 0$, and it is possible to get slightly better results for $E > 0$.) In this case, the bounds (7) resp. (8) simplify to

$$\Pr[\sum_{i=1}^{k} Y_i \ge \lambda k] \le e^{-k\left(\lambda-\sum_i \beta_i/k\right)^2/2} \tag{10}$$

$$\Pr[\sum_{i=1}^{k} Y_i \ge \lambda k] \le e^{-kD(1/2+\lambda/2||1/2+\beta/2)} \tag{11}$$

which are the same as the familiar Hoeffding's inequality (restricted to binary random variables) and the Chernoff bound, see Section 2. However, for binary variables our assumptions are weaker, as we do not assume independence of $Y_1, \dots, Y_k$ but only the weaker condition (6).

It is possible to prove similar concentration bounds for variables $Y_i \in [-1, +1]$, with slightly worse bounds.

Inequality (9) is easily seen to be optimal. Although not stated explicitly, this inequality can be distilled from the arguments in Section 3 of [12].

Proof. Set $p(y) = \Pr[Y_1 = y_1, \dots, Y_k = y_k]$. For any parameter $t \ge 0$ it holds

$$\sum_{y, \sum_i c_iy_i \ge \lambda k} p(y) \quad \le \quad e^{-t\lambda k}\sum_y e^{t\sum_i c_iy_i}p(y), \tag{12}$$

because $e^{-t\lambda k+t\sum_i c_iy_i}$ is always at least 0 and $e^{-t\lambda k+t\sum_i c_iy_i} \ge 1$ when $\sum_i c_iy_i \ge \lambda k$. Defining $c(y) = e^{t\sum_i c_iy_i}$ we can compute (using the definition in equation (3))

$$\begin{aligned}
\widehat{c}(S) &= \sum_{y\in\{-1,1\}^k}\prod_{i\in S} y_ie^{tc_iy_i}\prod_{i\notin S}e^{tc_iy_i} = \prod_{i\in S}(e^{tc_i}-e^{-tc_i})\prod_{i\notin S}(e^{tc_i}+e^{-tc_i}) \\
\widehat{p}(S) &= \sum_{y\in\{-1,1\}^k} p(y)\prod_{i\in S} y_i = \mathrm{E}[\prod_{i\in S} Y_i].
\end{aligned}$$

We now bound (12) further by using Plancherel's identity (equation (4)) and then the assumptions of our lemma together with the fact that $\forall S : \widehat{c}(S) \geq 0$:

$$\begin{aligned}
&= 2^{-k} e^{-t\lambda k} \sum_{S \subseteq [k]} \widehat{c}(S) \widehat{p}(S) && (13)\\
&\leq 2^{-k} e^{-t\lambda k} \sum_{S \subseteq [k]} \left(C \prod_{i \in S} (e^{tc_i} - e^{-tc_i}) \beta_i \prod_{i \notin S} (e^{tc_i} + e^{-tc_i}) + E \prod_{i \in S} (e^{tc_i} - e^{-tc_i}) \prod_{i \notin S} (e^{tc_i} + e^{-tc_i}) \right)\\
&= 2^{-k} e^{-t\lambda k} \left(C \prod_{i=1}^{k} ((1+\beta_i) e^{tc_i} + (1-\beta_i) e^{-tc_i}) + E 2^k e^{t \sum_i c_i} \right), && (14)
\end{aligned}$$

where in the last step we use the identity $\prod_{i=1}^{k} (a_i + b_i) = \sum_{S \subseteq [k]} \prod_{i \in S} a_i \prod_{i \notin S} b_i$ once with $a_i = (e^{tc_i} - e^{-tc_i}) \beta_i, b_i = (e^{tc_i} + e^{-tc_i})$ and once with $a_i = (e^{tc_i} - e^{-tc_i}), b_i = (e^{tc_i} + e^{-tc_i})$.

To establish inequality (7) we use Lemma 3 below to bound (14) and then choose $t := (\lambda k - \sum_i c_i \beta_i) / \sum_i c_i^2 \geq 0$.

$$\begin{aligned}
&\leq Ce^{-t\lambda k + t^2 \sum_i c_i^2/2 + t \sum_i c_i \beta_i} + Ee^{t(\sum_i c_i - \lambda k)}\\
&= Ce^{(-\lambda^2 k^2 + \lambda k \sum_i c_i \beta_i + \lambda^2 k^2/2 - \lambda k \sum_i c_i \beta_i + (\sum_i c_i \beta_i)^2/2 + \lambda k \sum_i c_i \beta_i - (\sum_i c_i \beta_i)^2))/\sum_i c_i^2}\\
&\quad + Ee^{(\lambda k - \sum_i \beta_i)(\sum_i c_i - \lambda k)/\sum_i c_i^2}
\end{aligned}$$

Transforming the numerator in the exponent of the first exponential as

$$\begin{aligned}
& -\lambda^2 k^2 + \lambda k \sum_i c_i \beta_i + \lambda^2 k^2/2 - \lambda k \sum_i c_i \beta_i + (\sum_i c_i \beta_i)^2/2 + \lambda k \sum_i c_i \beta_i - (\sum_i c_i \beta_i)^2\\
= & -\lambda^2 k^2/2 + \lambda k \sum_i c_i \beta_i - (\sum_i c_i \beta_i)^2)/2\\
= & (\lambda k - \sum_i c_i \beta_i)^2/2
\end{aligned}$$

we arrive at result (7).

To establish inequality (8) we set $c_i = 1$ and choose $t := \frac{1}{2} \ln \frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \geq 0$, which means $e^t = \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \right)^{1/2}$. Then (14) becomes

$$\begin{aligned}
= & \frac{C}{2^k} \left((1+\beta) \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \right)^{1/2-\lambda/2} + (1-\beta) \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \right)^{-1/2-\lambda/2} \right)^k\\
& + E \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \right)^{k(1/2-\lambda/2)}
\end{aligned}$$

Transforming the term in the first big bracket

$$\begin{aligned}
& (1+\beta) \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \right)^{1/2-\lambda/2} + (1-\beta) \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)} \right)^{-1/2-\lambda/2}\\
= & \frac{(1-\beta)^{1/2-\lambda/2} (1+\beta)^{1/2+\lambda/2}}{(1-\lambda)^{1/2-\lambda/2} (1+\lambda)^{1/2+\lambda/2}} ((1+\lambda) + (1-\lambda))\\
= & 2e^{-D(1/2+\lambda/2 || 1/2+\beta/2)}
\end{aligned}$$

gives claim (8) as desired.

Inequality (9) follows from (14) with $\lambda = 1$, $\forall i : c_i = 1$ and $t \longrightarrow \infty$ or the following simple derivation. Define the indicator function $I_k : \{-1,+1\}^k \to \{0,1\}$ as $I_k(y) = 1 \leftrightarrow y = (+1,\ldots,+1)$. From the definition in equation (3) we can quickly compute that $\forall S : \widehat{I}_k(S) = 1$. Then, using the same arguments as before we can bound

$$\begin{aligned}\Pr[\sum_{i=1}^{k} Y_i = k] &= \sum_{y\in\{-1,1\}^k} I_k(y)p(y) = \frac{1}{2^k}\sum_S \widehat{I}_k(S)\widehat{p}(S) \\ &\leq \frac{1}{2^k}\sum_S\left(C\prod_{i\in S}\beta_i + E\right) = C\prod_{i=1}^{k}\frac{1+\beta_i}{2} + E.\end{aligned}$$

□

Lemma 3. *For all $\beta, t \in \mathbb{R}$ it holds $(1+\beta)e^t + (1-\beta)e^{-t} \leq 2e^{t^2/2+\beta t}$.*

Proof. Fix any β. We will show that $f(t) := 2e^{t^2/2+\beta t} - (1+\beta)e^t - (1-\beta)e^{-t} \geq 0$ for all t. Note that f is analytic. We will use this repeatedly without further mention. Compute $f'(t) = \frac{\partial f}{\partial t} = (2t+2\beta)e^{t^2/2+\beta t} - (1+\beta)e^t + (1-\beta)e^{-t}$ and $f''(t) = \frac{\partial^2 f}{\partial t^2} = (2+2(\beta+t)^2)e^{t^2/2+\beta t} - (1+\beta)e^t - (1-\beta)e^{-t}$. It holds $f(0) = 0$, $f'(0) = 0$ and $f''(0) = 2 + 2\beta^2 - 1 \geq 1$. Hence, $\exists \epsilon > 0 \forall t \in [-\epsilon,\epsilon]\backslash\{0\} : f(t) > 0$.

We now apply a "bootstrapping" argument. Assume there was a $t > \epsilon$ for which $f(t) < 0$. Then there must be a smallest $\widehat{t} > \epsilon$ with $f(\widehat{t}) = 0$ and hence $\forall 0 < t < \widehat{t} : f(t) > 0$. Together with the fact that $\forall t : f''(t) - f(t) = 2(\beta+t)^2e^{t^2/2+\beta t} \geq 0$, this implies that f is strictly convex in $[0,\widehat{t}]$. But then $f(0) = 0$ and $f'(0) = 0$ imply $f(\widehat{t}) > 0$, giving a contradiction.

The argument for $t < -\epsilon$ is analogous. □

3.1 Extensions

Our Threshold Lemma can be extended in several ways. These extensions are not relevant for our later applications but might be interesting for other applications. First we notice that it is possible to weaken condition (6) by instead demanding that for all $l \in [k]$

$$\mathrm{E}[\sum_{S,|S|=l}\prod_{i\in S} Y_i] \leq C\sum_{S,|S|=l}\prod_{i\in S}\beta_i + E. \tag{15}$$

This is because the only time we use the bound on $\mathrm{E}[\prod_{i\in S} Y_i]$ in the proof of the lemma is in equation (13) and since $\widehat{c}(S)$ is constant for all S of the same cardinality, equation (13) also holds with the above weaker condition.

Furthermore, it is also not necessary that condition (6) holds strictly for all $S \subseteq [k]$. In particular, from equation (13) we see that if the bounds on $\mathrm{E}[\prod_{i\in S} Y_i]$ are only slightly worse, then the final bound will also be only slightly worse. It is also noteworthy that since $\widehat{c}(S)$ becomes smaller when $|S|$ increases, the bounds on $\mathrm{E}[\prod_{i\in S} Y_i]$ for large $|S|$ are less crucial.[4]

[4] To give a concrete example, consider a distribution on $k = 2n$ bits $Y_1,\ldots,Y_k$ which are a random permutation of n times $+1$ and n times -1. Clearly, $\sum_i Y_i = 0$, hence $\Pr[\sum_i Y_i > 0] = 0$. However, our lemma in its current form cannot even show that $\Pr[\sum_i Y_i > 0] \leq 2^{-\Omega(k)}$, because although for all $S \subset [k]$ it holds that $\mathrm{E}[\prod_{i\in S} Y_i] \leq 0$, we also have $\mathrm{E}[\prod_{i\in[k]} Y_i] = 1$, so the conditions of the lemma are formally not satisfied. However, since $2^{-k}e^{-t\lambda k}\widehat{c}([k])\widehat{p}([k]) \leq 2^{-k}$, one sees by looking at the derivation from line (13) to (14) that the term in (14) plus 2^{-k} is still an upper bound on (12). And following the rest of the proof one can show that $\Pr[\sum_i Y_i \geq \lambda k] \leq e^{-kD(1/2+\lambda/2||1/2)} + 2^{-k} \leq 2^{-\Omega(k)}$.

It is also not necessary to assume that $\forall i : Y_i \in \{-1, +1\}$ but it is enough to assume that $Y_i \in [-1, 1]$, since a distribution which maximizes the right-hand side of equation (12) and satisfies (6) must have $\forall i : Y_i \in \{-1, +1\}$, by convexity of e^x.

4 Applications

4.1 Quantum XOR games

We start by recalling the relevant definitions from [12].

Definition 1. *An* XOR game $G = (f, \pi)$ *is given by two sets* S, T*, a predicate* $f : S \times T \to \{-1, +1\}$ *and a probability distribution* π *on* $S \times T$*. The game is played between two provers (Alice and Bob), who cannot communicate with each other once the game has started, and a verifier. The verifier selects a pair of questions* $(s, t) \in S \times T$ *according to distribution* π *and send* s *to Alice and* t *to Bob. Alice (Bob) sends back a bit* $a \in \{-1, +1\}$ *(*$b \in \{-1, +1\}$*) to the verifier, who accepts if and only if* $a \cdot b = f(s, t)$*. We denote by* $\omega_c(G)$ *the maximum success probability (over the provers's strategy) and define the bias as* $\epsilon_c(G) := 2\omega_c(G) - 1$*. In the quantum case, the provers may additionally share an arbitrary quantum state, and we denote by* $\omega_q(G)$ *the maximum success probability (over the provers's strategy and their shared quantum state) and similarly,* $\epsilon_q(G) := 2\omega_q(G) - 1$

Our definition is for *non-degenerate* quantum XOR games and all quoted results are for non-degenerate quantum XOR games.[5] We want to analyze how these games behave under parallel repetition.

Definition 2. *For* k *XOR games* $G_1 = (f_1, \pi_1), \ldots, G_k = (f_k, \pi_k)$ *and* $g : \{-1, +1\}^k \to \{-1, +1\}$ *define the* g*-composition, denoted by* $g(G_1, \ldots, G_k)$*, as follows. The verifier chooses questions* $((s_1, t_1), \ldots, (s_k, t_k)) \in (S_1 \times T_1) \times \cdots \times (S_1 \times T_1)$ *according to the product distribution* $\pi_1 \times \cdots \times \pi_k$*, and sends* $(s_1, \ldots, s_k)$ *to Alice and* $(t_1, \ldots, t_k)$ *to Bob. Alice and Bob output bits* $a_1, \ldots, a_k \in \{-1, +1\}$ *and* $b_1, \ldots, b_k \in \{-1, +1\}$*, respectively. They win if* $g(a_1 b_1 f(s_1, t_1), \ldots, a_k b_k f(s_k, t_k)) = 1$ *and we denote the maximal winning probability by* $\omega_{c/q}(g(G_1, \ldots, G_k))$ *and analogously,* $\epsilon_{c/q} = 2\omega_{c/q}(g(G_1, \ldots, G_k)) - 1$*.*

A simple way for Alice and Bob to play $g(G)$ is to independently play each game individually optimally. For which g is this optimal? Theorem 1 from [12] implies that this is the case if g is the k-bit XOR function $XOR(y_1, \ldots, y_k) = \prod_i y_i$, i.e., they win if they loose an even number of individual games.

Theorem 4. *[Theorem 1 in [12]] For any XOR games* $G_1, \ldots, G_k$*,* $\epsilon_q(XOR(G_1, \ldots, G_k)) = \prod_i \epsilon_q(G_i)$*.*

Using the results from Section 3, we can turn this into a Chernoff-type direct-product Theorem. For $0 \leq \lambda \leq 1$ define the function $\theta[\lambda] : \{-1, +1\}^k \to \{-1, +1\}$ as

$$\theta[\lambda](y) = \begin{cases} +1 & \text{if } \sum_i y_i \geq \lambda k \\ -1 & \text{otherwise} \end{cases} ,$$

[5] In degenerate XOR games, the verifier is allowed to accept or reject independently of the value $a \cdot b$, whereas in non-degenerate XOR games the verifier has to accept exactly one value $a \cdot b$ for any pair of questions (s, t). For degenerate XOR games, the results stated in this section are simply not true, see Section 3.2.2 in [37].

Theorem 5. *Let $G_1, \dots, G_k$ be quantum XOR games, and let $\sum_i \epsilon_q(G_i)/k \leq \lambda \leq 1$. Then,*

$$\omega_q(\theta[\lambda](G_1, \dots, G_k)) \leq e^{-k\left(\lambda - \sum_i \beta_i/k\right)^2/2}.$$

In other words, it is highly unlikely to win significantly more than a $\omega_q(G)$-fraction of k XOR games.

Proof. Let $\theta[\lambda](G)$ be a composition of k XOR games and fix any strategy of Alice and Bob to play this game. For $i = 1, \dots, k$ define random variables $Y_i = a_i b_i f(s_i, t_i)$. By Theorem 4 for every $S \subseteq [k]$: $\mathrm{E}[\prod_{i \in S} Y_i] \leq \prod_{i \in S} \epsilon_q(G_i)$. Applying our Threshold Lemma (with parameters $c_i = 1$, $C = 1$ and $E = 0$ as in inequality (10)) yields the result. □

Note that if all G_i are the same inequality (11) gives slightly stronger results and for the extreme case $\lambda = 1$ the bound in (9) gives optimal bounds. This last result is the content of Theorem 2 in [12].

4.2 Quantum random access codes

In this section we extend results from [4] about quantum random access codes (QRAC). Roughly speaking, k-out-of-n-QRACs are mappings from n classical bits x to m qubits $|\psi_x\rangle$, such that it is possible to recover arbitrary k of the original n bits of x by an appropriate measurement on $|\psi_x\rangle$. They [4] show lower bounds on the size m of k-out-of-n-QRACs: If m is significantly smaller than n, then the success probability is exponentially small in k. We can extend this result by showing a similar lower bound on the size of (l/k)-out-of-n-QRACs, (pronounced as "l-out-of-k-out-of-n-QRAC"). These are quantum random access codes which are encodings of n-bit strings such that at least l out of k positions can be recovered correctly. See Definition 4.

In [4] the results about QRAC's are used to prove direct-product theorems for disjointness in the quantum one-way communication model. It turns out that using our bounds on (l/k)-out-of-n-QRACs their proofs can be simplified and strengthened, as we will show in the full version.

The following definition for XOR-QRACs is from [4].

Definition 3. *An XOR quantum random access code with parameters (k, n, m, ϵ) is a map $E : \{-1, +1\}^n \to \mathbb{C}^{2^m \times 2^m}$ from n classical bits to m-qubit quantum states and a decoding algorithm D, which takes m-qubit states as inputs and outputs one classical bit in $\{-1, +1\}$, with the property that*

$$\Pr_{\substack{S \sim \binom{[n]}{k} \\ x \in \{-1,+1\}^n}} [D(E(x), S) = \prod_{i \in S} x_i] \geq 1/2 + \epsilon/2,$$

i.e., the probability to correctly output the XOR of k of the n bits is at least $1/2 + \epsilon/2$.

Note that usually one demands that the decoding succeeds *for all x*. However, if one is interested in lower bounds, this weaker notion of a QRAC, where the probability is taken over random $x \in \{-1, +1\}^n$, is sufficient.

Theorem 6 (Theorem 7 from [4])**.** *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 it holds that*

$$\epsilon \leq C_\eta \left(\frac{\eta m}{n}\right)^{k/2}.$$

To give a feel of appropriate parameters, it turns out that for example choosing constants $\eta = 4$, $C_\eta = 1$, the theorem holds for $n/k > 2$. From this one can get a bound for quantum random access codes.

Definition 4. *An (l/k)-out-of-n quantum random access code (QRAC) on m qubits with success probability p is a map $E : \{-1,+1\}^n \to \mathbb{C}^{2^m \times 2^m}$ from classical bit strings $x \in \{0,1\}^n$ to m-qubit quantum states $|\psi_x\rangle$ and a decoding algorithm D, which given $|\psi_x\rangle$ reconstructs l out of k bits of x correctly, i.e.*

$$\Pr_{\substack{S\sim\binom{[n]}{k} \\ x\in\{-1,+1\}^n}} [\sum_{i=1}^{k} D(E(x),S)_i x_{S(i)} \geq 2l-k] \geq p. \tag{16}$$

Here $S(i)$ is the i-th largest element of S. The probability is over x, the choice of the k-subset $S \subseteq [n]$ and the randomness in the encoding and decoding algorithms.

If $l = k$ then our definition coincides with Definition 1 in [4]. Note that for $l < k$ we do not demand that the decoding algorithm also “knows” on which positions the output is correct.

Theorem 7. *For any $\eta > 2\ln 2$ there is a constant C_η such that if n/k is large enough and $l \geq (1/2 + \sqrt{\eta m/n}/2)k$ it holds that the success probability for any (l/k)-out-of-n QRAC on m qubits is at most*

$$p \leq C_\eta e^{-kD(l/k||1/2+\sqrt{\eta m/n}/2)}.$$

The case $l = k$ matches Theorem 2 in [4]. Our result extends this by showing that if an encoding of strings $x \in \{0,1\}^n$ does not use $\Omega(n)$ qubits, then it is also hard to correctly decode a large fraction of bits of x.

Proof. Fix any D, E. Define $Y_1, \ldots, Y_k \in \{-1,+1\}$ by

$$Y_i = D(E(x),S)_i \cdot x_{S(i)},$$

where $S(i)$ is the i-th largest element of S. Note that the Y_i depend on the (random) choice of the k-subset $S \subseteq [n]$. Theorem 6 implies that for any $T \subseteq [k]$ it holds: $\Pr[\prod_{i\in T} Y_i = 1] \leq 1/2 + C_\eta \beta^{|T|}/2$, with $\beta := \sqrt{\eta m/n}$. In other words, $\mathrm{E}[\prod_{i\in T} Y_i] \leq C_\eta \beta^{|T|}$. Plugging this into our Threshold Lemma (or equation (11)) yields $\Pr[\sum_{i=1}^{k} Y_i \geq 2l-k] \leq e^{-kD(l/k||1/2+\sqrt{\eta m/n}/2)}$ and noticing that $\Pr[\sum_{i=1}^{k} Y_i \geq 2l-k]$ is equal to the lhs of (16) yields the result. □

4.2.1 Application: Quantum one-way communication complexity of disjointness

In this section we sketch an application of the previous results to quantum *one-way* communication complexity of the disjointness function. (In section 4.3 we will present applications to *two-way* communication complexity.) We start by defining the task. See [24] for an introduction to communication complexity. In this section we will use the $\{0,1\}$-basis for bits.

In the n-bit Disjointness problem Alice gets a string $x \in \{0,1\}^n$ and Bob $y \in \{0,1\}^n$. The task is (for Bob) to output $DISJ_n(x,y)$, which is 0 if $\exists i : x_i = y_i = 1$, and 1 otherwise. In the k-fold disjointness problem $DISJ^{(k)}$, Alice and Bob each receive k strings $x^1, \ldots, x^k \in \{0,1\}^n$ resp. $y^1, \ldots, y^k\{0,1\}^n$ and the task is (for Bob) to output $DISJ_n(x^1,y^1), \ldots, DISJ_n(x^k,y^k)$. We say that $DISJ_n^{(k)}$ can be solved with c qubits and probability p in the quantum one-way model if there is a protocol in which Alice sends c qubits to Bob, who outputs $DISJ_n(x^1, y^1), \ldots, DISJ_n(x^k, y^k)$ with probability at least p for every choice of $x^1, \ldots, x^k, y^1, \ldots, y^k \in \{0,1\}^n$. In the (l/k) threshold version we demand that Bob outputs the correct result on at least l of the k instances. It is known [5] that the quantum one-way communication complexity of $DISJ_n(x, y)$ is $\Theta(n)$ (for any $p > 1/2$).

We want to show that there is some constant $\alpha > 0$ such that any protocol which uses fewer than αkn qubits of communication has success probability exponentially small in k to compute all of $DISJ_n(x^1, y^1), \ldots, DISJ_n(x^k, y^k)$ correctly. Furthermore, we will show that no protocol can compute more than a large (but constant fraction) of the output bits $DISJ(x^1, y^1), \ldots, DISJ(x^k, y^k)$ correctly, except with probability $2^{-\Omega(k)}$. As mentioned earlier, this result is already present in [4], but using our new result about (l/k)-out-of-n quantum random access codes our proof becomes simpler and gives stronger bounds. The key is the following lemma.

Lemma 8. *Assume there is a quantum one-way protocol $\mathcal{P}$ for $DISJ_n^{(k)}$ that communicates at most m qubits and with probability at least σ there are λk instances on which the protocol is correct. Let $0 < \tau \leq 1$ be any number. Then for $k' = \tau k$ there is also a $(\lambda(1-\tau)k'/k')$-out-of-kn QRAC of m qubits with success probability at least $p = \tau\lambda\sigma/2$.*

The proof goes along similar lines as Lemma 8 in [4].

Proof. From the protocol $\mathcal{P}$ we construct a QRAC with the desired properties. Let $x = x^1 \ldots x^k \in \{0,1\}^{nk}$ be a string with k blocks, each of length n. The encoding of the QRAC is just the message $|\psi_x\rangle$ Alice would send in the protocol $\mathcal{P}$ on input $x = x^1, \ldots, x^k$. Let $S = \{i_1, \ldots, i_{k'}\} \sim \binom{[nk]}{k'}$. We now describe a decoding procedure, which can extract at least $\lambda(1-\tau)k'$ of the bits $x_{|S} = x_{i_1}, \ldots, x_{i_{k'}}$ from $|\psi_x\rangle$ with probability at least $\Omega(\sigma)$.

For each $j \in [k']$ set $b_j = \lceil \frac{i_j}{k} \rceil$, i.e., i_j points into block b_j of x. Then, going from $j = 1$ to $j = k$ do the following: If there is no $j' < j$ with $b_j = b_{j'}$ then set $y^{b_j} = e^{rem_n(i_j)}$, where $rem_n(i)$ is the remainder of i divided by n and $e^i \in \{0,1\}^n$ is the i-th standard basis vector.[6] We say that block b_j has been "hit" and that j is "good". Otherwise, if there is a $j' < j$ with $b_j = b_{j'}$ (i.e., block b_j has already been hit previously), then call j "bad". For all blocks b which are not hit by this procedure set y^b to something arbitrary, say $y^b = 0^n$. Note that this procedure sets each string $y_1, \ldots, y_k$ to some n-bit string.

The decoding procedure now outputs a guess $\tilde{x}_j$ for the bit x_{i_j} as follows: If j was bad, then output a random bit $\tilde{x}_j$. If j was good, then output Bob's guess for $DISJ_n(x^{b_j}, y^{b_j})$ in $\mathcal{P}$.

Note, that if $\mathcal{P}$ succeeds on $DISJ_n(x^{b_j}, y^{b_j})$ and j is good then $\tilde{x}_j = x_{i_j}$, as wanted. Otherwise, the output is correct with probability 1/2.

First note that given that $\mathcal{P}$ succeeds on a λ-fraction of the indices, we have for each j: $\Pr[\tilde{x}_j = x_{i_j}] = \Pr[j \text{ is good and points into a block on which } \mathcal{P} \text{ succeeds}] + \Pr[j \text{ is bad}]/2 \geq (1-\tau)\lambda + \tau/2 \geq (1-\tau/2)\lambda$. Hence, $\mathrm{E}[|\{j : \tilde{x}_j = x_{i_j}\}|] \geq k'(1-\tau/2)\lambda$. Using the fact that $|\{j : \tilde{x}_j = x_{i_j}\}| \leq k'$ and setting $s = \Pr[|\{j : \tilde{x}_j = x_{i_j}\}| \geq (1-\tau)\lambda k']$ we can use an argument similar to Markov's inequality to get $s \cdot 1 + (1-s)(1-\tau)\lambda \geq (1-\tau/2)\lambda$, which implies $s(1-(1-\tau)\lambda) + (1-\tau)\lambda \geq (1-\tau/2)\lambda$ and then $s \geq \frac{\tau\lambda}{2(1-(1-\tau)\lambda)} \geq \tau\lambda/2$. □

As mentioned above, this proof closely follows the proof of Lemma 8 in [4]. However, comparing our proof with theirs we note that our reduction still works if some of the j are not "good" and we only need to make the weaker assumption that a λ-fraction of the k instances of $DISJ_n$ succeed,

[6] The entries of e^i are all zero, apart from the i-th, which is 1.

instead of all k of them. This makes the argument slightly simpler and will also yield stronger bounds in the next Theorem.

Note further that the construction actually works for all $x = x^1 \dots x^k \in \{0,1\}^{nk}$ (but uniformly chosen S), which is a stronger statement then what we need for our definition of QRAC. Also, the reduction is of course not tight, but sufficient for our purposes.

This Lemma and Theorem7 immediately imply standard and threshold DPTs for the quantum one-way communication complexity of the disjointness function.

Theorem 9. *Fix any $0 \leq \alpha < 1$. Let $\eta > 2\ln 2$. Then for large enough n, any k and any quantum one-way protocol with at most $m \leq \alpha kn$ qubits it holds*

1. *The probability to correctly compute $DISJ_n^{(k)}$ is $\sigma = 2^{-\Omega(k)}$.*
2. *For any $1/2 + \sqrt{\alpha\eta}/2 < \lambda \leq 1$ the success probability to compute $DISJ_n^{(k)}$ correctly on at least λk instances has success probability $\sigma = 2^{-\Omega(k)}$.*

Proof. Clearly, statement 2 implies 1. Assume statement 2 was not true. Choose $\tau > 0$ such that $\lambda(1-\tau) > 1/2 + \sqrt{\alpha\eta}/2$. Lemma 8 guarantees the existence of a $(\lambda(1-\tau)\tau k/\tau k)$-out-of-$kn$-QRAC with αkn qubits with success probability $\tau\lambda\sigma/2$. However, Theorem 7 states that such a QRAC can have success probability at most $C_\eta e^{-\tau k D(\lambda(1-\tau)||1/2+\sqrt{\alpha\eta}/2)} = 2^{-\Omega(k)}$. □

This reproves Theorems 9 and 10 in [4]. In particular, we can get slightly better bounds by optimizing τ from the proof, i.e. $\max_\tau \tau D(\lambda(1-\tau)||1/2 + \sqrt{\alpha\eta}/2)$ s.t. $0 < \tau$ and $\lambda(1-\tau) > 1/2 + \sqrt{\alpha\eta}/2$ (or by using a better argument in the proof of Lemma 8).

4.3 Communication complexity

We will now sketch several notions from communication complexity. More extensive explanations are provided in [24] and for our applications in particular also in [26].

Two-party communication complexity In the standard model of (distributional) communication complexity two parties, Alice and Bob, want to compute the value $f(x,y)$ of some function $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$, but initially $x \in \{0,1\}^n$ is only known to Alice and $y \in \{0,1\}^n$ to Bob. The question is how many bits do they need to communicate in order for Alice to learn $f(x,y)$ with probability $1-\epsilon$, when each input pair x,y is chosen with probability $P_{x,y}$. Following [24] we define $D_\epsilon^P(f)$ as the minimum number of bits any deterministic protocol needs to communicate in order to compute $f(x,y)$ correctly with success probability at least $1-\epsilon$. We can further define $R_\epsilon(f) = \max_P D_\epsilon^P(f)$. By Yao's principle this is the minimum number of bits any randomized protocol needs in order to correctly output $f(x,y)$ with probability at least $1-\epsilon$ for every x,y.

Discrepancy method Lower bounds on $D_\epsilon^P(f)$ are often obtained via the discrepancy method, which we quickly sketch now. Given f and a probability distribution $p_{x,y}$ define the communication matrix $M_f \in \mathbb{R}^{2^n \times 2^n}$ as

$$(M_f)_{x,y} := f(x,y)$$

For matrices $A, B \in \mathbb{R}^{\{0,1\}^n \times \{0,1\}^n}$ we can define their point-wise product as $(A \circ B)_{x,y} = A_{x,y}B_{x,y}$ and regarding P as a matrix we can define the discrepancy of M_f with respect to P as

$$\mathrm{disc}_P(M_f) = \max_{x,y\in\{0,1\}^n} |x^T(M_f \circ P)y|$$

and general discrepancy as

$$\mathrm{disc}(M_f) = \min_P \mathrm{disc}_P(M_f).$$

discrepancy and distributional complexity are related via the following fact [24]:

$$D^P_\epsilon(f) \geq \log_2 \frac{1-2\epsilon}{\mathrm{disc}_P(M_f)}, \tag{17}$$

which can be written as

$$D^P_{1/2-\beta/2}(f) \geq \log_2 \frac{\beta}{\mathrm{disc}_P(M_f)} \tag{18}$$

by setting $\beta := 1-2\epsilon$. Similarly

$$R_{1/2-\beta/2}(f) \geq \log_2 \frac{\beta}{\mathrm{disc}(M_f)}, \tag{19}$$

Note that $M^{f_1 \cdot f_2} = M^{f_1} \otimes M^{f_2}$, and hence the following theorem (extending earlier work in [34]) gives XOR-lemmas for randomized communication complexity.

Theorem 10 (Theorems 19 and 20 in [26])**.** *For communication matrices A, B and corresponding probability distributions P, Q as above it holds*

$$\begin{aligned} \mathrm{disc}_P(A)\mathrm{disc}_Q(B) &\leq \mathrm{disc}_{P\otimes Q}(A \otimes B) \leq 64\mathrm{disc}_P(A)\mathrm{disc}_Q(B). & (20)\\ \mathrm{disc}(A)\mathrm{disc}(B) &\leq \mathrm{disc}(A \otimes B) \leq 64\mathrm{disc}(A)\mathrm{disc}(B). & (21) \end{aligned}$$

DPT for 2-party communication complexity The following theorem can be seen as a (threshold) DPT for discrepancy.

Theorem 11. *Let $f_1, \ldots, f_k : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ be functions. Assume that $\forall i : \beta_i := 64\mathrm{disc}(M^{f_i}) < 1$. Choose any $\sum_i \beta_i/k \leq \lambda < 1$. Then there is a distribution $P_{x_1,\ldots,x_k,y_1,\ldots,y_k}$ on $(\{0,1\}^n \times \{0,1\}^n)^k$ of inputs for Alice and Bob such that for any c-bit protocol $\mathcal{P}$ (trying to compute $f_1(x_1,y_1),\ldots,f_k(x_k,y_k)$ simultaneously) it holds*

$$\Pr[\mathcal{P}(x_1,y_1,\ldots,x_k,y_k) = f_1(x_1,y_1),\ldots,f_k(x_k,y_k)] \leq 2^{c-6}\prod_{i=1}^k (1/2+\beta_i/2) \tag{22}$$

$$\Pr[\sum_{i=1}^k \mathcal{P}(x_1,y_1,\ldots,x_k,y_k)_i f_i(x_i,y_i) \geq \lambda k] \leq 2^{c-6} e^{-k\left(\lambda - \sum_i \beta_i/k\right)^2/2}, \tag{23}$$

where the probability is over inputs $x_1, y_1, \ldots, x_k, y_k \sim P$ and the randomness of the protocol $\mathcal{P}$.

Proof. As usual, define random variables $Y_i = \mathcal{P}(x_1,y_1,\ldots,x_k,y_k)_i f(x_i,y_i)$ and for $S \subseteq [k]$ define $\beta_S = \mathrm{E}[\prod_{i\in S} Y_i]$. Since $\mathcal{P}$ is a c-cit protocol we have $R^{1/2-\beta_S/2}(\prod_{i\in S} f_i) \leq c$ and by inequality (19) and inequality (21) in Theorem 10 also $\log 64\beta_S \prod_{i\in S} \frac{1}{64\mathrm{disc}(M^{f_i})} \leq R^{1/2-\beta_S/2}(\prod_{i\in S} f_i)$. Combining these last two inequalities yields

$$\beta_S \leq 2^{c-6}\prod_{i\in S} 64\mathrm{disc}(M^{f_i}) = 2^{c-6}\prod_{i\in S}\beta_i$$

From this we see that by using parameters $C = 2^{c-6}$, $E = 0$ and $\forall_i c_i = 1$ in our Threshold Lemma inequality (9) yields (22) and inequality (7) yields (23). □

If all f_i are the same it is possible to prove slightly stronger bounds. Furthermore, we stated our results for $R^{\epsilon}(f)$. The same proof gives analogous results for $D^{P}_{\epsilon}(f)$ by using (18) instead of (19) and inequality (20) instead of (21). Our results are probably not tight, since in the proof we allow each individual instance to use all c bits.

NOF-model An extension of the 2-party model is the so called number-on-the-forehead model (NOF-model), introduced by Chandra, Furst, and Lipton [6]. In this model q parties want to compute a function $f : \{0,1\}^{n\times q} \to \{-1,+1\}$ which depends on q n-bit strings $x^1,\dots,x^q \in \{0,1\}^n$ and each party is given one of the q input strings. The i-th party sees all inputs, apart from its own (which is "written on its forehead"). The question is how many bits c of communication are needed such that in the end one party (say the first) can output the correct result $f(x^1,\dots,x^q)$ with some probability p, when inputs $x^1,\dots,x^q \in \{0,1\}^n$ are chosen from some worst-case distribution μ.[7] The communication is via a broadcast, i.e., every party sees every bit communicated. Note that by Yao's principle we may assume that the q parties employ a deterministic protocol.

This generalization might look contrived at first, but lower bounds in this model also imply lower bounds in various other models of computation, such as ACC circuits, multi-tape Turing machines, branching programs and Lovasz-Schrijver proof systems. Some lower bounds are known in this model if $q \le \log_2 n$, see [25, 37, 24] and references therein. In particular, in a recent break-through result [25, 7] it was shown that the disjointness function (i.e., the first party has to announce whether $x^1 \wedge \dots \wedge x^q = 0^n$ or not and $\wedge$ is the bit-wise AND) requires $\Omega\left(n^{1/(q+1)}2^{-2^{q-1}}\right)$ bits of communication and [3] show that the inner-product function (i.e., the first party has to announce whether $x^1 \wedge \dots \wedge x^q$ contains an even or odd number of ones) needs communication at least $\Omega(n/2^{2q})$.

The following theorem for the NOF-model extends Corollary 1.7 in [37] to a threshold direct-product theorem.

Theorem 12. *Let $f : \{0,1\}^{n\times q} \to \{-1,+1\}$ be a function such that no q-party protocol which also uses q bits of communication can compute f better than with probability $1/2+\epsilon/2$, for $0 \le \epsilon \le 1$, when the inputs are chosen uniformly at random. Choose λ such that $\epsilon^{1/2^q} \le \lambda \le 1$. Then for any q-party protocol $\mathcal{P}$ with uniformly random inputs $x^{1,1},\dots,x^{q,k} \in \{0,1\}^n$ (the i-th player gets inputs $x^{i,1},\dots,x^{i,k}$) and k output bits, which uses c bits of communication, it holds: The probability that the output of $\mathcal{P}$ agrees with $f(x^{1,1},\dots,x^{q,1})\dots f(x^{1,k},\dots,x^{q,k})$ on at least $(1/2+\lambda/2)k$ positions is at most*

$$\Pr[\sum_{i=1}^{k} \mathcal{P}(x^{1,1},\dots,x^{q,k})_i f(x^{1,i},\dots,x^{q,i}) \ge \lambda k] \le 2^c e^{-kD(1/2+\lambda/2||1/2+\epsilon^{1/2^q})/2}. \tag{24}$$

Proof. Let $\mathcal{P}$ be any protocol. Define random variables $Y_i = \mathcal{P}(x^{1,1},\dots,x^{q,k})_i f(x^{1,i},\dots,x^{q,i})$. By Theorem 1.3 in [37] (see also [10]) it follows that for any set $S \subseteq [k]$: $\mathrm{E}[\prod_{i\in S} Y_i] \le 2^c \epsilon^{|S|/2^q}$. Hence, the theorem follows with our threshold lemma, with the parameters $\beta = \epsilon^{1/2^q}$ and $C = 2^c$. □

For $\lambda = 1$ the bound in (24) simplifies to $2^c \epsilon^{k/2^q}$, which is ≤ 1 if $\epsilon \le 2^{-c2^q}$.

[7] In the deterministic setting –which we will not consider– we ask for the amount of communication necessary such that for any input the first party always outputs the correct result.

4.4 Hardness amplification

In this section we will show a threshold DPT for circuits. Assume that for some boolean function $f : \{-1,+1\}^n \to \{-1,+1\}$ it holds that no circuit of size s can compute f better than with probability p (for randomly chosen inputs). We want to show that then also no circuit (of size comparable to s) can compute the k-fold composition of $f^{(k)}(x_1,\ldots,x_k) = f(x_1),\ldots,f(x_k)$ better than with probability $e^{-\Omega(k)}$. Even more, no such circuit can get significantly more than pk of the instances correct, except with probability $e^{-\Omega(k)}$. This result was recently obtained in [20]. We give a simpler proof. As pointed out in the introduction, a similar result also follows from the hardcore Lemma [19], using the stronger recent version in [18], so our result is not at all new. We have included this statement because our proof is very simple and only uses Yao's XOR Lemma, which is simpler and historically earlier than the hardcore Lemma, and will also work for classes where XOR Lemmas are known, but no hardcore Lemma, for example polynomials over GF(2), see Theorem 1.2 in [37] and the remark at the end of Section 1.1. Furthermore, in this example the error constant E of our main Lemma is essential.

We will start with Levin's version [27] of Yao's XOR Lemma [38], as stated in [15].

Lemma 13. *Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be a function and let $0 \le \beta \le 1$ be such that for every circuit C of size at most s it holds $\Pr_{x\in\{-1,+1\}^n}[C(x) = f(x)] \le 1/2 + \beta/2$. Then for every $E > 0$, any k and any circuit C' of size $s' \le s \cdot poly(E, 1/k, 1/n)$ it holds $\Pr_{x_1,\ldots,x_k\in\{-1,+1\}^n}[C'(x_1,\ldots,x_k) = \prod_{i=1}^{k} f(x_i)] \le 1/2 + \beta^k/2 + E$.*

Note that in order to make the correctness probability exponentially close to 1/2, it is necessary to make E exponentially small and hence also $s' \le se^{\Omega(-k)}$. This will also show up in the following threshold DPT, and we do not know of any way to circumvent this problem, even using different techniques.

Theorem 14. *Let $f : \{-1,+1\}^n \to \{-1,+1\}$ be a function and let $0 \le \beta \le 1$ such that for every circuit C of size at most s it holds $\Pr_{x\in\{-1,+1\}1^n}[C(x) = f(x)] \le 1/2 + \beta/2$. Let $\beta \le \lambda \le 1$. Then for any k and any circuit C' of size $s' \le s \cdot poly(e^{-\Omega(k)}, 1/n)$ (which takes k n-bit inputs and outputs k bits) it holds*

$$\Pr_{x_1,\ldots,x_k\in\{-1,+1\}^n}[\sum_{i=1}^{k} C'(x_1,\ldots,x_k)_i f(x_i) \ge \lambda k] \le e^{-\Omega(k)}.$$

Proof. The result follows by Yao's XOR Lemma with parameter $E = \left(\frac{(1+\lambda)(1-\beta)}{(1+\beta)(1-\lambda)}\right)^{-k(1/2-\lambda/2)} e^{-\Omega(k)}$ and then applying inequality (8) of our threshold Lemma. □

Acknowledgements

I thank Madhur Tulsiani, Thomas Vidick, Troy Lee, Ronald de Wolf and Mario Szegedy.

References

[1] Mark Adcock and Richard Cleve. A quantum Goldreich-Levin theorem with cryptographic applications. In *STACS '02: Proceedings of the 19th Annual Symposium on Theoretical Aspects of Computer Science*, pages 323–334, 2002.

[2] L. Babai, L. Fortnow, and C. Lund. Non-deterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[3] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45(2):204–232, 1992.

[4] Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and ldcs. *Foundations of Computer Science, Annual IEEE Symposium on*, 0:477–486, 2008.

[5] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual Conference on Computational Complexity*, page 120, 2001.

[6] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *Proc. 15th STOC*, pages 94–99, Boston, Massachusetts, 1983. ACM Press.

[7] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(002), 2008.

[8] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Math. Stat.*, 23:493–509, 1952.

[9] Benny Chor, Oded Goldreich, Johan Hastad, Joel Freidmann, Steven Rudich, and Roman Smolensky. The bit extraction problem or t-resilient functions. *Symposium on Foundations of Computer Science*, pages 396–407, 1985.

[10] Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM Journal on Discrete Mathematics*, 6(1):110–123, 1993.

[11] R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *IEEE Conference on Computational Complexity*, pages 236–249, 2004.

[12] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. *Special Issue of 22nd IEEE Conference on Computational Complexity*, 17(2):282–299, 2008.

[13] Yevgeniy Dodis, Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Security amplification for interactive cryptographic primitives. In *Sixth Theory of Cryptography Conference*, 2009.

[14] Oded Goldreich and Hugo Krawczyk. On the composition of zero-knowledge proof systems. *SIAM Journal on Computing*, 25:169–192, 1996.

[15] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR lemma. Technical Report TR95-050, Electronic Colloquium on Computational Complexity, March 1995.

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

[17] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[18] Thomas Holenstein. Key agreement from weak bit agreement. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 664–673, 2005.

[19] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proc. 36th FOCS*, pages 538–545, Los Alamitos, CA, USA, 1995. IEEE Computer Society.

[20] Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Chernoff-type direct product theorems. In *In Proceeding of the Twenty-Seventh Annual International Cryptology Conference (CRYPTO 2007*, pages 500–516, 2007.

[21] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, Ben Toner, and Thomas Vidick. Entangled games are hard to approximate. *Foundations of Computer Science, Annual IEEE Symposium on*, 0:447–456, 2008.

[22] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, and Thomas Vidick. Using entanglement in quantum multi-prover interactive proofs. In *CCC '08: Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, pages 211–222, 2008.

[23] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007. Earlier version in FOCS'04.

[24] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, Cambridge, 1997.

[25] Troy Lee and Adi Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. In *CCC '08: Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, pages 81–91, 2008.

[26] Troy Lee, Adi Shraibman, and Robert Špalek. A direct product theorem for discrepancy. In *CCC '08: Proceedings of the 2008 IEEE 23rd Annual Conference on Computational Complexity*, pages 71–80, 2008.

[27] Leonid A. Levin. One way functions and pseudorandom generators. *Combinatorica*, 7(4):357–363, 1987.

[28] Ueli Maurer and Stefano Tessaro. Computational indistinguishability amplification: Tight product theorems for system composition. In *CRYPTO 2009*, pages 350–368, 2009.

[29] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993.

[30] Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. *SIAM J. Comput.*, 28(3):1035–1050, 1999.

[31] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of 40th ACM STOC*, 2008.

[32] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[33] Ran Raz. A counterexample to strong parallel repetition. In *Foundations of Computer Science*, pages 369–373, 2008.

[34] Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003.

[35] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. In *Proc. 40th STOC*, pages 589–598, Victoria, Canada, 2008. ACM Press.

[36] U. Vazirani. Efficiency considerations in using semi-random sources. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 160–168, 1987.

[37] Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

[38] Andrew Yao. Theory and applications of trapdoor functions (extended abstract). In *Proc. 23rd FOCS*, pages 80–91, Los Alamitos, CA, USA, 1982. IEEE Computer Society.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de