comput. complex. **12** (2003), 1 – 22
1016-3328/03/020001–22
DOI 10.1007/s00037-003-0175-x

computational complexity

TOWARDS PROVING STRONG DIRECT PRODUCT THEOREMS

Ronen Shaltiel

Abstract. A fundamental question of complexity theory is the direct product question. A famous example is Yao's XOR-lemma, in which one assumes that some function f is hard on average for small circuits (meaning that every circuit of some fixed size s which attempts to compute f is wrong on a non-negligible fraction of the inputs) and concludes that every circuit of size s' only has a small advantage over guessing randomly when computing $f^{\oplus k}(x_1, \ldots, x_k) = f(x_1) \oplus \cdots \oplus f(x_k)$ on independently chosen $x_1, \ldots, x_k$. All known proofs of this lemma have the property that $s' < s$. In words, the circuit which attempts to compute $f^{\oplus k}$ is **smaller** than the circuit which attempts to compute f on a single input! This paper addresses the issue of proving **strong** direct product assertions, that is, ones in which $s' \approx ks$ and is in particular **larger** than s. We study the question of proving strong direct product question for decision trees and communication protocols.

Keywords. Product theorems, XOR-lemma, hardness amplification, average case complexity.

Subject classification. 68Q17, 68Q15.

1. Introduction

1.1. The direct product question. Suppose you are given a biased coin (that is, one in which the probability to get heads is $1/2 + \alpha$), and you toss it k times and compute the exclusive-or of the outcomes. It is easy to see that the bias of the "new coin" you obtain goes to zero exponentially fast with k. (More precisely, the "new coin" will have probability $1/2 + (2\alpha)^k/2$.)[1] The direct product question asks to what extent this is true in the computational world.

[1] A nice measure of the bias of a random coin is the *advantage* of the coin which is defined to be the difference between the probabilities of heads and tails, or 2α in the notation above. The nice thing about this measure is that the advantage of the "new coin" is exactly the advantage of the original coin raised to the kth power. To see this one encodes heads as 1 and tails as -1. In this encoding the advantage of a coin Z is given by $E(Z)$, and exclusive-or is multiplication. It now follows that the advantage of $\prod Z_i$ is given by $E(\prod Z_i) = \prod E(Z_i)$.

Suppose you have a boolean function which is hard on average for some complexity class. This intuitively means that from the point of view of any algorithm from that class the outcome of the function on a uniformly chosen input looks like a (biased) random coin. Is it true that XORing k independent copies of this "biased coin" produces a bit which is much less biased from the point of view of algorithms in the given complexity class? Such assertions are referred to as direct product assertions. It turns out that proving such assertions is much more involved than one might expect at first glance.

1.1.1. Informal statement of the question. Given a boolean function f over domain X and an integer k, we define a function $f^{\oplus k} : X^k \to \{0,1\}$ by

$$f^{\oplus k}(x_1, \dots, x_k) = f(x_1) \oplus \cdots \oplus f(x_k).$$

The intuition presented before suggests that if f is hard on average (say f can be computed correctly by some complexity class on at most a $(1/2+\alpha)$-fraction of the inputs), then we expect $f^{\oplus k}$ to be computed correctly on a fraction of inputs which is about $1/2 + \alpha^k$. In other words, we expect $f^{\oplus k}$ to become "exponentially harder" on average. What makes this problem much harder than the information theoretic problem about the exclusive-or of independent random coins presented above is that the algorithm attempting to compute $f^{\oplus k}(x_1, \dots, x_k)$ may correlate computations on different inputs, which intuitively corresponds to making the coins correlated. We will use an "abstract model of computation" to model the complexity class, as the direct product question can be stated in many computational models.

1.1.2. An abstract model of computation. Consider some computational resource (such as circuit size, decision tree depth, number of bits exchanged in a communication protocol…). Let Res_r denote the class of all functions computable using r "units" of the resource. In this paper we consider the following concrete classes:

- Size_s, the family of functions computed by circuits of size s.
- Comm_c, the family of functions (on two inputs) computed by a deterministic communication protocol which exchanges c bits.
- Depth_d, the family of functions computed by decision trees of depth d.

Saying that a function f is "hard on average" for Res_r means that every algorithm from Res_r computes f correctly on a fraction of the inputs which is bounded away from one[2]. We use the following notation:

$$\mathrm{Suc}_r^{\mathrm{Res}}(f) = \max_{P \in \mathrm{Res}_r} \Pr_{x \in_R X}[P(x) = f(x)].$$

Since f is boolean it can always be computed correctly on at least half the inputs. Thus, we will be interested in the advantage the algorithm can get over guessing randomly, which is given by $\mathrm{Suc}_r^{\mathrm{Res}}(f) - 1/2$. It turns out that it is nicer to work with this quantity when it is normalized. We define the *advantage* Res_r has on f in the following way:

$$\mathrm{Adv}_r^{\mathrm{Res}}(f) = 2(\mathrm{Suc}_r^{\mathrm{Res}}(f) - 1/2).$$

When normalized this way, the advantage also has the following useful interpretation (see also footnote 1):

$$\mathrm{Adv}_r^{\mathrm{Res}}(f) = \max_{P \in \mathrm{Res}_r} \Pr_{x \in_R X}[P(x) = f(x)] - \Pr_{x \in_R X}[P(x) \neq f(x)].$$

1.1.3. Formal statement of the direct product question. The direct product question can now be presented as follows:

The direct product problem: Is it true that for all f and r, k:

$$\mathrm{Adv}_r^{\mathrm{Res}}(f) \leq p \Rightarrow \mathrm{Adv}_{r'}^{\mathrm{Res}}(f^{\oplus k}) \leq p',$$

where r' and p' are parameters which may depend on r, p and k? In words, one supposes that f is hard on average to algorithms with r units of the resource and concludes that $f^{\oplus k}$ is hard on average to algorithms having r' units of the resource. Naturally, the assertion is stronger when r' is large and p' is small. It seems reasonable to allow the algorithm attempting to compute $f^{\oplus k}$ to use $r' = kr$ units of the resource. This will at least enable it to run k copies of the best algorithm for f in Res_r on the k independent inputs. This strategy indeed computes f with advantage $\mathrm{Adv}_r^{\mathrm{Res}}(f)^k$. Thus, we say that the assertion is *optimal* when $r' = kr$ and $p' = p^k$.

However, it turns out that such assertions are often proven for much smaller r'. As we will see, in some cases only results with $r' \ll r \ll kr$ are known. In this paper we are interested in proving direct product assertions for *large* r'. We will call such assertions *strong* if $r' = \Omega(kr)$ and $p' = p^{\Omega(k)}$.

The strong-direct product problem: Is it true that for all f and r, k,

$$\mathrm{Adv}_r^{\mathrm{Res}}(f) \leq p \Rightarrow \mathrm{Adv}_{\Omega(kr)}^{\mathrm{Res}}(f^{\oplus k}) \leq p^{\Omega(k)}?$$

[2] In this paper we restrict ourselves to average case hardness relative to the uniform distribution. Some of our results do not generalize to arbitrary probability distributions. See the open problems in Section 6.

The choice of $r' = \Omega(kr)$ is not that important. Most of our results give tradeoffs between r' and p'. What is important (and different than most of the previous work in this area) is that we are interested in $r' \gg r$.

1.1.4. The concatenation variant. A different variant of the direct product problem that is often considered is the *concatenation variant.* It involves replacing the function $f^{\oplus k}$ with $f^{(k)}(x_1, \ldots, x_k) = (f(x_1), \ldots, f(x_k))$. In words, rather than trying to compute the exclusive-or of the outputs of the function on independent inputs, the algorithm is asked to compute *all* the outputs correctly simultaneously. In this setup the wanted assertion is the following:

$$\mathrm{Suc}_r^{\mathrm{Res}}(f^{(k)}) \leq p \Rightarrow \mathrm{Suc}_{r'}^{\mathrm{Res}}(f) \leq p'.$$

Once again, it is desired to have $r' = kr$ and $p' = p^k$, and one can define optimal and strong such assertions in an analogous way. The two variants of the direct product question are related. In particular, an optimal direct product theorem in the xor variant implies an optimal direct product theorem in the concatenation variant.

1.2. Previous work. The most studied computational model for direct product results is boolean circuits. The so-called "Yao's XOR-lemma" (Yao 1982) can be stated this way in our terminology:

$$\mathrm{Adv}_s^{\mathrm{Size}}(f) \leq p \Rightarrow \mathrm{Adv}_{s'}^{\mathrm{Size}}(f^{\oplus k}) \leq p^k + \epsilon,$$

where $s' = s(\epsilon/n)^{O(1)}$, and n is the number of inputs of f. Note that in this result s' is actually *smaller* than s. In other words, the circuit which tries to compute f on many instances is smaller than the one which tries to compute f on one instance. This is unavoidable in the sense that all known proofs of this lemma (Goldreich *et al.* 1995; Impagliazzo 1995; Impagliazzo & Wigderson 1997; Levin 1987) work by proving the contrapositive claim: $\mathrm{Adv}_{s'}^{\mathrm{Size}}(f^{\oplus k}) > p^k + \epsilon \Rightarrow \mathrm{Adv}_s^{\mathrm{Size}}(f) > p$, and use the circuit which computes $f^{\oplus k}$ too well as a subcircuit in the circuit that computes f too well. (See Goldreich *et al.* 1995 for a survey on Yao's XOR-lemma.)

Another weakness of this result is that p' is always larger than $1/s$, which means that one does not benefit from taking $k > \log s$. An unpublished result which is commonly attributed to Steven Rudich shows that all "black box"[3] proofs of the XOR-lemma suffer from this flaw. Thus, proving a result in

[3] "Black box" refers to proofs like the ones mentioned above, that use a circuit which computes $f^{\oplus k}$ too well as a black box in a circuit that computes f too well.

which $s' > s$ or $p' < 1/s$ seems to be beyond our current ability, as we do not know how to handle boolean circuits other than using them as black boxes. It should be noted that despite these weaknesses Yao's XOR-lemma has many applications in complexity theory.

The direct product question was also studied in other computational models. (We mention only previous work which is relevant to this paper, and the interested reader may find more references in Impagliazzo *et al.* 1994). Nisan *et al.* (1999) study the concatenation variant of the direct product question in decision trees. They consider a specific variant of decision trees which they call "decision forests". A k-*decision forest* of depth d consists of k decision trees of depth d. Each is allowed to query all k inputs, and the ith tree is supposed to compute $f(x_i)$. The final output of the decision forest is the concatenation of outputs of individual trees. Let us denote the class of all functions computable by depth d k-decision forests by $\mathrm{Forest}_{k,d}$. With this terminology their result could be stated this way:

$$\mathrm{Suc}_d^{\mathrm{Depth}}(f) \le p \Rightarrow \mathrm{Suc}_{k,d}^{\mathrm{Forest}}(f^{(k)}) \le p^k.$$

(Here $f^{(k)}(x_1, \ldots, x_k) = (f(x_1), \ldots, f(x_k))$, see Section 1.1.4). This result is optimal in the sense that a decision forest of depth d can run k decision trees of depth d in parallel and compute $f^{(k)}$ with success $\mathrm{Suc}_d^{\mathrm{Depth}}(f)^k$. Parnafes *et al.* (1997) used the technique of Raz's parallel repetition theorem (Raz 1998) to study the concatenation variant of the direct product question for communication protocols. They prove a product theorem for "forests of c-bit communication protocols". (A *forest of c-bit communication protocols* is a collection of k c-bit communication protocols, each is over all k inputs, and the ith protocol is supposed to compute the function on the ith input.) Their result is similar in flavor to that of Nisan *et al.* (1999) with the exception that $p' = p^{\Omega(k/c)}$. This dependence on c comes from the technique of Raz, but whereas a dependence on c is unavoidable in the parallel repetition theorem (as was shown by Feige & Verbitsky 2002), it is open whether the result of Parnafes *et al.* (1997) is best possible for forests of communication protocols.

1.3. Our results. Our first result is a general counterexample which shows that strong direct product assertions (or even ones with r' sufficiently larger than r) are simply not true. This counterexample applies to many models of computation and in particular to boolean circuits, communication protocols and decision trees.

While this counterexample rules out the possibility of proving strong direct product assertions, it seems to exploit defects in the formulation of the problem rather than show that our general intuition for direct product assertions is false. Intuitively, the algorithm of the counterexample is able to compute $f^{\oplus k}$ correctly with high probability by using its resources in an unbalanced way allocating a lot of its resources to specific instances. This does not contradict our intuition for why strong direct product assertions are true as this is not a counterexample to our belief that it is not beneficial for the algorithm to correlate computations on different inputs. We elaborate on this point in Section 3.3. In any case, as the assertion is not true as is, in order to capture our intuition and prove results with a strong direct product flavor we have to either strengthen the assumption or weaken the conclusion.

1.3.1. Strengthening the assumption: demanding more information on the function. The function presented in the counterexample has a large subset of "easy inputs", and it is feasible to check whether a given input is "easy". It is natural to ask what kind of restrictions can be placed on the function in order to make a strong direct product assertion hold. We give such a restriction for communication protocols. This is done by insisting that the function f has low discrepancy. (The discrepancy of $f(x, y)$, denoted by $\mathrm{disc}(f)$ measures how unbalanced f is in large rectangles.) It is standard that low discrepancy entails that the function is hard on average for communication protocols. More precisely,

$$\mathrm{Adv}_c^{\mathrm{Comm}}(f) \le \mathrm{disc}(f)2^c.$$

However, having low discrepancy is stronger than being hard on average and it intuitively says that the function has no large recognizable subset of easy inputs. The main result of this paper is the following inequality:

$$\mathrm{Adv}_{kc/3}^{\mathrm{Comm}}(f^{\oplus k}) \le O(\mathrm{disc}(f)2^c)^{k/3}.$$

(We stress that the constant hidden in the $O(\cdot)$ notation is a universal constant and does not depend on the choice of f.) This inequality has the following interpretation: If the fact that $\mathrm{Adv}_c^{\mathrm{Comm}}(f) \le p$ follows from the fact that f has low discrepancy ($\mathrm{disc}(f) \le p2^{-c}$) then a strong direct product assertion holds for f. We would like to point out that the "discrepancy method" is the most common way to prove that f is hard on average for communication protocols.

We prove the above statement by proving a "product theorem" for discrepancy. More precisely, we prove that

$$\mathrm{disc}(f^{\oplus k}) = O(\mathrm{disc}(f))^{k/3}.$$

(Again, the constant hidden in the $O(\cdot)$-notation does not depend on f.) The main step in this proof establishes a connection between the discrepancy of a matrix and its spectral norm.

1.3.2. Discrepancy and spectral norm. A communication complexity problem f can be encoded as a matrix A where $A_{xy} = (-1)^{f(x,y)}$. The main lemma of the paper gives the following connection between the discrepancy of A and its spectral norm $\|A\|_2$:

$$\Omega\left(\frac{\|A\|_2}{N}\right)^3 \leq \operatorname{disc}(A) \leq \frac{\|A\|_2}{N}.$$

The proof uses the method of Nisan & Wigderson (1995). This lemma immediately implies the "product theorem" for discrepancy as $\|A^{\otimes k}\|_2$ (the spectral norm of the tensor product of A with itself k times) is equal to $\|A\|_2^k$. Thus, once one can switch between discrepancy and spectral norm, the multiplicativity of the spectral norm gives the "product theorem" for discrepancy.

1.3.3. Weakening the conclusion: imposing restrictions on the algorithm. Another way to prove a strong direct product assertion is to weaken the conclusion. We suggest proving the assertion only for algorithms with certain restrictions. In a way the "forest model" of Nisan *et al.* (1999) is such a restriction. However, the forest model is only suitable for the "concatenation variant" of the direct product question and makes no sense in the "XOR variant". In this paper we suggest a different restriction. The algorithm presented in the counterexample has the property that it uses its resource in an "unfair" way spending more than r units on particular inputs. We suggest a "fairness" restriction on the algorithm. Some evidence for the potential of this direction is that we can prove an optimal direct product assertion for "fair" decision trees.

A decision tree of depth kd over variables $x_1, \ldots, x_k$ is *fair* if on every path from the root to a leaf at most d bits from each variable are queried. Let us denote the class of fair decision trees of depth kd by $\mathrm{FairDepth}_{kd}$. It is not hard to prove that

$$\mathrm{Adv}_d^{\mathrm{Depth}}(f) \leq p \Rightarrow \mathrm{Adv}_{kd}^{\mathrm{FairDepth}}(f^{\oplus k}) \leq p^k.$$

It is our hope that the two directions we present here can be extended to prove strong direct product assertions for stronger computational models.

1.4. Organization of the paper. In Section 3 we present our counterexample. In Section 4 we prove a strong direct product assertion for communication protocols assuming the function has low discrepancy. In Section 5 we prove a strong direct product theorem for fair decision trees. In Section 6 we present open problems and possible generalizations.

2. Preliminaries

We use $\oplus$ to denote the exclusive-or. For two matrices A and B of size $N \times N$, we use $A \otimes B$ to denote their *tensor product*. More precisely $A \otimes B$ is an $N^2 \times N^2$ matrix. We think of it as an $N \times N$ matrix with entries being matrices of size $N \times N$ and place a copy of the matrix $A_{ij} \cdot B$ in the ith row and jth column. The tensor product of A with itself k times is denoted by $A^{\otimes k}$.

We use Size_s to denote the class of all functions (over arbitrary number of inputs) computable by boolean circuits of size s.

We use Comm_c to denote the class of all functions of two arguments $f(x, y)$ which can be computed by a c-bit communication protocol. The exact definition of a communication protocol can be found in any textbook on this subject (e.g., Kushilevitz & Nisan 1997). Loosely speaking, a communication protocol is a protocol for two players which works in steps. At each step the protocol specifies a player, and this player sends a bit which may depend on his input and previously sent bits. The only property of such protocols used in this paper is that such a protocol induces a partition of the inputs into 2^c rectangles, and on each such rectangle the answer of the protocol is constant.

We use Depth_d to denote the class of all functions computed by a decision tree of depth d. A *decision tree* is a binary tree in which every internal node is labeled with a specific bit of the input, and leafs are labeled with outputs. An input to the decision tree defines a path from root to leaf in the obvious way (at each node the variable which is labeled by the node is inspected and the path continues to the left son if the value is zero and to the right one if the value is one). The output of the tree on this input is the leaf label.

3. A general counterexample

In this section we give a general counterexample to direct product assertions with $r' \gg r$ which works for boolean circuits, decision trees and communication protocols. We show that given a function which is hard given r units of the resource and easy given slightly more units, we can construct a function which is hard given r units of the resource, and yet computing it on k independent inputs is easy. We present the example using our general notation in Section 3.1 and then draw conclusions for specific models in Section 3.2. In Section 3.3 we discuss the implications of this counterexample.

3.1. The general setting. We will present a function which is hard on average given r units of the resource, yet $f^{\oplus k}$ can be computed correctly with probability $1 - 2^{-\Omega(k)}$ given r' units of the resource, for r' sufficiently larger than r.

The counterexample works assuming the existence of a function which is hard given r units of the resource, and easy given slightly more units. Formally, we assume the existence of a function $g : \{0,1\}^n \to \{0,1\}$ and $r < \bar{r}$ such that:

- $\mathrm{Suc}_r^{\mathrm{Res}}(g) \leq 3/4$,
- $g \in \mathrm{Res}_{\bar{r}}$.

Another ingredient is an easy function (over few inputs) which answers one on a prescribed fraction of its inputs. Formally, given a rational constant $q < 1$ we assume the existence of a function $h : \{0,1\}^l \to \{0,1\}$ and a small number r^* such that

- $h \in \mathrm{Res}_{r^*}$,
- $\Pr_{y \in_R \{0,1\}^l}[h(y) = 1] = q$.

Our counterexample function is a combination of the easy and hard functions.

Definition 3.1. We define a function $f_q : \{0,1\}^n \times \{0,1\}^l \to \{0,1\}$ in the following way:

$$f(x,y) = \begin{cases} g(x) & \text{if } h(y) = 1, \\ 0 & \text{if } h(y) = 0. \end{cases}$$

An algorithm for computing $f^{\oplus k}$ can utilize its resource smartly by spending a lot of the resource on the (expectedly few) inputs in which f involves the hard function, and spend a very small amount on other inputs. This is made formal in the following lemma.

"Lemma" 3.2. [4] *The following inequalities hold:*

- $\mathrm{Suc}_r^{\mathrm{Res}}(f) \leq 1 - q/4$.
- *If* $r' \geq 2qk\bar{r} + kr^*$ *then* $\mathrm{Suc}_{r'}^{\mathrm{Res}}(f^{\oplus k}) \geq 1 - 2^{-\Omega(k)}$.

[4] The lemma is put in quotes because the formal statement requires some natural properties of the "abstract" computational model. Stating these properties precisely is tedious. The reader can verify that the algorithm described in the proof can be carried out in any of the models studied in this paper: boolean circuits, communication protocols and decision trees (as well as in any computational model that comes to mind).

Proof. For the first item note that an algorithm which is correct on f with probability greater than $1-q/4$ induces an algorithm that is correct on g with probability greater than $3/4$. More precisely, given an input for g, one can fix a y such that $h(y)=1$ and run the algorithm on the pair of inputs. For the second item, note that when $(y_1,\dots,y_k)$ are randomly chosen, we expect qk of them to have $h(y_i)=1$. By Chernoff's inequality the probability that more than $2qk$ of them have $h(y_i)=1$ is bounded by $2^{-\Omega(k)}$. We can check which of the y_i's have $h(y_i)=1$ using kr^* units of the resource. Assuming the constant function zero can be computed using 0 units of the resource, we can compute the function f on (x_i,y_i)'s such that $h(y_i)=0$. Assuming that there are at most $2qk$ of y_i's such that $h(y_i)=1$ we can use $2qk\bar{r}$ units to compute the outputs of the "hard inputs". (Here we use some natural closure properties of the computational model.) Thus, $\mathrm{Suc}^{\mathrm{Res}}_{kr^*+2qk\bar{r}}(f^{\oplus k}) \geq 1-2^{-\Omega(k)}$. □

Remark 3.3. It should be noted that the function f constructed here is not as pathological as it may seem at first glance. Impagliazzo's hard core theorem (Impagliazzo 1995) shows that (at least in the boolean circuit model) every function f with $\mathrm{Suc}^{\mathrm{Size}}_{s}(f) \leq 1-q$ has a large subset of the inputs on which any (slightly smaller) circuit succeeds with probability roughly 1/2. In our example the "hard core" of f is the function g. The unnatural state of affairs in our example is that the function is easy outside of the hard core, and deciding whether an input is in the hard core is an easy computational task.

3.2. Conclusions for specific models.

In order to use the counterexample from the previous section we will show the existence of the required "building block" functions g and h for various computational models.

We will use the same function as "h" in all constructions. Namely we choose $q=2^{-l}$ for integer l, and define $h:\{0,1\}^l \to \{0,1\}$ to take the value one if all its inputs are zeroes, and zero otherwise. It is immediate to verify that h is in $\mathrm{Comm}_l, \mathrm{Depth}_l, \mathrm{Size}_{O(l)}$, and accepts a q-fraction of its inputs.

3.2.1. Communication protocols.

For communication complexity we use the inner product function as "g". More precisely, consider the function $g:\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined by $g(x,y)=\sum_{1\leq i\leq n} x_i y_i \bmod 2$. As do all functions over n bit inputs, $g \in \mathrm{Comm}_n$. It is known that g is very hard on average given $n/4$ bits of communication[5].

[5] Preparing for Section 4, we remark that the proof of that statement works by showing that g has very low discrepancy.

Theorem 3.4 (Chor & Goldreich 1988). $\mathrm{Adv}^{\mathrm{Comm}}_{n/4}(g) \leq 2^{-\Omega(n)}$.

Using "Lemma" 3.2 with a sufficiently small constant q we conclude that for $c = n/4$ and large enough k, $\mathrm{Adv}^{\mathrm{Comm}}_{\Omega(kc)}(f^{\oplus k}) > \mathrm{Adv}^{\mathrm{Comm}}_{c}(f)$. Thus, there are no strong direct product assertions for communication protocols.

3.2.2. Decision trees. For decision trees we choose g to be the parity function $g(x) = \bigoplus_{1 \leq i \leq n} x_i$. It is immediate that $g \in \mathrm{Depth}_n$ and $\mathrm{Adv}^{\mathrm{Depth}}_{n-1}(g) = 0$. Once again we can use "Lemma" 3.2 to get a counterexample with the same behavior of parameters as the counterexample for communication protocols.

3.2.3. Boolean circuits. The counterexample applies also to boolean circuits. That is, it is possible to prove the existence of a function g which is hard on average for some size and easy for slightly larger size. (Any function can be computed in size roughly $2^n/n$ and a counting argument could be used to show that there is a function which is hard on average for slightly smaller size.) However, in this model, there are much stronger counterexamples. This is shown by the following "mass production" theorem by Uhlig.

Theorem 3.5 (Uhlig 1974). *For every function* $f : \{0,1\}^n \to \{0,1\}$ *and* $k = 2^{o(n/\log n)}$, *we have* $f^{\oplus k} \in \mathrm{Size}_{2^n n^{-1} + o(2^n n^{-1})}$.

This means that if f has near maximal circuit complexity then huge savings can be made when computing f on many inputs. This is much stronger than our counterexample as in this setup savings can actually be made on the worst $x_1, \ldots, x_k$ whereas our example only shows that savings can be made on "average" $x_1, \ldots, x_k$.

It should be pointed out that this is not the case in decision trees. It is easy to show that if $f \notin \mathrm{Depth}_d$ then $f^{\oplus k} \notin \mathrm{Depth}_{kd}$, and no savings can be made on the worst case. It is not known whether significant savings can be made on the worst case in communication protocols. This question is known as the "direct sum conjecture" of Karchmer *et al.* (1995). In some sense (see Karchmer *et al.* 1995), a negative answer will supply a lower bound for boolean circuits with small depth.

3.3. Interpretation of the counterexample. It seems that the counterexample does not contradict our intuition as to why direct product assertions are true. When proving a direct product assertion, the main task is to show that the algorithm does not benefit from correlating computations on different inputs. In the counterexample we presented no real correlations occur. Instead, the algorithm uses its resource in an unbalanced way, spending a lot of it on particular inputs. This is best explained when considering the concatenation variant of the direct product question. In this case, the algorithm outputs answers for each input. Note that with probability $1 - 2^{\Omega(k)}$ each of these answers depend only on its corresponding input. Moreover, the algorithm is able to compute f on single instances with advantage greater than p. This is something which is ruled out when $r' \leq r$.

We would like to change the formulation of the direct product problem to rule out such cases. In the next two sections we suggest two such approaches. One involves adding assumptions on the function f, in hope that such assumptions can prevent the situation of the counterexample. Intuitively, if f is hard in a "robust way", any additional resources spent on f on one instance will result in a "loss" in another instance, and it will not be beneficial to treat the inputs unfairly. The other approach involves restricting the algorithm in a way that ensures that it cannot have advantage greater than p when attempting to compute f on any single coordinate.

4. A discrepancy product theorem

In the previous section we have seen that a direct product assertion for communication protocols is not true if we allow the protocol trying to compute $f^{\oplus k}$ to pass slightly more bits than the protocol attempting to compute f. In this section we show that if f has "low discrepancy" then a strong direct product theorem holds for f.

A communication complexity problem $f(x, y)$ can be viewed as a matrix A such that $A_{xy} = f(x, y)$. It will be convenient to think of the outputs of a communication complexity problem as $\{-1, 1\}$ rather than $\{0, 1\}$, thus the matrix will have entries in $\{-1, 1\}$. The choice of $\{-1, 1\}$ is made so that the tensor product of A with itself k times (denoted by $A^{\otimes k}$) is exactly the matrix of the communication problem $f^{\oplus k}$. We assume the inputs of both players are of the same size, which means that A is a square matrix.

For a set $C \subseteq [N]$, we use χ_C to denote the *characteristic vector* of C, that is, $(\chi_C)_i = 1$ for $i \in C$ and $(\chi_C)_i = 0$ for $i \notin C$.

Definition 4.1. Let A be an N by N matrix with entries in $\{-1, 1\}$. For a rectangle $R = C \times D$, where $C, D \subseteq [N]$, we define

$$\mathrm{disc}_R(A) = \frac{|\chi_C^t A \chi_D|}{N^2}.$$

The *discrepancy* of A is defined in the following way:

$$\mathrm{disc}(A) = \max_R \mathrm{disc}_R(A),$$

where the maximum is taken over all rectangles $R = C \times D$ with $C, D \subseteq [N]$.

For a fixed rectangle $R = C \times D$, $|\chi_C^t A \chi_D|/|C \times D|$ measures how unbalanced the matrix A is in the rectangle R. If R is a rectangle reached in a leaf of a communication protocol then this quantity is the advantage the protocol gets in the rectangle R. The definition of $\mathrm{disc}_R(A)$ multiplies this quantity by $|C \times D|/N^2$ to take into account the volume of the rectangle. More precisely, the advantage is multiplied by the volume of the rectangle R to give the contribution of R to the overall advantage of the protocol. This normalization is made so that low discrepancy will imply that the problem is hard on average. This is made formal in the following lemma.

Lemma 4.2. $\mathrm{Adv}_c^{\mathrm{Comm}}(A) \leq \mathrm{disc}(A)2^c$.

Proof. Let P be the c-bit communication protocol which achieves the maximum in the definition of $\mathrm{Adv}_c^{\mathrm{Comm}}(A)$. A c-bit communication protocol partitions A into 2^c disjoint rectangles. On each rectangle $R_i = C_i \times D_i$ the advantage of the protocol in the rectangle is given by $|\chi_{C_i}^t A \chi_{D_i}|/|C_i \times D_i|$. We can now bound the advantage of P:

$$\mathrm{Adv}_c^{\mathrm{Comm}}(A) = \sum_{1 \leq i \leq 2^c} \frac{|C_i \times D_i|}{N^2} \cdot \frac{|\chi_{C_i}^t A \chi_{D_i}|}{|C_i \times D_i|} \leq 2^c \mathrm{disc}(A). \qquad \square$$

The requirement that $\mathrm{disc}(A)$ is small is stronger than that A is hard on average. Still, the most common way of showing that communication problems are hard on average is by showing that they have low discrepancy.

Remark 4.3. In Remark 3.3 we pointed out that the function of the counterexample has a large set of easy inputs, and it is possible to check whether a given input is easy. Intuitively, low discrepancy exactly avoids this kind of scenario as it enforces that in any large rectangle the function is hard. It is impossible that the two players pass few bits and reach a large set on which the function is easy.

We show that the discrepancy of the tensor product of A with itself k times goes down exponentially with k.

Theorem 4.4. *There exists some constant a such that for every matrix A and integer $k > 1$, $\mathrm{disc}(A^{\otimes k}) \leq (a \cdot \mathrm{disc}(A))^{k/3}$.*

This has the following interpretation: Suppose (as is often the case) that the fact that $\mathrm{Adv}_c^{\mathrm{Comm}}(A) \leq p$ follows from the fact that $\mathrm{disc}(A) \leq p2^{-c}$. In that case

$$\mathrm{Adv}_{kc/3}^{\mathrm{Comm}}(A^{\otimes k}) \leq \mathrm{disc}(A^{\otimes k})2^{kc/3} \leq (a \cdot \mathrm{disc}(A) \cdot 2^c)^{k/3} \leq (ap)^{k/3}.$$

In words, we get a strong direct product theorem for A. This is stated with more generality in the next corollary.

Corollary 4.5. *There exists some constant a such that for every matrix A and integers k, c', $\mathrm{Adv}_{c'}^{\mathrm{Comm}}(A^{\otimes k}) \leq (a \cdot \mathrm{disc}(A))^{k/3}2^{c'}$.*

Remark 4.6. While we do not know whether Theorem 4.4 is tight, it is impossible to get an estimate $\mathrm{disc}(A^{\otimes k}) \leq \mathrm{disc}(A)^{\Omega(k)}$. In fact, we now show an example of a matrix A where $\mathrm{disc}(A^{\otimes k})$ is constant and does not depend on k. Consider an N by N matrix B for $N = 2^n$, which corresponds to the following communication game: Each of the two players gets an n-bit vector and the players want to compute the exclusive-or of the $2n$ bits. It is easy to see that $\mathrm{disc}(B) = 1/4$ and does not depend on N. (To show that $\mathrm{disc}(B) \geq 1/4$ note that B has a trivial 2-bit communication protocol in which each player sends the exclusive-or of its bits). However, $B^{\otimes k}$ is equal to the matrix B of size $N^k \times N^k$. Thus, $\mathrm{disc}(B^{\otimes k}) = \mathrm{disc}(B) = 1/4$ for all k, and does not go down when k is increased.

The proof of the theorem will require the definition of the spectral norm of a matrix A.

Definition 4.7. For a vector x we use $\|x\|_2$ to denote the L_2-norm of x. For a matrix A, $\|A\|_2$ is defined to be $\max_{x:\|x\|_2=1} \|Ax\|_2$.

It will be useful to consider equivalent definitions of this norm.

Fact 4.8. *Equivalent definitions for $\|A\|_2$ are:*

(i) $\|A\|_2 = \max_{x:\|x\|_2=1,\, y:\|y\|_2=1} x^tAy$,

(ii) $\|A\|_2 = \max\{\sqrt{\lambda} \mid \lambda$ *is an eigenvalue of* $A^tA\}$.

A useful property of $\|A\|_2$ is that it is multiplicative under tensor product.

FACT 4.9. $\|A^{\otimes k}\|_2 = \|A\|_2^k$.

PROOF. The proof of Fact 4.9 consists of two steps. The first is to show this is true for symmetric matrices. If A is symmetric then it can be diagonalized. Moreover, $\|A\|_2 = |\lambda|$, where λ is the maximal eigenvalue of A in absolute value. It is easy to see that $A^{\otimes k}$ can also be diagonalized, and that its eigenvalues are exactly all products of eigenvalues of A. Thus, $\|A^{\otimes k}\|_2 = |\lambda|^k$. The fact now follows for non-symmetric matrices by using the second item of Fact 4.8. Note that the matrix A^tA is symmetric and that $(A^{\otimes k})^t A^{\otimes k} = (A^tA)^{\otimes k}$. □

In the remainder of this section we prove Theorem 4.4. The first step is to express the discrepancy in terms of the spectral norm. We then use the multiplicativity of the spectral norm to get the conclusion. The second item of Fact 4.8 enables us to upper bound the discrepancy using the spectral norm.

LEMMA 4.10. $\mathrm{disc}(A) \leq \|A\|_2/N$.

PROOF. Let $R = C \times D$ be a rectangle such that $\mathrm{disc}(A) = \mathrm{disc}_R(A)$. We have $\mathrm{disc}(A) = |\chi_C^t A \chi_D|/N^2$. We define $x = \chi_C/\sqrt{|C|}$ and $y = \chi_D/\sqrt{|D|}$. Note that $\|\chi_C\|_2 = \|\chi_D\|_2 = 1$. It follows from the first item of Fact 4.8 that

$$\|A\|_2 \geq |x^tAy| = \frac{|\chi_C^t A \chi_D|}{\sqrt{|C|}\sqrt{|D|}} \geq \frac{|\chi_C^t A \chi_D|}{N} \geq \mathrm{disc}(A)N. \qquad \square$$

Nisan & Wigderson (1995) address the so-called "log-rank conjecture" and show that $\mathrm{disc}(A) = \Omega(1/\mathrm{rank}(A)^{3/2})$. We use the technique from that paper to lower bound the discrepancy using the spectral norm.

LEMMA 4.11. $\mathrm{disc}(A) = \Omega(\|A\|_2/N)^3$.

We start by showing that Theorem 4.4 easily follows from Lemma 4.11.

PROOF (of Theorem 4.4).

$$\mathrm{disc}(A^{\otimes k}) \leq \frac{\|A^{\otimes k}\|_2}{N^k} = \left(\frac{\|A\|_2}{N}\right)^k = O(\mathrm{disc}(A))^{k/3}.$$

We have applied consecutively Lemma 4.10, Fact 4.9, and Lemma 4.11. □

We want to prove Lemma 4.11 in a similar way to the previous lemma. The first step is a way to transform a bilinear form with arbitrary vectors into one with characteristic vectors. Such a transformation was given in Nisan & Wigderson (1995).

Lemma 4.12 (Nisan & Wigderson 1995). *Let u, v be vectors such that $\|u\|_\infty, \|v\|_\infty \leq 1$. Then there exists a rectangle $R = C \times D$ such that $|\chi_C^t A \chi_D| \geq u^t Av/4$.*

For completeness we give the proof of this lemma.

Proof. Let $z = Av$, so that $u^t Av = u^t z$. There is a subset C of the coordinates such that $\sum_{i \in C} u_i z_i \geq u^t Av/2$. (This subset is either the indices at which both are positive or the indices at which both u and z are negative). In both cases, $|\chi_C^t Av| \geq \sum_{i \in C} u_i z_i \geq u^t Av/2$. We now repeat this argument one more time to find a subset D. If $\chi_C^t Av$ is negative, then we replace v by $v' = -v$ and otherwise we set $v' = v$. In both cases $\chi_C^t Av' = |\chi_C^t Av|$. Let $z = \chi_C^t A$, so that $\chi_C^t Av' = z^t v'$. By the same argument we find a set D so that $|\chi_C^t A \chi_D| \geq \chi_C^t Av'/2 = |\chi_C^t Av|/2 \geq u^t Av/4$. □

We are now tempted to use Lemma 4.12 directly to prove Lemma 4.11. That is, start from u, v with $\|u\|_2 = \|v\|_2 = 1$ such that $\|A\|_2 = u^t Av$ and get a rectangle $R = C \times D$ with roughly the same value. This will not do since to get a bound on $\mathrm{disc}(A)$ we have to divide by N^2. Thus, the above argument only gives the non-impressive estimate $\mathrm{disc}(A) \geq \|A\|_2/4N^2$. To do better we note that had it been the case that u and v had $\|u\|_\infty, \|v\|_\infty \leq \rho < 1$ we could deduce that $\rho^2 \mathrm{disc}(A) \geq \|A\|_2/4N^2$ and do better. Following Nisan & Wigderson (1995) we show that $\|A\|_2$ can be approximated by such u and v.

Lemma 4.13. *For every $\rho > 0$ there are vectors u, v with $\|u\|_\infty, \|v\|_\infty \leq \rho$ such that $u^t Av \geq \|A\|_2 - 2\sqrt{N}/\rho$.*

Proof. Let x and y be vectors such that $\|x\|_2 = \|y\|_2 = 1$ and $\|A\|_2 = x^t Ay$. Let $I = \{i \mid x_i > \rho\}$ and $J = \{j \mid y_j > \rho\}$. Let u be the vector obtained from x by setting the coordinates in I equal to zero, and v be the vector obtained from y by setting the coordinates in J equal to zero. Note that $|I|, |J| \leq 1/\rho^2$ (since otherwise the contribution of elements in I (resp. J) to the norm of x (resp. y) is greater than one). We now have

$$u^t Av \geq \|A\|_2 - \Big| \sum_{i \in I,\, j \in [N]} x_i a_{ij} y_j + \sum_{i \in [N],\, j \in J} x_i a_{ij} y_j \Big|.$$

We will now argue that the two terms on the right hand side are small because they involve small rectangles. We will bound the first term, and the second can be bounded the same way:

$$\Big|\sum_{i\in I,\, j\in[N]} x_i a_{ij} y_j\Big| \le \sum_{i\in I,\, j\in[N]} |x_i y_j| = \sum_{i\in I} |x_i| \sum_{j\in[N]} |y_j|$$

$$\le \sqrt{|I|}\sqrt{N}\, \|x\|_2 \|y\|_2 \le \frac{\sqrt{N}}{\rho}$$

(where the second inequality follows from the Cauchy–Schwarz inequality). Plugging this in the previous calculation we obtain

$$u^t A v \ge \|A\|_2 - \frac{2\sqrt{N}}{\rho}. \qquad \square$$

Lemma 4.11 now follows from the previous lemmas.

Proof (of Lemma 4.11). Given a matrix A, we set

$$\rho = \frac{3\sqrt{N}}{\|A\|_2}.$$

Let u, v be the vectors whose existence is given by Lemma 4.13. We now define new vectors $\bar{u} = u/\rho$ and $\bar{v} = v/\rho$. Note that $\|\bar{u}\|_\infty, \|\bar{v}\|_\infty \le 1$. By Lemma 4.12 there exists a rectangle $R = C \times D$ such that

$$|\chi_C^t A \chi_D| \ge \frac{\bar{u}^t A \bar{v}}{4} = \frac{u^t A v}{4\rho^2} = \frac{\|A\|_2 - 2\sqrt{N}/\rho}{4\rho^2} = \frac{\|A\|_2^3}{108N}.$$

We conclude that $\mathrm{disc}(A) = \Omega(\|A\|_2/N)^3$. □

5. Fair decision trees

In this section we prove an optimal direct product theorem for fair decision trees. The proof is quite simple and mimics the technique of Nisan *et al.* (1999). The purpose of this section is to promote the notion of fairness which can perhaps be extended to more interesting models of computation. We start with the definition of fairness for decision trees.

Definition 5.1. A decision tree over inputs $x_1, \ldots, x_k$ is $(d_1, \ldots, d_k)$-*fair* if for every $1 \le i \le k$ and on every path from the root to a leaf the decision tree queries at most d_i bits from x_i. A decision tree is d-*fair* if it is $(d, \ldots, d)$-fair. Let $\mathrm{FairDepth}_{d_1,\ldots,d_k}$ denote the class of functions over inputs $x_1, \ldots, x_k$ computed by $(d_1, \ldots, d_k)$-fair decision trees, and $\mathrm{FairDepth}_{kd} = \mathrm{FairDepth}_{d,\ldots,d}$.

Theorem 5.2. $\mathrm{Adv}_{kd}^{\mathrm{FairDepth}}(f^{\oplus k}) \le \mathrm{Adv}_{d}^{\mathrm{Depth}}(f)^k$.

The proof of this theorem is by induction and is similar to that of Nisan *et al.* (1999). To simplify the notation we will prove it for $k = 2$. The proof for general k follows in the same way. To have a stronger induction hypothesis we will prove the following stronger version.

Lemma 5.3. *For any two functions f_1, f_2 and numbers d_1, d_2,*

$$\mathrm{Adv}_{d_1,d_2}^{\mathrm{FairDepth}}(f \oplus g) \le \mathrm{Adv}_{d_1}^{\mathrm{Depth}}(f_1) \cdot \mathrm{Adv}_{d_2}^{\mathrm{Depth}}(f_2).$$

Proof. We prove the lemma by induction on $d_1 + d_2$. If $d_1 + d_2 = 0$ then the decision tree does not base its answer on the inputs. It is standard to check that indeed $\mathrm{Adv}_{0,0}^{\mathrm{FairDepth}}(f_1 \oplus f_2) = \mathrm{Adv}_{0}^{\mathrm{Depth}}(f_1) \cdot \mathrm{Adv}_{0}^{\mathrm{Depth}}(f_2)$. (From the point of view of a tree which makes no queries the outputs of f_1 and f_2 are *independent* random variables.) To bound $\mathrm{Adv}_{d_1,d_2}^{\mathrm{FairDepth}}(f_1 \oplus f_2)$ for $d_1 + d_2 > 0$, let T be a decision tree which achieves this advantage. Without loss of generality the first query of T is from x_1. We will use the notation $x_1 = (y, b)$ where b is the bit queried by T and y is the remaining bits. We denote the two subtrees of T by T^0 and T^1 respectively. For $b \in \{0, 1\}$, we define functions $g^b(y) = f_1((y, b))$. We now have

$$\mathrm{Adv}_{d_1,d_2}^{\mathrm{FairDepth}}(f_1 \oplus f_2) = \sum_{b \in \{0,1\}} \frac{1}{2} \mathrm{Adv}^{T_b}(g^b \oplus f_2).$$

Here, $\mathrm{Adv}^T(f)$ is used to denote the advantage of T on f. However, T^0 and T^1 are of depth $d_1 - 1 + d_2$, and are in $\mathrm{FairDepth}_{d_1-1,d_2}$. Applying our induction hypothesis we find that the above is

$$\begin{aligned} &\le \sum_{b \in \{0,1\}} \frac{1}{2} \mathrm{Adv}_{d_1-1,d_2}^{\mathrm{FairDepth}}(g^b \oplus f_2) \\ &\le \mathrm{Adv}_{d_2}^{\mathrm{Depth}}(f_2) \sum_{b \in \{0,1\}} \frac{1}{2} \cdot \mathrm{Adv}_{d_1-1}^{\mathrm{Depth}}(g^b). \end{aligned}$$

Consider trees P^0, P^1 which achieve the advantage on g^0, g^1. We now construct a tree P of depth d_1 which starts by querying b and depending on the outcome activates P^b. We then have

$$\sum_{b \in \{0,1\}} \frac{1}{2} \mathrm{Adv}_{d_1-1}^{\mathrm{Depth}}(g^b) = \mathrm{Adv}^{p}(f_1) \le \mathrm{Adv}_{d_1}^{\mathrm{Depth}}(f_1).$$

The equality follows from the definition of P, and the inequality from the fact that P is of depth d_1. Plugging this in our previous calculation we prove the lemma. □

6. Discussion and open problems

We have suggested two approaches for obtaining direct product assertions. The first is to strengthen the assumption and prove the assertion only for functions f with a certain property (as we did for communication games using low discrepancy). The second approach is to weaken the conclusion and prove the assertion only for restricted algorithms (as we did for fair decision trees).

The most interesting open problem is to extend these approaches to other computational models. There are also some interesting concrete open problems regarding communication games. For example, can the notion of discrepancy be altered to match other probability distributions on the inputs of the two players (not just the uniform distribution) while still enabling a strong direct product assertion? Another problem is whether the approach of this paper can be extended to one-sided discrepancy, which removes the absolute value in Definition 4.1. These two problems are motivated by Razborov's proof (Razborov 1992) of the lower bound on the disjointness function (Babai *et al.* 1986; Kalyanasundaram & Schnitger 1992) which uses one-sided discrepancy in a non-uniform distribution.

More generally, an interesting problem is whether the notion of "fairness" suggested here can be used for other models of computation. A natural place to start is communication protocols or even one-round communication protocols. In the following we discuss how to generalize the definition of fairness to other computational models.

6.1. Fairness for general models of computation. While it is easy to define fairness for decision trees, how does one define a fair communication protocol or a fair computation in general? Before making our suggestions, let us examine what is the role of fairness in the argument. Recall the information theoretic analogue presented at the beginning of the introduction. The goal there is to predict the outcome of an exclusive-or of k independent biased coins. The direct product question is the analogue of this question in the sense that from of the point of view of any algorithm in $\mathrm{Res}_{r'}$ the function (or more precisely the exclusive-or of the outcome of the function with the answer of the algorithm) is a biased coin. What makes the computational version difficult is that the algorithm may make these biased coins correlated. However, we still want to have that each individual coin is not fully determined. In other words, that the algorithm attempting to compute $f^{\oplus k}$ cannot compute any of the individual outcomes of $f(x_i)$. When $r' \leq r$ (as is the case in Yao's XOR-lemma), this happens simply because the protocol attempting to compute $f^{\oplus k}$ is "small" enough to be bounded by the hypothesis. The role of the fairness restriction is to ensure this situation when $r' > r$. Having this in mind we can now suggest two approaches to impose a fairness restriction on general models of computation.

6.1.1. A syntactic approach. We want that a protocol using $r' > r$ units of the resource will not be able to compute any individual outcome of $(f(x_1), \ldots, f(x_k))$ too well. In fair decision trees this is guaranteed because once you fix $k-1$ inputs of a d-fair decision tree, the induced tree is of depth d, and is thus bounded by the hypothesis of the direct product assertion. This leads us to the following definition of syntactic fairness for a general model of computation.

Definition 6.1. An algorithm $P' \in \mathrm{Res}_{r'}$ with inputs $x_1, \ldots, x_k$ is *r-fair* if for every $1 \le i \le k$ and every fixing $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k$ for the variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k$, algorithm $P'(a_1, \ldots, a_{i-1}, x, a_{i+1}, \ldots, a_k)$ can be simulated by an algorithm $P(x)$ in Res_r.

Syntactic fairness seems a very strong restriction. However, at this point we do not know whether syntactic fairness gives a strong direct product assertion in communication protocols.

6.1.2. A semantic approach. The intuition above is that syntactic fairness guarantees that the algorithm cannot compute any individual outcome too well. Consider the concatenation variant of the direct product question defined in Section 1.1.4. In this variant the algorithm P' is required to give outputs for all k inputs. We can replace the syntactic approach above by imposing that for every $1 \le i \le k$, the algorithm P' succeeds on x_i with probability at most p.

Definition 6.2. An algorithm $P' \in \mathrm{Res}_{r'}$ with inputs $x_1, \ldots, x_k$ and outputs in $\{0,1\}^k$ is (r,p)-*semantically fair* for f if for every $1 \le i \le k$,

$$\Pr_{x_1,\ldots,x_k}[P'_i(x_1, \ldots, x_k) = f(x_i)] \le p.$$

Syntactic fairness automatically implies semantic fairness. However, semantic fairness is far less restricting.

Acknowledgements

I would like to thank Avi Wigderson for introducing me to this area and for his endless guidance and support. I would like to thank Russell Impagliazzo and Salil Vadhan for helpful discussions. I am also grateful to anonymous referees for pointing out some relevant previous work and helpful comments.

This research was performed in part at Hebrew University and Institute for Advanced Study. It was supported in part by the Koshland Scholar program.

References

L. Babai, P. Frankl & J. Simon (1986). Complexity classes in communication complexity theory (preliminary version). In *27th Annual Symposium on Foundations of Computer Science* (Toronto), IEEE, 337–347.

B. Chor & O. Goldreich (1988). Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.* **17**, 230–261.

U. Feige & O. Verbitsky (2002). Error reduction by parallel repetition—a negative result. *Combinatiorica* **22**, 461–478.

O. Goldreich & L. A. Levin (1989). A hard-core predicate for all one-way functions. In *Proc. 21st Annual ACM Symposium on Theory of Computing* (Seattle, WA), 25–32.

O. Goldreich, N. Nisan & A. Wigderson (1995). On Yao's XOR-lemma. Technical report, Electronic Colloquium on Computational Complexity. `http://www.eccc.uni-trier.de/eccc`.

R. Impagliazzo (1995). Hard-core distributions for somewhat hard problems. In *36th Annual Symposium on Foundations of Computer Science* (Milwaukee, WI), IEEE, 538–545.

R. Impagliazzo, R. Raz & A. Wigderson (1994). A direct product theorem. In *9th Structure in Complexity Theory Conference*, 88–96.

R. Impagliazzo & A. Wigderson (1997). $P = BPP$ if E requires exponential circuits: derandomizing the XOR lemma. In *Proc. 29th Annual ACM Symposium on Theory of Computing* (El Paso, TX), 220–229.

B. Kalyanasundaram & G. Schnitger (1992). The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.* **5**, 545–557.

M. Karchmer, R. Raz & A. Wigderson (1995). Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Comput. Complexity* **5**, 191–204.

E. Kushilevitz & N. Nisan (1997). *Communication Complexity*. Cambridge Univ. Press.

L. A. Levin (1987). One way functions and pseudorandom generators. *Combinatorica* **7**, 357–363.

N. Nisan, S. Rudich & M. Saks (1999). Products and help bits in decision trees. *SIAM J. Comput.* **28**, 1035–1050.

N. Nisan & A. Wigderson (1995). On rank vs. communication complexity. *Combinatorica* **15**, 557–565.

I. Parnafes, R. Raz & A. Wigderson (1997). Direct product results and the GCD problem, in old and new communication models. In *Proc. 29th Annual ACM Symposium on Theory of Computing* (El Paso, TX), 363–372.

R. Raz (1998). A parallel repetition theorem. *SIAM J. Comput.* **27**, 763–803.

A. A. Razborov (1992). On the distributional complexity of disjointness. *Theoret. Comput. Sci.* **106**, 385–390.

D. Uhlig (1974). On the synthesis of self-correcting schemes from functional elements with a small number of reliable elements. *Math. Notes Acad. Sci. USSR* **15**, 558–562.

A. C. Yao (1982). Theory and applications of trapdoor functions (extended abstract). In *23rd Annual Symposium on Foundations of Computer Science* (Chicago, IL), IEEE, 80–91.

Manuscript received 10 July 2003

Ronen Shaltiel
Department of Applied Mathematics
and Computer Science
Weizmann Institute of Science
Rehovot, 76100, Israel
ronens@wisdom.weizmann.ac.il
http://www.wisdom.weizmann.ac.il/~ronens