Constructive Proofs of Concentration Bounds

Russell Impagliazzo*
Institute for Advanced Study & University of California, San Diego
russell@cs.ucsd.edu

Valentine Kabanets†
Institute for Advanced Study & Simon Fraser University
kabanets@cs.sfu.ca

August 24, 2010

Abstract

We give a combinatorial proof of the Chernoff-Hoeffding concentration bound [Che52, Hoe63], which says that the sum of independent $\{0,1\}$-valued random variables is highly concentrated around the expected value. Unlike the standard proofs, our proof does not use the method of higher moments, but rather uses a simple and intuitive counting argument. In addition, our proof is constructive in the following sense: if the sum of the given random variables is not concentrated around the expectation, then we can efficiently find (with high probability) a subset of the random variables that are statistically dependent. As simple corollaries, we also get the concentration bounds for $[0,1]$-valued random variables and Azuma's inequality for martingales [Azu67].

We interpret the Chernoff-Hoeffding bound as a statement about Direct Product Theorems. Informally, a Direct Product Theorem says that the complexity of solving all k instances of a hard problem increases exponentially with k; a Threshold Direct Product Theorem says that it is exponentially hard in k to solve even a significant fraction of the given k instances of a hard problem. We show the equivalence between optimal Direct Product Theorems and optimal Threshold Direct Product Theorems. As an application of this connection, we get the Chernoff bound for expander walks [Gil98] from the (simpler to prove) hitting property [AKS87], as well as an optimal (in a certain range of parameters) Threshold Direct Product Theorem for weakly verifiable puzzles from the optimal Direct Product Theorem [CHS05]. We also get a simple constructive proof of Unger's result [Ung09] saying that XOR Lemmas imply Threshold Direct Product Theorems.

Keywords: Concentration bounds, Chernoff-Hoeffding bound, Azuma's inequality, expander walks, Direct Product Theorems, Threshold Direct Product Theorems, XOR Lemmas

1 Introduction

Randomized algorithms and random constructions have become common objects of study in modern computer science. Equally ubiquitous are the basic tools of probability theory used for their analysis. Some of the most widely used such tools are various *concentration bounds*. Informally, these are statements saying that the outcome of a random experiment is likely to be close to what is expected (concentrated near the expectation). The well-known Chernoff bound [Che52] is a prime example, and is probably one of the most-often used such concentration bounds. Basically, it says that repeating a random experiment many times independently and taking the average of the outcomes results in a value that is extremely likely to be very close to the expected outcome of the experiment, with the probability of deviation diminishing exponentially fast with the number of repetitions.

*Supported by NSF grants DMS-0835373, CNS-0716790, CCF-0832797, DMS-0635607, Simonyi Foundation, the Bell Company Fellowship, and Fund for Math.

†Supported by an NSERC Discovery grant and NSF grant DMS-0635607.

A computational analogue of concentration bounds in complexity are *Direct Product* Theorems. Informally, these are statements saying that solving a somewhat hard problem on many independent random instances becomes extremely hard, with the hardness growing at an exponential rate with the number of repetitions. The main application of direct product theorems is to hardness amplification: taking a problem that is somewhat hard-on-average to solve, and turning it into a problem that is extremely hard-on-average to solve. Such hardness amplification is important for cryptography and complexity; for example, in cryptography, the increased hardness of a function translates into the increased security of a cryptographic protocol.

In this paper, we show a close connection between probability-theoretic and complexity-theoretic concentration bounds. We give a new, constructive proof of the Chernoff bound, and use this proof to establish an equivalence between two versions of direct product theorems: the standard Direct Product Theorem and the *Threshold* Direct Product. In the standard direct product, we want to upperbound the probability of efficiently solving *all* given instances of a somewhat hard problem, whereas in the threshold direct product, we want to upperbound the probability of solving more than a certain *fraction* of the instances.

To motivate the need for Threshold Direct Product Theorems, we give an example of its typical use in cryptography. CAPTCHAs [ABHL03] are now widely used to distinguish human users from artificially intelligent "bots". Here a user is issued a random puzzle, say distorted text, and is asked to decipher the text. Say that a legitimate user succeeds with probability $c \leqslant 1$, whereas an attacker succeeds with probability at most $s < c$. To boost our confidence that we are dealing with a legitimate user, we will issue k random puzzles in parallel, and see how many of them get answered correctly. If $c = 1$, then we know that the legitimate user will answer all k instances correctly. A standard Direct Product Theorem for CAPTCHAs [BIN97, CHS05] could then be used to argue that it's very unlikely that an attacker will answer *all* k instances. In reality, however, even a legitimate user can make an occasional mistake, and so $c < 1$. Thus we can't distinguish between legitimate users and attackers by checking if all k instances are answered correctly. Intuitively, though, we still expect that a legitimate user should answer almost all instances (close to c fraction), whereas the attacker can't answer significantly more than s fraction of them. This intuition is formalized in the Threshold Direct Product Theorem for CAPTCHAs [IJK09], which thus allows us to make CAPTCHAs reliably easy for humans but reliably hard for "bots".

The probability-theoretic analogue of a Direct Product Theorem is the statement that if a random experiment succeeds with probability at most p, then the probability that it succeeds in k independent trials is at most p^k. The analogue of a Threshold Direct Product is the Chernoff bound saying that the probability of getting significantly more than the expected pk successes is exponentially small in k. We give a *constructive* proof of the equivalence between these two probability-theoretic statements. Namely, we show that if the probability of getting more than pk successes is *noticeably larger* than it should be (by the Chernoff bound), then we can *efficiently* find a subset S of the k trials such that the random experiment succeeds in all trials $i \in S$ with probability *noticeably larger* than $p^{|S|}$.

In the language of direct products, this means that there is an equivalence between standard direct product theorems and threshold direct product theorems. Moreover, the constructive nature of the proof of this equivalence means that it applies to the *uniform* setting of computation, where the hardness (security) is measured with respect to uniform algorithms (rather than non-uniform circuits). In particular, we get that for a wide variety of classes of cryptographic protocols, there is a Direct Product Theorem for the class iff there is a Threshold Direct Product theorem.

The formalized equivalence between standard and threshold direct products also allows us to quantify the information-theoretic limitations of simple reductions between the two. We then show how to overcome this limitation with slightly more complicated reductions (using conditioning).

1.1 Chernoff-Hoeffding bounds, martingales and expander walks

The well-known Chernoff-Hoeffding bound [Che52, Hoe63] states that the sum of independent $\{0,1\}$-valued random variables is highly concentrated around the expected value. Numerous variants of this concentration bound have been proved, with Bernstein's inequalities from 1920's and 1930's being probably the earliest [Ber64]. The known proofs of these bounds rely on the idea of Bernstein to use the moment-generating function of the given sum of independent random variables $X_1 + \cdots + X_n$; recall that the moment-generating

function of a random variables X is $M_X(t) = \mathbf{Exp}[e^{t \cdot X}]$, where $\mathbf{Exp}[\cdot]$ denotes the expectation.

While not difficult technically, the standard proof, in our opinion, does not provide intuition why concentration is likely. One of the main results of our paper is a different proof of the Chernoff bound, using a simple combinatorial argument (and, in particular, avoiding any use of the moment-generating functions). We actually prove a generalization of the Chernoff bound, originally due to Panconesi and Srinivasan [PS97] (who also used the standard method of moment-generating functions in their proof). In this generalization, the assumption of independence of the variables $X_1, \ldots, X_n$ is replaced with the following weaker assumption: There exists some $\delta > 0$ such that, for all subsets $S \subseteq [n]$ of indices, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \delta^{|S|}$. Observe that if the variables X_i's are independent, with each $\mathbf{Exp}[X_i] \leqslant \delta$, then, for all $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \delta^{|S|}$.

Theorem 1.1 (Generalized Chernoff bound [PS97])**.** *Let $X_1, \ldots, X_n$ be Boolean random variables such that, for some $0 \leqslant \delta \leqslant 1$, we have that, for every subset $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \delta^{|S|}$. Then, for any $0 \leqslant \delta \leqslant \gamma \leqslant 1$, $\mathbf{Pr}\left[\sum_{i=1}^{n} X_i \geqslant \gamma n\right] \leqslant e^{-nD(\gamma \| \delta)}$, where $D(\cdot \parallel \cdot)$ is the relative entropy function (defined in Section 2 below), satisfying $D(\gamma \parallel \delta) \geqslant 2(\gamma - \delta)^2$.*

We now sketch our proof of Theorem 1.1. Imagine sampling a random subset $S \subseteq [n]$ where each index $i \in [n]$ is put in S independently with some probability q (to be optimally chosen). We compute, in two ways, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1]$, where the probability is over S and $X_1, \ldots, X_n$.

On one hand, since $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \delta^{|S|}$ for *all* $S \subseteq [n]$, the probability of choosing $S \subseteq [n]$ with $\wedge_{i \in S} X_i = 1$ is *small*. On the other hand, if $p = \mathbf{Pr}[\sum_{i=1}^{n} X_i \geqslant \gamma n]$ is relatively large, we are likely to sample a n-tuple $X_1, \ldots, X_n$ with very many (at least γn) 1's. Given such a tuple, we are then likely to sample a subset $S \subseteq [n]$ with $\wedge_{i \in S} X_i = 1$. Thus the overall probability of choosing $S \subseteq [n]$ with $\wedge_{i \in S} X_i = 1$ is relatively *large*. The resulting contradiction shows that p must be small. (The complete proof is given in Section 3.1.)

We also get several other concentration bounds as simple corollaries of Theorem 1.1. First, we get a version of Theorem 1.1 in the setting of real-valued random variables that take their values in the interval $[0, 1]$, the Hoeffding bound [Hoe63] (Theorem 3.3). Then we prove a concentration bound for martingales, known as Azuma's inequality [Azu67] (Theorem 3.4). In another application of our Theorem 1.1, we obtain a Chernoff-type concentration bound for random walks on expander graphs (Theorem 3.8), almost matching the parameters of [Gil98, Hea08].

1.2 Applications to Direct Product Theorems

We interpret Theorem 1.1 as giving an equivalence between certain versions of Direct Product Theorems (DPTs), which are statements of the form "k-wise parallel repetition increases the complexity of a problem at an exponential rate in the number of repetitions k". Such theorems are known for a variety of models: Boolean circuits [Yao82, GNW95], 2-prover games [Raz98], decision trees [NRS94], communication complexity [PRW97], polynomials [VW08], puzzles [BIN97], and quantum XOR games [CSUU07], just to mention a few. However, there are also examples where a direct product statement is false (see, e.g., [BIN97, PW07, Sha03]).

More formally, for a function $F \colon U \to R$, its k-wise direct product is the function $F^k \colon U^k \to R^k$, where $F^k(x_1, \ldots, x_k) = (F(x_1), \ldots, F(x_k))$. The main application of this construction is to *hardness amplification*. Intuitively, if $F(x)$ is easy to compute on at most p fraction of inputs x (by a certain resource-bounded class of algorithms), then we expect $F^k(x_1, \ldots, x_k)$ to be easy on at most (close to) p^k fraction of k-tuples $(x_1, \ldots, x_k)$ (for a related class of algorithms).

A DPT may be viewed as a computational analogue of the following (obvious) probabilistic statement: Given k random independent Boolean variables $X_1, \ldots, X_k$, where each $X_i = 1$ with probability at most p, we have $\mathbf{Pr}[\wedge_{i=1}^{k} X_i = 1] \leqslant p^k$. The Chernoff bound says that with all but exponentially small probability at most about pk of the random variables $X_1, \ldots, X_k$ will be 1. The computational analogue of this concentration bound is often called a *Threshold Direct Product Theorem (TDPT)*, saying that if a function F is easy to compute on at most p fraction of inputs (by a certain class of algorithms), then computing $F^k(x_1, \ldots, x_k)$ correctly in significantly more than pk positions $1 \leqslant i \leqslant k$ is possible for at most a (negligibly more than) exponentially small fraction of k-tuples $(x_1, \ldots, x_k)$ (for a related class of algorithms). TDPTs are also known for a number of models, e.g., Boolean circuits (follows from [Imp95, Hol05]), 2-prover games [Rao08], puzzles [IJK09], and quantum XOR games [CSUU07].

Observe that Theorem 1.1 says that the Chernoff concentration bound for random variables $X_1, \ldots, X_n$ follows from the assumption that $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant p^{|S|}$ for all subsets S of $[n]$. In the language of direct products, this means that Threshold Direct Product Theorems follow from Direct Product Theorems. We explain this connection in more detail next.

1.2.1 Equivalence between DPTs and TDPTs

Let us call a DPT *optimal* if has perfect exponential increase in complexity: A function F that is computable on at most p fraction of inputs gives rise to the function F^k that is computable on at most p^k fraction of inputs. Similarly, we call a TDPT optimal, if its parameters match exactly its probabilistic analogue, the Chernoff-Hoeffding bound.

As an immediate application of Theorem 1.1, we get that an optimal DPT implies an optimal TDPT. We illustrate it for the case of the DPT for Boolean circuits. Suppose F is a Boolean function that can be computed on at most p fraction of inputs (by circuits of certain size s). The optimal DPT for circuits (provable, e.g., using [Imp95, Hol05]) says that for any k, the function F^k is computable on at most p^k fraction of inputs (by any circuit of appropriate size $s' < s$).

Towards a contradiction, suppose there is an algorithm A that computes $F^k(x_1, \ldots, x_k)$ in significantly more than pk positions $1 \leqslant i \leqslant k$, for more than the exponentially small fraction of inputs $(x_1, \ldots, x_k)$. Define Boolean random variables $X_1, \ldots, X_k$, dependent on F, A, and a random k-tuple $(x_1, \ldots, x_k)$, so that $X_i = 1$ iff $A(x_1, \ldots, x_k)_i = F(x_i)$. By our assumption, these variables $X_1, \ldots, X_k$ fail the Chernoff concentration bound. Hence, by Theorem 1.1, there is a subset $S \subseteq \{1, \ldots, k\}$ such that $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] > p^{|S|}$. But the latter means that our algorithm A, restricted to the positions $i \in S$, computes $F^{|S|}$ with probability greater than $p^{|S|}$, contradicting the optimal DPT.

In an analogous way, we get an optimal TDPT for every non-uniform model where an optimal DPT is known: e.g., decision trees [NRS94] and quantum XOR games [CSUU07]; for the latter model, an optimal TDPT was already proved in [CSUU07].

1.2.2 A constructive version of Theorem 1.1

For non-uniform models (as in the example of Boolean circuits considered above), it suffices to use Theorem 1.1 which only says that if the random variables $X_1, \ldots, X_n$ fail to satisfy the concentration bound, then there *must exist* a subset S of them such that $\wedge_{i \in S} X_i = 1$ with large probability. To obtain the Direct Product Theorems in the uniform model of computation, it is important that such a subset S be efficiently computable by a *uniform* algorithm.

Our combinatorial proof of Theorem 1.1 immediately yields such an algorithm. Namely, we just randomly sample a subset S by including each index i, $1 \leqslant i \leqslant n$, into S with probability q, where q is chosen as a function of how far the variables $X_1, \ldots, X_n$ are from satisfying the concentration bound. We then output S if $\wedge_{i \in S} X_i = 1$ has "high" probability; otherwise we sample another set S. Here we assume that our algorithm has a way to sample from the distribution $X_1, ..., X_n$. This reasoning yields the following.

Theorem 1.2. *There is a randomized algorithm $\mathcal{A}$ such that the following holds. Let $X_1, \ldots, X_n$ be 0-1-valued random variables. Let $0 < \delta < \gamma \leqslant 1$ be such that $\mathbf{Pr}[\sum_{i=1}^n X_i \geqslant \gamma n] = p > 2\alpha$, for some $\alpha \geqslant e^{-nD(\gamma \| \delta)}$. Then, on inputs $n, \gamma, \delta, \alpha$, the algorithm $\mathcal{A}$, using oracle access to the distribution $X_1, \ldots, X_n$, runs in time $\mathrm{poly}(\alpha^{-1/((\gamma-\delta)\delta)}, n)$ and outputs a set $S \subseteq [n]$ such that, with probability at least $1 - o(1)$, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] > \delta^{|S|} + \Omega(\alpha^{4/((\gamma-\delta)\delta)})$.*

Using this constructive version, we prove an optimal TDPT also for uniform models. In particular, we get such a result for the case of CAPTCHA-like puzzles, called weakly verifiable puzzles [CHS05] (see Theorem 4.2).[1] DPTs for puzzles are known [BIN97, CHS05], with [CHS05] giving an optimal DPT. Also TDPTs are known [IJK09, Jut10], but they are not optimal. Here we immediately get an optimal TDPT for puzzles, using the optimal DPT of [CHS05], when the success probabilities of the legitimate user and the attacker are constant.

[1] Unger [Ung09] claims to get a TDPT for puzzles, but in fact only proves a TDPT for circuits from Yao's XOR Lemma. Actually, no XOR Lemma for puzzles is known, and so Unger's methods don't apply.

We also show that the limitation on the success probabilities being constant is *unavoidable* for the *naive* reductions between DPTs and TDPTs, as those in Theorem 1.2. Namely, we give an example of a distribution $X_1, \ldots, X_n$ where the dependence on $\gamma - \delta$ in the exponent of α (stated in Theorem 1.2) is necessary.

Lemma 1.3. *There are Boolean random variables $X_1, \ldots, X_n$, and parameters $0 < \delta < \gamma < 1$ such that $\mathbf{Pr}[\sum_{i=1}^{n} X_i \geqslant \gamma n] = p/2 > 2\alpha$, for $\alpha \geqslant e^{-nD(\gamma\|\delta)}$, but, for every subset $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i\in S} X_i = 1] - \delta^{|S|} \leqslant (4\alpha)^{\delta(\ln 1/\delta)/(\gamma-\delta)}$.*

We also show that this limitation can be overcome by *conditioned* reductions which are allowed to use conditioning (albeit the concentration bound we get in this case is not as tight as before).

Theorem 1.4. *There is a randomized algorithm $\mathcal{A}$ satisfying the following. Let $X_1, \ldots, X_n$ be Boolean-valued random variables, and let $0 \leqslant \delta < \gamma \leqslant 1$. Suppose that $\mathbf{Pr}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geqslant \gamma\right] > \alpha$, where $\alpha > (32/(\gamma-\delta)) \cdot e^{-(\gamma-\delta)^2 n/64}$. Then, the algorithm $\mathcal{A}$ on inputs $n, \gamma, \delta, \alpha$, using oracle access to the conditional distribution $(X_1, \ldots, X_n \mid \sum_{j\in S} X_j \geqslant \gamma n/2)$, runs in time $\mathrm{poly}(n, 1/\alpha, 1/\gamma, 1/\delta)$ and outputs a subset $S \subset [n]$ (of size $n/2$) and an index $i_0 \in \bar{S}$ (where $\bar{S} = [n] - S$) such that, with probability at least $1 - o(1)$, $\mathbf{Pr}\left[X_{i_0} = 1 \mid \sum_{j\in S} X_j \geqslant \gamma n/2\right] > \delta + (\gamma - \delta)/16$.*

Remark 1.5. *Naive reductions between DPT and TDPT (as in Theorem 1.2) are applicable in any setting, whereas conditioned reductions (as in Theorem 1.4) need an additional assumption (sampleability from a conditional distribution). The universality of naive reductions, however, comes at an unavoidable cost, as witnessed by Lemma 1.3. Together with Theorem 1.4, this shows that there is an actual quantitative difference between naive and conditioned reductions. In particular, while, surprisingly, naive reductions are optimal in terms of quantitative hardness amplification, they are suboptimal in terms of preserving the adversary's advantage, which is only polynomailly preserved if $\gamma - \delta = \Omega(1)$. In contrast, conditioned reductions can preserve this advantage linearly.*

Finally, our Theorem 1.1 implies some TDPT even when we only have a weak (suboptimal) DPT for the model. For example, we can get some version of a TDPT for 2-prover games, using the best available DPT for such games [Raz98, Hol07, Rao08];[2] however, a better TDPT for 2-prover games is known [Rao08]. Also, as shown by Haitner [Hai09], for a wide class of cryptographic protocols (interactive arguments), even if the original protocol doesn't satisfy any DPT, there is a slight modification of the protocol satisfying some weak DPT. Then, our results imply that these modified protocols also satisfy some weak TDPT.

1.2.3 Direct Product Theorems vs. XOR Lemmas

A close relative of DPTs is an XOR Theorem. For a Boolean function $F\colon \{0,1\}^n \to \{0,1\}$, its k-wise XOR function is $F^{\oplus k}\colon (\{0,1\}^n)^k \to \{0,1\}$, where $F^{\oplus k}(x_1, \ldots, x_k) = \oplus_{i=1}^{k} F(x_i)$. Intuitively, taking XOR of the k independent copies of a function F, where F can be computed on at most p fraction of inputs, is similar to taking the XOR of k independent random Boolean variables $X_1, \ldots, X_k$, where each $X_i = 1$ with probability at most p. In the latter case, it is easy to compute that $\mathbf{Pr}[\oplus_{i=1}^{k} X_i = 1] \leqslant 1/2 + (2p-1)^k/2$, i.e., the k-wise XOR approaches a fair coin flip exponentially fast in k. In the computational setting, one would like to argue that $F^{\oplus k}$ becomes essentially unpredictable. Such XOR results are also known, the most famous being Yao's XOR Lemma for Boolean circuits [Yao82, Lev87, GNW95] (many proofs of this lemma have been given over the years, see, e.g., [IJKW10] for the most recent proof, and the references).

We call an XOR lemma *optimal* if its parameters exactly match the probabilistic analogue given above. Recently, Unger [Ung09] essentially showed that an optimal XOR result implies an optimal TDPT (and hence also an optimal DPT). More precisely, he proved the following generalization of the Chernoff-Hoeffding bound: Let $X_1, \ldots, X_k$ be Boolean random variables such that for some $-1 \leqslant \beta \leqslant 1$, we have that, for every subset $S \subseteq \{1, \ldots, k\}$, $\mathbf{Pr}[\oplus_{i\in S} X_i = 1] \leqslant 1/2 + \beta^{|S|}/2$. Then for any $\beta \leqslant \rho \leqslant 1$, $\mathbf{Pr}[\sum_{i=1}^{k} X_i \geqslant (1/2 + \rho/2)k] \leqslant e^{-kD(1/2+\rho/2\|1/2+\beta/2)}$.

Unger's original proof uses the method of moment-generating functions and some basic tools from Fourier analysis. In contrast, we give a simple reduction showing that the assumption in Unger's theorem implies

[2] In fact, for 2-prover games, it is impossible to achieve the "optimal" decrease in the success probability from p to p^k, for k parallel repetitions of the game [Raz08].

the assumption in Theorem 1.1, and thus we immediately get an alternative (and simpler) proof of Unger's result. For a random variable $X \in \{0,1\}$, we define $bias(X) = \mathbf{Pr}[X = 0] - \mathbf{Pr}[X = 1]$. We show the following.

Theorem 1.6. *Let $X_1, \ldots, X_n$ be 0-1-valued random variables. Suppose that there is $-1 \leqslant \beta \leqslant 1$ such that, for every $S \subseteq [n]$, $bias(\oplus_{i \in S} X_i) \leqslant \beta^{|S|}$. Then, for every $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i \in S}(X_i = 0)] \leqslant (1/2 + \beta/2)^{|S|}$.*

Moreover, the reduction in the proof of Theorem 1.6 is constructive. Combining it with the constructive version of Theorem 1.1, we get a *constructive* version of Unger's result: if the variables $X_1, \ldots, X_n$ fail to satisfy the concentration bound, then we can efficiently find (using a randomized algorithm) a subset S of indices such that $\oplus_{i \in S} X_i$ has "large" bias. Such a constructive version is not implied by the original proof of [Ung09].

1.3 Related work

1.3.1 Chernoff bounds for negatively correlated random variables

The assumption on the random variables $X_1, \ldots, X_n$ used in Theorem 1.1 is similar to the assumption that the X_i's are *negatively correlated*; the latter means that for every subset $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \prod_{i \in S} \mathbf{Pr}[X_i = 1]$. The only difference between the negative correlation assumption and the assumption in Theorem 1.1 is that the latter upperbounds $\mathbf{Pr}[\wedge_{i \in S} X_i = 1]$ by some $\delta^{|S|}$, where δ is an upper bound on $\mathbf{Pr}[X_i = 1]$. Panconesi and Srinivasan [PS97] observed that the Chernoff-Hoeffding bound continues to hold for the case of random variables that satisfy this generalized version of negative correlation. The proof in [PS97] follows the standard, Bernstein-style, proof of the Chernoff-Hoeffding bound.

1.3.2 TDPTs from DPTs, and DPTs from XOR lemmas

A simple idea for converting DPTs into TDPTs by randomly sampling a subset of a given n-tuple of instances was also suggested by Ben-Aroya et al. [BARW08, Theorem 10], but their reduction doesn't give the optimal parameters. In the setting of interactive protocols, Chung and Liu [CL10] show how to obtain an almost-optimal TDPT from an optimal DPT, also using a very similar sampling-based argument. The fact that XOR Lemma implies DPT was also shown by Viola and Wigderson [VW08, Proposition 1.4]. Our proof of Theorem 1.6 (showing that optimal XOR Lemma implies optimal DPT) is a very similar argument.

While the idea of using sampling to get weak versions of TDPTs from DPTs has been used in earlier works, the difference in our paper is to use it in the *abstract setting* of probability-theoretic concentration bounds, and achieve *tight parameters*. It is actually surprising that such a simple idea is powerful enough to yield tight concentration bounds. The advantage of the abstract framework is also that it suggests applications in settings where one doesn't usually think in terms of standard direct products and threshold direct products. For example, we use our Theorem 1.1 to prove the Chernoff concentration bound for expander walks [Gil98] from the hitting property of [AKS87]. We also show the *information-theoretic limitations* of simple reductions between DPTs and TDPTs, and suggest a way to overcome these limitations with stronger reductions.

We consider the new proof of Chernoff-type concentration bounds more revealing and intuitive than the standard Bernstein-style proofs, and hope that its constructiveness will have other applications in computer science.

2 Preliminaries

For a natural number n, we denote by $[n]$ the set $\{1, 2, \ldots, n\}$. For $0 \leqslant \rho, \sigma \leqslant 1$, let $D(\rho \parallel \sigma)$ be the binary relative entropy defined as $D(\rho \parallel \sigma) = \rho \ln \frac{\rho}{\sigma} + (1 - \rho) \ln \frac{1-\rho}{1-\sigma}$, with $0 \ln 0 = 0$. We shall also use the following simple estimate: $D(\sigma + \epsilon \parallel \sigma) \geqslant 2\epsilon^2$ (obtained by considering the Taylor expansion of the function $g(x) = D(p + x \parallel p)$ up to the second derivative).

For parameters $0 \leqslant \delta \leqslant \gamma \leqslant 1$, we define the function $f_{\delta,\gamma}(q) = \frac{1-q(1-\delta)}{(1-q)^{1-\gamma}}$; we shall be interested in the case where $0 \leqslant q < 1$. When δ, γ are clear from the context, we drop the subscripts and simply write $f(q)$.

Taking the derivative of the function $f(q)$, we get that $f(q)$ achieves its minimum at $q^* = \frac{\gamma-\delta}{\gamma(1-\delta)}$. It is easy to see that $f(q^*) = \left(\frac{\delta}{\gamma}\right)^{\gamma} \left(\frac{1-\delta}{1-\gamma}\right)^{1-\gamma} = e^{-D(\gamma\|\delta)}$.

For parameters $n \in \mathbb{N}$ and $0 \leqslant q \leqslant 1$, we denote by $Bin(n,q)$ the *binomial distribution* on sets $S \subseteq [n]$, where a set S is obtained by picking each index $1 \leqslant i \leqslant n$, independently, with probability q. We will denote by $S \sim Bin(n,q)$ the random choice of $S \subseteq [n]$ according to $Bin(n,q)$.

We use the following "mean is median" result of Jogdeo and Samuels [JS68] for general binomial distributions (where the probabilities of choosing an index i may be different for different i's).

Lemma 2.1 ([JS68])**.** *For every n-tuple of real numbers $p_1, \ldots, p_n$, $0 \leqslant p_i \leqslant 1$ for all $1 \leqslant i \leqslant n$, and for the Boolean random variables $X_1, \ldots, X_n$ where each $X_i = 1$ with probability p_i, and $X_i = 0$ with probability $1 - p_i$, let $S = \sum_{i=1}^n X_i$ and let $\mu = \sum_{i=1}^n p_i$. Then the median of the distribution S is either $\lfloor \mu \rfloor$ or $\lceil \mu \rceil$ (and is equal to μ if μ is an integer). In particular, we have $\mathbf{Pr}[S \geqslant \lfloor \mu \rfloor] \geqslant 1/2$.*

3 Concentration bounds

3.1 Boolean random variables

Theorem 1.1 is the special case of the following theorem (when $\delta_1 = \cdots = \delta_n$).

Theorem 3.1. *Let $X_1, \ldots, X_n$ be 0-1-valued random variables. Suppose that there are $0 \leqslant \delta_i \leqslant 1$, for $1 \leqslant i \leqslant n$, such that, for every set $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \prod_{i \in S} \delta_i$. Let $\delta = (1/n) \sum_{i=1}^n \delta_i$. Then, for any γ such that $\delta \leqslant \gamma \leqslant 1$, we have $\mathbf{Pr}[\sum_{i=1}^n X_i \geqslant \gamma n] \leqslant e^{-nD(\gamma\|\delta)}$.*

Proof. For a parameter $0 \leqslant q \leqslant 1$ to be chosen later, consider the following random experiment. Pick a random n-tuple $(x_1, \ldots, x_n)$ from the given distribution $X_1, \ldots, X_n$. Pick a set $S \sim Bin(n,q)$ (i.e., each position $1 \leqslant i \leqslant n$, independently, is in S with probability q).

Let $\mathcal{E}$ be the event that $\sum_{j=1}^n X_j \geqslant \gamma n$, and let $p = \mathbf{Pr}[\mathcal{E}]$. By conditioning,

$$\mathbf{Exp}[\wedge_{i \in S} X_i = 1] \geqslant \mathbf{Exp}[\wedge_{i \in S} X_i = 1 \mid \mathcal{E}] \cdot p, \tag{1}$$

where the expectations are over random choices of $S \sim Bin(n,q)$ and $X_1, \ldots, X_n$.

For every $S \subseteq [n]$, we have $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] \leqslant \prod_{i \in S} \delta_i$. Hence,

$$\mathbf{Exp}[\wedge_{i \in S} X_i = 1] \leqslant \sum_{S \subseteq [n]} \left[q^{|S|} (1-q)^{n-|S|} \prod_{i \in S} \delta_i \right]. \tag{2}$$

Let us denote by $(z_1, \ldots, z_n) \in \{0,1\}^n$ the characteristic vector of a set S chosen in the random experiment above. That is, each z_i is 1 with probability q, and 0 with probability $1-q$; all z_i's are independent. In this new notation, the expression in (2) equals $\mathbf{Exp}_{z_1,\ldots,z_n}[\prod_{i=1}^n \delta_i^{z_i}] = \prod_{i=1}^n \mathbf{Exp}_{z_i}[\delta_i^{z_i}] = \prod_{i=1}^n (q\delta_i + 1 - q)$, where the first equality is by the independence of the z_i's. By convexity, $(1/n) \sum_{i=1}^n \ln(q\delta_i + 1 - q) \leqslant \ln(q\delta + 1 - q)$, and hence $\prod_{i=1}^n (q\delta_i + 1 - q) \leqslant (q\delta + 1 - q)^n$. (When $\delta_1 = \cdots = \delta_n$, the same upper bound on the r.h.s. of (2) follows immediately from the binomial formula.)

On the other hand, $\mathbf{Exp}[\wedge_{i \in S} X_i = 1 \mid \mathcal{E}]$ is the probability that a random $S \sim Bin(n,q)$ misses all the 0 positions in the chosen sample from $X_1, \ldots, X_n$, conditioned on $\mathcal{E}$. Since there are at most $n - \gamma n$ such 0 positions, we get $\mathbf{Exp}[\wedge_{i \in S} X_i = 1 \mid \mathcal{E}] \geqslant (1-q)^{n-\gamma n}$. Combining this with Eqs. (1)–(2), we get $p \leqslant \left(\frac{q\delta + 1 - q}{(1-q)^{(1-\gamma)}}\right)^n = (f(q))^n$, where $f(q)$ is the function defined in Sect. 2 above. Choosing $q = q^*$ to minimize $f(q)$ (see Sect. 2), we get $p \leqslant e^{-nD(\gamma\|\delta)}$. □

Remark 3.2. *For $\gamma = 1$, Theorem 3.1 is tight, as $e^{-nD(1\|\delta)} = \delta^n$.*

3.2 Real-valued random variables, and martingales

We prove a version of Theorem 1.1 for the case of real-valued random variables.

Theorem 3.3. *Let $X_1, \ldots, X_n \in [0,1]$ be real-valued random variables. Suppose that there is a $0 \leqslant \delta \leqslant 1$ such that, for every set $S \subseteq [n]$, $\mathbf{Exp}\left[\prod_{i \in S} X_i\right] \leqslant \delta^{|S|}$. Then, for any γ such that $\delta \leqslant \gamma \leqslant 1$, $\mathbf{Pr}\left[\sum_{i=1}^n X_i \geqslant \lceil \gamma n \rceil\right] \leqslant 2 \cdot e^{-nD(\gamma \| \delta)}$.*

Proof. Let $p = \mathbf{Pr}\left[\sum_{i=1}^n X_i \geqslant \lceil \gamma n \rceil\right]$. Suppose that $p > 2 \cdot \exp(-nD(\gamma \parallel \delta))$. Our proof is by a reduction to the Boolean case. Consider Boolean random variables $Y_1, \ldots, Y_n$, where $\mathbf{Pr}[Y_i = 1] = X_i$, for all $1 \leqslant i \leqslant n$; that is, we think of the real value X_i as the probability that a Boolean variable Y_i is 1. Suppose we sample $x_1, \ldots, x_n$ from the distribution $X_1, \ldots, X_n$. Conditioned on $\sum_{i=1}^n x_i \geqslant \lceil \gamma n \rceil$, we have by Lemma 2.1 that $\mathbf{Pr}[\sum_{i=1}^n Y_i \geqslant \lceil \gamma n \rceil] \geqslant 1/2$. Lifting the conditioning (and using the assumed lower bound on the probability p), we get $\mathbf{Pr}\left[\sum_{i=1}^n Y_i \geqslant \lceil \gamma n \rceil\right] \geqslant p/2 > e^{-nD(\gamma \| \delta)}$, where the probability is over X_i's and Y_i's.

By Theorem 1.1, we have that there is a subset $S \subseteq [n]$ such that $\mathbf{Pr}[\wedge_{i \in S} Y_i = 1] > \delta^{|S|}$. Denote $\vec{X} = (X_1, \ldots, X_n)$, and similarly for $\vec{Y}$. We can equivalently write $\mathbf{Pr}[\wedge_{i \in S} Y_i = 1] = \mathbf{Exp}_{\vec{X}}\left[\mathbf{Exp}_{\vec{Y}}\left[\prod_{i \in S} Y_i\right]\right] = \mathbf{Exp}_{\vec{X}}\left[\prod_{i \in S} \mathbf{Exp}_{\vec{Y}}\left[Y_i\right]\right] = \mathbf{Exp}_{\vec{X}}\left[\prod_{i \in S} X_i\right]$, where the second equality is by the independence of Y_i's (given any fixing of X_i's), and the last equality by the definition of Y_i's. Thus, $\mathbf{Exp}[\prod_{i \in S} X_i] > \delta^{|S|}$, which is a contradiction. □

A sequence of random variables $X_0, \ldots, X_n$ is a *martingale* if $\mathbf{Exp}[X_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = X_i$, for all $0 \leqslant i < n$. Suppose that $X_0 = 0$. The concentration bound for martingales (Azuma's inequality [Azu67]) says that if $|X_{i+1} - X_i| \leqslant 1$ for all $1 \leqslant i \leqslant n$, then X_n is unlikely to deviate from 0 by more than $\sqrt{n}$. More precisely, for any $\lambda > 0$, $\mathbf{Pr}[X_n \geqslant \lambda\sqrt{n}] \leqslant \exp(-\lambda^2/2)$.

Theorem 3.4. *Let $0 = X_0, X_1, \ldots, X_n$ be a martingale such that $|X_{i+1} - X_i| \leqslant 1$ for all $0 \leqslant i < n$. Then, for any $\lambda > 0$, $\mathbf{Pr}[X_n \geqslant \lceil \lambda\sqrt{n} \rceil] \leqslant 2 \cdot \exp(-\lambda^2/2)$.*

Proof. Define new random variables $Y_i = X_i - X_{i-1}$, for all $1 \leqslant i \leqslant n$; the sequence $Y_1, \ldots, Y_n$ is a martingale difference sequence. Note that each $Y_i \in [-1, 1]$. Clearly, $\mathbf{Exp}[Y_{i+1} \mid Y_i, Y_{i-1}, \ldots, Y_1] = \mathbf{Exp}[Y_{i+1} \mid X_i, X_{i-1}, \ldots, X_0] = 0$. Let us also define the random variables $Z_i = (1 + Y_i)/2$, for $1 \leqslant i \leqslant n$. Observe that each $Z_i \in [0, 1]$. We want to apply Theorem 3.3 to the Z_i's. To this end, we show that, for every subset $S \subseteq [n]$, $\mathbf{Exp}[\prod_{i \in S} Z_i] = (1/2)^{|S|}$. The proof of this is by induction on $|S|$, and using the martingale property of Y_i's.

Applying Theorem 3.3 to the Z_i's (with $\delta = 1/2$ and $\gamma = 1/2 + \epsilon$), we get that, for every $0 \leqslant \epsilon \leqslant 1/2$, $\mathbf{Pr}[\sum_{i=1}^n Z_i \geqslant \lceil (1/2 + \epsilon)n \rceil] \leqslant 2 \cdot \exp(-nD(1/2 + \epsilon \parallel 1/2)) \leqslant 2 \cdot \exp(-2\epsilon^2 n)$. Since $\sum_{i=1}^n Z_i = n/2 + (\sum_{i=1}^n Y_i)/2$, we get $\mathbf{Pr}[\sum_{i=1}^n Y_i \geqslant \lceil 2\epsilon n \rceil] \leqslant 2 \cdot \exp(-2\epsilon^2 n)$. Using the fact that $\sum_{i=1}^n Y_i = X_n$ and choosing ϵ so that $\lambda = 2\epsilon\sqrt{n}$, we conclude that $\mathbf{Pr}[X_n \geqslant \lceil \lambda\sqrt{n} \rceil] \leqslant 2 \cdot \exp(-\lambda^2/2)$. □

3.3 Expander walks

We recall some basic definitions (for more details on expanders, see the excellent survey [HLW06]). For a d-regular undirected graph $G = (V, E)$ on n vertices, let $A = (a_{i,j})$ be its normalized adjacency matrix (where each entry of the adjacency matrix is divided by d). All eigenvalues of A are between -1 and 1, with the largest eigenvalue being equal to 1. Order all eigenvalues according to their absolute values. For $0 \leqslant \lambda \leqslant 1$, we call G a λ-*expander* if the second largest (in absolute value) eigenvalue of A is at most λ.

Expanders have numerous applications in computer science and mathematics (cf. [HLW06]), in particular, due to the following sampling properties. The *hitting* property of expanders, first shown by Ajtai, Komlos, and Szemeredi [AKS87], and later improved by Alon et al. [AFWZ95], is the following.

Theorem 3.5 (Hitting property of expander walks [AKS87, AFWZ95])**.** *Let $G = (V, E)$ be a λ-expander, and let $W \subset V$ be any vertex subset of measure μ, with $\mu \geqslant 6\lambda$. Then the probability that a $(t-1)$-step random walk started from a uniformly random vertex stays inside W is at most $\mu(\mu + 2\lambda)^{t-1}$. Moreover, for any subset $S \subseteq [t]$, the probability that, in each of the time steps $i \in S$, the random walk hits a vertex in W is at most $(\mu + 2\lambda)^{|S|}$.*

The second sampling property, originally proved by Gillman [Gil98], is similar to the Chernoff-Hoeffding concentration bound, and is sometimes called the *Chernoff bound for expander walks*.

Theorem 3.6 (Chernoff bound for expander walks [Gil98])**.** *Let $G = (V, E)$ be a λ-expander, and let $W \subset V$ be any vertex subset of measure μ. Then the probability that a $(t-1)$-step random walk started from a uniformly random vertex contains at least $(\mu + \epsilon)t$ vertices from W is at most $e^{-\epsilon^2(1-\lambda)t/4}$.*

The hitting property of Theorem 3.5 is fairly easy to prove, using basic linear algebra. In contrast, the original proof of Theorem 3.6 relied on some tools from perturbation theory and complex analysis. Subsequently, the proof was significantly simplified by Healy [Hea08], who used only basic linear algebra.

We first observe the following.

Theorem 3.7. *Let $G = (V, E)$ be a λ-expander, and let $W \subset V$ be of measure μ, where $\mu \geqslant 6\lambda$. Let $1 > \epsilon > 2\lambda$. Then the probability that $(t-1)$-step random walk started from a uniformly random vertex contains at least $(\mu + \epsilon)t$ vertices from W is at most $e^{-tD(\mu+\epsilon\|\mu+2\lambda)} \leqslant e^{-2(\epsilon-2\lambda)^2 t}$.*

Proof. Define the 0-1-valued random variables $X_1, \ldots, X_t$ where $X_i = 1$ if the ith step of a random walk in G lands in W, and $X_i = 0$ otherwise. By Theorem 3.5, we have that for every subset $S \subseteq [t]$, $\mathbf{Pr}[\wedge_{i\in S} X_i = 1] \leqslant (\mu + 2\lambda)^{|S|}$. By Theorem 1.1, the probability that a random walk in G contains at least $(\mu+\epsilon)t$ vertices from W is at most $e^{-tD(\mu+\epsilon\|\mu+2\lambda)}$. Using $D(\sigma + \rho \parallel \sigma) \geqslant 2\rho^2$, we can upperbound this probability by $e^{-2(\epsilon-2\lambda)^2 t}$. □

We can lift the assumption of Theorem 3.7 that $\epsilon > 2\lambda$, thereby getting

Theorem 3.8. *Let $G = (V, E)$ be a λ-expander, and let $W \subset V$ be of measure μ. Then the probability that a $(t-1)$-step random walk started from a uniformly random vertex contains at least $(\mu + \epsilon)t$ vertices from W (where $\epsilon \leqslant (2/3)\mu$) is at most $e^{-\epsilon^2(1-\lambda)t/(2\ln 4/\epsilon)}$.*

sketch. The idea is to view random t-vertex walks in the graph G also as t/c-vertex walks in the graph G^c (the cth power of the graph G), for a suitably chosen integer c. The second largest eigenvalue of G^c is at most λ^c. By choosing c so that $\lambda^c < \epsilon/2$, we will satisfy the assumptions of Theorem 3.7, for walks of length t/c, thus getting an exponentially small upper bound on the fraction of imbalanced walks in G. Since this probability is computed based on walks of length t/c rather than t, we lose an extra factor (namely, $(1-\lambda)/(\ln 1/\epsilon)$) in the exponent. □

4 Application: Uniform TDPTs for CAPTCHAs

CAPTCHAs are a special case of weakly verifiable puzzles defined by [CHS05]. A *weakly verifiable puzzle* has two components: *(1)* a polynomial-time sampleable distribution ensemble $D = \{D_n\}_{n\geqslant 1}$ on pairs (x, α), where x is called the puzzle and α the check string (n is the security parameter); and *(2)* a polynomial-time computable relation $R((x, \alpha), y)$, where y is a string of a fixed polynomially-related length. Here we think of α as a uniform random string used to generate the puzzle x. The *k-wise direct product puzzle* P^k is defined in the obvious way.

A puzzle P is called *δ-hard* (for some $0 \leqslant \delta \leqslant 1$) if, for every randomized polynomial-time algorithm A, there is a negligible function *negl* so that the success probability of A on a random P-instance is at most $(1-\delta) + negl$.

Theorem 4.1 ([CHS05])**.** *If a puzzle P is $(1-\rho)$-hard, for some $0 \leqslant \rho \leqslant 1$, then P^k is $(1-\rho^k)$-hard.*

We show the following optimal threshold direct-product result for P^k.

Theorem 4.2. *Suppose a puzzle P is $(1-\rho)$-hard, for a constant $0 \leqslant \rho \leqslant 1$. Let $\gamma = \rho + \nu \leqslant 1$, for any constant $0 \leqslant \nu \leqslant 1$. Then, for every randomized polynomial-time algorithm A, there is a negligible function negl such that the following holds: The fraction of k-tuples $\vec{x} = (x_1, \ldots, x_k)$ of instances of P^k where A solves correctly at least γk of the x_i's, is at most $e^{-kD(\gamma\|\rho)} + negl$.*

Proof. Suppose A is a randomized polynomial-time algorithm that violates the conclusion of the theorem. For random strings $\alpha_1, \ldots, \alpha_k$, define the 0-1-valued random variables $Z_1, \ldots, Z_k$ so that, for each $1 \leqslant i \leqslant k$, $Z_i = 1$ iff the algorithm $A(x_1, \ldots, x_k)$ is correct on x_i, where $x_1, \ldots, x_k$ are the puzzles determined by the random strings $\alpha_1, \ldots, \alpha_k$. Note that the distribution of $Z_1, \ldots, Z_k$ is efficiently sampleable since A is efficient (and since the puzzle P is defined for a polynomial-time sampleable distribution D).

By assumption, there is some nonnegligible function $\eta \geqslant e^{-kD(\gamma\|\rho)}$ so that $\mathbf{Pr}[\sum_{i=1}^{k} Z_i \geqslant \gamma k] \geqslant e^{-kD(\gamma\|\rho)} + 2\eta$. By Theorem 1.2, we can efficiently find a subset $S \subseteq [k]$ such that $\mathbf{Pr}[\wedge_{i\in S} Z_i = 1] > \rho^{|S|} + \eta'$, where $\eta' = \Omega(\eta^{4/(\nu\rho)})$ is nonnegligible. Thus we have an efficient algorithm that solves $P^{|S|}$ with success probability noticeably higher than $\rho^{|S|}$, contradicting Theorem 4.1. □

Remark 4.3. *The proof argument of Theorem 4.2 applies to* any *cryptographic interactive protocol as long as the protocol can be* efficiently simulated *(so that the corresponding distribution $Z_1, \ldots, Z_k$ is efficiently sampleable). Hence, for every class of protocols that can be efficiently simulated,* there is an optimal DPT for the class iff there is an optimal TDPT*; here the hardness parameters (as ρ and ν in Theorem 4.2) are assumed to be constants.*

Theorem 4.2 provides an optimal concentration bound, but under the assumption that the probabilities γ and ρ are constant; the same assumption is also needed for the similar result of [CL10]. The earlier bounds of [IJK09, Jut10] do not make such an assumption, but they are not optimal. Using conditioning in the reductions, we can remove the said limitation on γ and δ, albeit at the expense of losing the tightness of the probability bound.

5 Summary

Let $X_1, \ldots, X_n$ be Boolean random variables such that, for some $0 \leqslant \delta \leqslant 1$, $\mathbf{Pr}[X_i = 0] \leqslant \delta$, for $1 \leqslant i \leqslant n$. Let $bias(X_i) = \mathbf{Pr}[X_i = 0] - \mathbf{Pr}[X_i = 1] \leqslant \beta = 2\delta - 1$, for $1 \leqslant i \leqslant n$. Consider the following statements.

1. $X_1, \ldots, X_n$ are independent.
2. $\forall S \subseteq [n]$, $bias(\oplus_{i\in S} X_i) \leqslant \beta^{|S|}$.
3. $\forall S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i\in S}(X_i = 0)] \leqslant \delta^{|S|}$.
4. $\forall S \subseteq [n]$, $\forall 0 \leqslant \delta \leqslant \gamma \leqslant 1$, $\mathbf{Pr}[\{X_i\}_{i\in S}$ has $\geqslant \gamma|S|$ zeros $] \leqslant e^{-|S|\cdot D(\gamma\|\delta)}$.

Theorem 5.1. (1) ⇒ (2) ⇒ (3) ⇔ (4).

Proof. (1) ⇒ (2) is trivial. For (2) ⇒ (3), see Theorem 1.6. For (3) ⇒ (4), see Theorem 3.1 (the implication (4) ⇒ (3) is trivial). □

The analogous statement for direct product theorems is: optimal XOR Theorems ⇒ optimal DPTs ⇔ optimal TDPTs. Moreover, the implications have *constructive* proofs.

References

[ABHL03] L. von Ahn, M. Blum, N.J. Hopper, and J. Langford. CAPTCHA: Using hard AI problems for security. In *Advances in Cryptology - EUROCRYPT 2003, International Conference on the Theory and Applications of Cryptographic Techniques*, pages 294–311, 2003.

[AFWZ95] N. Alon, U. Feige, A. Wigderson, and D. Zuckerman. Derandomized graph products. *Computational Complexity*, 5(1):60–75, 1995.

[AKS87] M. Ajtai, J. Komlos, and E. Szemeredy. Deterministic simulation in LOGSPACE. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 132–140, 1987.

[Azu67] K. Azuma. Weighted sums of certain dependent random variables. *Tohoku Math. Journal*, 19:357–367, 1967.

[BARW08] A. Ben-Aroya, O. Regev, and R. de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science*, pages 477–486, 2008.

[Ber64] S.N. Bernstein. *Collected works, Vol 4. The probability theory, Mathematical Statistics (1911–1946)*. Nauka, Moscow, 1964. (in Russian).

[BIN97] M. Bellare, R. Impagliazzo, and M. Naor. Does parallel repetition lower the error in computationally sound protocols? In *Proceedings of the Thirty-Eighth Annual IEEE Symposium on Foundations of Computer Science*, pages 374–383, 1997.

[Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–509, 1952.

[CHS05] R. Canetti, S. Halevi, and M. Steiner. Hardness amplification of weakly verifiable puzzles. In *Theory of Cryptography, Second Theory of Cryptography Conference, TCC 2005*, pages 17–33, 2005.

[CL10] K.M. Chung and F.H. Liu. Tight parallel repetition theorems for public-coin arguments. In *Theory of Cryptography Conference*, pages 19–36, 2010.

[CSUU07] R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Perfect parallel repetition theorem for quantum XOR proof systems. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, pages 109–114, 2007.

[Gil98] D. Gillman. A Chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[GNW95] O. Goldreich, N. Nisan, and A. Wigderson. On Yao's XOR-Lemma. *Electronic Colloquium on Computational Complexity*, TR95-050, 1995.

[Hai09] I. Haitner. A parallel repetition theorem for any interactive argument. In *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science*, pages 241–250, 2009.

[Hea08] A. Healy. Randomness-efficient sampling within NC^1. *Computational Complexity*, 17(1):3–37, 2008.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *American Statistical Journal*, pages 13–30, 1963.

[Hol05] T. Holenstein. Key agreement from weak bit agreement. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, pages 664–673, 2005.

[Hol07] T. Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pages 411–419, 2007.

[IJK09] R. Impagliazzo, R. Jaiswal, and V. Kabanets. Chernoff-type direct product theorems. *J. Cryptology*, 22(1):75–92, 2009.

[IJKW10] R. Impagliazzo, R. Jaiswal, V. Kabanets, and A. Wigderson. Uniform direct-product theorems: Simplified, optimized, and derandomized. *SIAM Journal on Computing*, 39(4):1637–1665, 2010.

[Imp95] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of the Thirty-Sixth Annual IEEE Symposium on Foundations of Computer Science*, pages 538–545, 1995.

[JS68] K. Jogdeo and S. Samuels. Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation. *Annals of Mathematical Statistics*, 39:1191–1195, 1968.

[Jut10] C.S. Jutla. Almost optimal bounds for direct product threshold theorem. In *Theory of Cryptography Conference*, pages 37–51, 2010.

[Lev87] L.A. Levin. One-way functions and pseudorandom generators. *Combinatorica*, 7(4):357–363, 1987.

[NRS94] N. Nisan, S. Rudich, and M. Saks. Products and help bits in decision trees. In *Proceedings of the Thirty-Fifth Annual IEEE Symposium on Foundations of Computer Science*, pages 318–329, 1994.

[PRW97] I. Parnafes, R. Raz, and A. Wigderson. Direct product results and the GCD problem, in old and new communication models. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 363–372, 1997.

[PS97] A. Panconesi and A. Srinivasan. Randomized distributed edge coloring via an extension of the Chernoff-Hoeffding bounds. *SIAM Journal on Computing*, 26(2):350–368, 1997.

[PW07] K. Pietrzak and D. Wikstrom. Parallel repetition of computationally sound protocols revisited. In *Theory of Cryptography, Fourth Theory of Cryptography Conference, TCC 2007*, pages 86–102, 2007.

[Rao08] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 1–10, 2008.

[Raz98] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[Raz08] R. Raz. A counterexample to strong parallel repetition. In *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science*, 2008.

[Sha03] R. Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1-2):1–22, 2003.

[Ung09] F. Unger. A probabilistic inequality with applications to threshold direct-product theorems. In *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science*, pages 221–229, 2009.

[VW08] E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

[Yao82] A.C. Yao. Theory and applications of trapdoor functions. In *Proceedings of the Twenty-Third Annual IEEE Symposium on Foundations of Computer Science*, pages 80–91, 1982.