Electronic Colloquium on Computational Complexity, Report No. 72 (2010)

Constructive Proofs of Concentration Bounds

Russell Impagliazzo* Valentine Kabanets†

April 8, 2010

Abstract

We give a simple 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

*Institute for Advanced Study & University of California, San Diego, russell@cs.ucsd.edu. Supported by the NSF grant DMS-0635607. Any opinions expressed in this work are those of the authors and do not necessarily reflect the views of the NSF.

†Institute for Advanced Study & Simon Fraser University, kabanets@cs.sfu.ca. Supported by the NSF grant DMS-0635607. Any opinions expressed in this work are those of the authors and do not necessarily reflect the views of the NSF.

ISSN 1433-8092

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.

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 [IJK09b], 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 $p^{|S|}$.

Translated into 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 this limitation may be circumvented if one allows slightly more complicated reductions (which use conditioning).

We give a more detailed description of our contributions next.

1.1 Chernoff-Hoeffding bounds

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.

More precisely, for any given parameter $\lambda \geqslant 0$, the probability of the event $\sum_{i=1}^n X_i \geqslant \lambda$ is equal to that of the event $e^{t \cdot \sum_{i=1}^n X_i} \geqslant e^{t \cdot \lambda}$, for any $t > 0$. Applying Markov's inequality, one gets

$$\mathbf{Pr}\left[e^{t \cdot \sum_{i=1}^n X_i} \geqslant e^{t \cdot \lambda}\right] \leqslant \mathbf{Exp}\left[e^{t \cdot \sum_{i=1}^n X_i}\right] / e^{t \cdot \lambda}. \tag{1}$$

Using the independence of the random variables X_i's, the numerator of the above expression can be written as $\prod_{i=1}^n \mathbf{Exp}[e^{t \cdot X_i}]$. Then the remainder of the proof consists in upperbounding $\mathbf{Exp}[e^{t \cdot X_i}]$, and choosing the best value for t to minimize the right-hand side of Eq. (1).

While the proof argument sketched above is not difficult technically, we feel that it does not provide an intuitive explanation why the sums of independent random variables are likely to be concentrated around their expectations. 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 $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$ 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 below.)

1.2 Hoeffding's bound, Azuma's inequality, and the Chernoff bound for expanders

We 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). Here, for a given subset W of an expander graph G, where W has density μ, we take a t-step random walk in G and see how many times we visited the set W. The Chernoff bound for expanders [Gil98, Hea08] says that the number of visits to W will be sharply concentrated around the expectation $t\mu$. The hitting property for expanders says that the probability that all t steps end up in the set W is at most about μ^t, and it is much easier to prove [AKS87, AFWZ95]. We give a simple proof of the Chernoff property for expander walks, from the hitting property, almost matching the best known parameters of [Gil98, Hea08].

1.3 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-Hoeffding concentration bound [Che52, Hoe63] 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 [IJK09b], 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.

1.3.1 Equivalence between Direct Product and Threshold Direct Product Theorems

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].

1.3.2 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 $X_1, ..., X_n$. (See Section 4 for the precise statement.)

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 5.2).[1] DPTs for puzzles are known [BIN97, CHS05], with [CHS05] giving an optimal DPT. Also TDPTs are known [IJK09b, 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.

We also show that the limitation on the success probabilities being constant is *unavoidable* for the simple reductions between DPTs and TDPTs (see Section 4.1), and suggest a way to overcome this information-theoretic limitation using a more general class of reductions (see Section 4.2).

Finally, we want to point out that our Theorem 1.1 would imply 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.

1.3.3 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 $F^{\oplus k}(x_1,\dots,x_k) = \oplus_{i=1}^k F(x_i)$. 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 $X_1,\dots,X_k$, 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 $X_1,\dots,X_k$ be Boolean random variables such that for some $-1 \leqslant \beta \leqslant 1$, we have that, for every subset $S \subseteq \{1,\dots,k\}$, $\mathbf{Pr}[\oplus_{i\in S} X_i = 1] \leqslant 1/2 + \beta^{|S|}/2$. Then for any $\beta \leqslant \rho \leqslant 1$, $\mathbf{Pr}[\sum_{i=1}^k X_i \geqslant (1/2+\rho/2)k] \leqslant e^{-kD(1/2+\rho/2\|1/2+\beta/2)}$, for $D(\cdot \parallel \cdot)$ the relative entropy.

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 the assumption in Theorem 1.1, and thus we immediately get an alternative (and simpler) proof of Unger's result. Moreover, our reduction is constructive. Combining it with the constructive version of Theorem 1.1, we get a *constructive* version of Unger's result: if the variables $X_1,\dots,X_n$ 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] 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.

[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].

1.4 Related work

1.4.1 Chernoff bounds for negatively correlated random variables

The assumption on the random variables $X_1, \dots, X_n$ 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 given in [PS97] follows the standard proof of the Chernoff-Hoeffding bound based on upperbounding the expectation $\mathbf{Exp}[e^{t \sum_{i=1}^{n} X_i}]$ (for some parameter t), with an extra idea to use the Taylor expansion of the function e^x (to which the assumed negative correlation bounds can be applied).

1.4.2 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 3.10 (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.

Remainder of the paper We give basics in Section 2. In Section 3, we prove Theorem 1.1 and other concentration bounds, including Azuma's bound for martingales, the Chernoff bound for expander walks, and Unger's theorem. In Section 4, we state and prove the constructive version of Theorem 1.1, show the information-theoretic limitations of simple reductions between TDPTs and DPTs, and suggest a way to overcome these limitations. In Section 5, we use our constructive concentration bounds to give some examples of deriving TDPTs from DPTs. We give a brief summary in Section 6.

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 $\mathbf{Pr}\left[S \geqslant \lfloor \mu \rfloor\right] \geqslant 1/2$.*

A similar result holds also for the case of hypergeometric distributions (see, e.g., [Sie01] for a proof), where one chooses a random subset from among all subsets of a given fixed size.

Lemma 2.2 ([Sie01])**.** *Let U be a universe of size n, let $A \subseteq U$ be any subset of density $\mu = |A|/|U|$, and let $1 \leqslant t \leqslant n$ be a parameter. Then $\mathbf{Pr}_T[|T \cap A| \geqslant \lfloor t\mu \rfloor] \geqslant 1/2$, where the probability is over a random choice of t-size subset $T \subseteq U$, and $t\mu = \mathbf{Exp}_T[|T \cap A|]$ is the expected intersection size of a random t-size set T with A.*

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 $S \subseteq [n]$, $\mathbf{Pr}\left[\wedge_{i \in S} X_i = 1\right] \leqslant \prod_{i \in S} \delta_i$. Let $\delta = (1/n) \sum_{i=1}^n \delta_i$. Then, for any γ such that $\delta \leqslant \gamma \leqslant 1$, $\mathbf{Pr}\left[\sum_{i=1}^n X_i \geqslant \gamma n\right] \leqslant e^{-nD(\gamma \| \delta)}$.*

Proof. For simplicity, we first prove the special case $\delta_1 = \cdots = \delta_n$, and then show how the general case is reduced to this special case.

Case of equal δ_i's: 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, we get that

$$\mathbf{Exp}[\wedge_{i \in S} X_i = 1] \geqslant \mathbf{Exp}[\wedge_{i \in S} X_i = 1 \mid \mathcal{E}] \cdot p, \tag{2}$$

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 $\mathbf{Pr}[\wedge_{i\in S} X_i = 1] \leqslant \delta^{|S|}$. Hence,

$$\mathbf{Exp}[\wedge_{i\in S} X_i = 1] \leqslant \sum_{k=0}^{n} \binom{n}{k} q^k (1-q)^{n-k} \delta^k = (q\delta + 1 - q)^n, \tag{3}$$

where we sum over all $\binom{n}{k}$ subsets S of size $0 \leqslant k \leqslant n$, each chosen with probability $q^k(1-q)^{n-k}$; the last equality is by 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 $X_1, \dots, X_n$, conditioned on $\mathcal{E}$. Since there are at most $n - \gamma n$ such 0 positions, we get

$$\mathbf{Exp}[\wedge_{i\in S} X_i = 1 \mid \mathcal{E}] \geqslant (1-q)^{n-\gamma n}. \tag{4}$$

Finally, combining Eqs. (2)–(4), 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 Section 2 above. Choosing $q = q^*$ to minimize $f(q)$ (see Section 2), we get $p \leqslant e^{-nD(\gamma\|\delta)}$.

Case of possibly distinct δ_i's: In the general case, the success probability of the random experiment is easily seen to be at most

$$\sum_{S\subseteq[n]} \left[q^{|S|}(1-q)^{n-|S|} \prod_{i\in S} \delta_i \right]. \tag{5}$$

Let us denote by $(z_1, \dots, z_n) \in \{0,1\}^n$ 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 Eq. (5) equals $\mathbf{Exp}_{z_1,\dots,z_n}\left[\prod_{i=1}^n \delta_i^{z_i}\right] = \prod_{i=1}^n \mathbf{Exp}_{z_i}[\delta_i^{z_i}] = \prod_{i=1}^n (q\delta_i + 1 - q)$, 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 $\prod_{i=1}^n (q\delta_i + 1 - q) \leqslant (q\delta + 1 - q)^n$.

The rest of the proof is identical to that of the previous case. □

Remark 3.2. Note that if $\gamma = 1$ in the theorem above, then the resulting probability that all $X_i = 1$ is at most $e^{-nD(1\|\delta)} = \delta^n$, which is tight.

3.2 Real-valued random variables

We prove a version of Theorem 1.1 for the case of real-valued random variables. (A generalization of Theorem 3.1 to the real-valued case can be proved using a similar argument.)

Theorem 3.3. *There is a universal constant $c \geqslant 1$ satisfying the following. Let $X_1, \dots, X_n$ be real-valued random variables, with each $X_i \in [0,1]$. 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 $\delta \leqslant \gamma \leqslant 1$, $\mathbf{Pr}\left[\sum_{i=1}^n X_i \geqslant \gamma n\right] \leqslant c \cdot e^{-nD(\gamma\|\delta)}$.*

Proof. Let $p = \mathbf{Pr}\left[\sum_{i=1}^n X_i \geqslant \gamma n\right]$. Suppose that $p > c \cdot \exp(-nD(\gamma \parallel \delta))$. Our proof is by a reduction to the Boolean case. Consider Boolean random variables $Y_1, \dots, Y_n$, 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 $x_1, \dots, x_n$ from the distribution $X_1, \dots, X_n$. Conditioned on $\sum_{i=1}^n x_i \geqslant \gamma n$, we have by Lemma 2.1 that $\mathbf{Pr}[\sum_{i=1}^n Y_i \geqslant \gamma n] \geqslant 1/c$, for a universal constant

$c \geqslant 1$. [3] Lifting the conditioning (and using the assumed lower bound on the probability p), we get $\mathbf{Pr}\left[\sum_{i=1}^n Y_i \geqslant \gamma n\right] \geqslant p/c > e^{-nD(\gamma\|\delta)}$, 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 $\vec{X} = (X_1, \dots, X_n)$, 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. □

3.3 Martingales

Here we use Theorem 3.3 to derive Azuma's inequality for martingales. Intuitively, a martingale is a sequence of random variables $X_0, X_1, \dots, X_n$, where we think of X_i as the value of the random process at time i, such that each time step increases the current value by a random amount whose expectation is zero. More formally, a sequence of random variables $X_0, \dots, X_n$ is a *martingale* if $\mathbf{Exp}[X_{i+1} \mid X_i, X_{i-1}, \dots, X_0] = X_i$, for all $0 \leqslant i < n$.

Suppose that $X_0 = 0$, for simplicity. 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 $\lambda > 0$, $\mathbf{Pr}[X_n \geqslant \lambda\sqrt{n}] \leqslant \exp(-\lambda^2/2)$.

We will prove the following.

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

Proof. Define new random variables $Y_i = X_i - X_{i-1}$, for all $1 \leqslant i \leqslant n$; the sequence $Y_1, \dots, Y_n$ is a martingale difference sequence. Note that each $Y_i \in [-1, 1]$. Clearly, $\mathbf{Exp}[Y_{i+1} \mid Y_i, Y_{i-1}, \dots, Y_1] = \mathbf{Exp}[Y_{i+1} \mid X_i, X_{i-1}, \dots, X_0] = 0$.

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 will show that, for every subset $S \subseteq [n]$, $\mathbf{Exp}[\prod_{i\in S} Z_i] = (1/2)^{|S|}$.

We argue by induction over the size of the set S. When $S = \varnothing$, there is nothing to prove. Consider any nonempty subset S, and let $j \in S$ be the largest element in S. Define $S' = S - \{j\}$. The expectation $\mathbf{Exp}[\prod_{i\in S} Z_i]$ is equal to the average of the conditional expectations $\mathbf{Exp}[Z_j \cdot \prod_{i\in S'} Z_i \mid Y_{j-1}, \dots, Y_1]$ over all possible values of the variables $Y_{j-1}, \dots, Y_1$. For every fixing of the $Y_{j-1}, \dots, Y_1$, the conditional expectation of Y_j is 0, and hence, the conditional expectation of Z_j is 1/2. It follows that $\mathbf{Exp}[\prod_{i\in S} Z_i] = (1/2) \cdot \mathbf{Exp}[\prod_{i\in S'} Z_i]$. By the induction hypothesis, $\mathbf{Exp}[\prod_{i\in S'} Z_i] = (1/2)^{|S'|}$, and so, $\mathbf{Exp}[\prod_{i\in S} Z_i] = (1/2)^{|S|}$.

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

[3] More precisely, Lemma 2.1 implies that $\mathbf{Pr}[\sum_{i=1}^n Y_i \geqslant \lfloor\gamma n\rfloor] \geqslant 1/2$ and so instead of γn we should actually use $\lfloor\gamma n\rfloor \geqslant \gamma n - 1$ in our estimates. However, this will only affect the final probability bound by a constant factor, which we take into account by the constant c in the statement of the theorem.

3.4 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, i.e., $a_{i,j}$ is $(1/d)$ times the number of edges between vertices i and j. 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. For any set W of vertices of a λ-expander G, say W of measure 1/2, if we pick a random vertex in G and then walk randomly in G for $t-1$ steps (for some integer $t \geqslant 1$), then the probability our walk stays within W is at most $\exp(-\Omega(t))$ [AKS87], and the probability the walk contains the number of vertices from W that is significantly different from the expected number $tn/2$ is also at most $\exp(-\Omega(t))$ [Gil98].

The first sampling property is called the *hitting* property of expanders, and was first shown by Ajtai, Komlos, and Szemeredi [AKS87]. We state the improved version due to Alon et al. [AFWZ95].

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 that Theorem 1.1 implies a version of the Chernoff bound for expanders from the hitting property.

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 $X_1, \dots, X_t$ 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 $e^{-2(\epsilon-2\lambda)^2 t}$. □

Next we show how to lift the assumption of Theorem 3.7 that $\epsilon > 2\lambda$, thereby getting the following.

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 $e^{-\epsilon^2(1-\lambda)t/(2\ln 4/\epsilon)}$.*

Proof. 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.

Let $\gamma = \mu + \epsilon$. Let p be the fraction of walks with at least γt vertices from W. Choose an integer $c \geqslant 1$ such that $\lambda^c \leqslant \min\{\epsilon/4, \mu/6\} = \epsilon/4$; set $c = (\ln 4/\epsilon)/(\ln 1/\lambda)$. Let $G' = G^c$, the graph G raised to the power c. For $t' = t/c$ (assumed integer, for simplicity), we will also be taking t'-vertex random walks in G'. Note that the second largest (in absolute value) eigenvalue of G' is $\lambda' \leqslant \lambda^c$. Let $\delta = \mu + \epsilon/2$. By Theorem 3.5, we have that for every subset $S \subseteq [t']$, the probability that a random t'-vertex walk in G' hits W in each time step $j \in S$ is at most $\delta^{|S|}$.

We will follow the argument in the proof of Theorem 3.1. Consider the following random experiment. Pick a uniformly random t-vertex walk $\bar{v} = (v_0, v_1, \dots, v_{t-1})$ in G. Pick a uniformly random integer i where $0 \leqslant i < c$. Let $\bar{w}^i$ be the subsequence of exactly those vertices v_j of $\bar{v}$ with $j = i \mod c$. For each vertex in $\bar{w}^i$, independently, choose this vertex with probability q (for some parameter $0 \leqslant q \leqslant 1$ to be determined). Let S be the set of chosen vertices. Declare the experiment a success if $\wedge_{j\in S}(v_j \in W)$.

Next we bound the success probability of this experiment. For a t-vertex walk $\bar{v}$ in G, let $\gamma(\bar{v})$ denote the fraction of vertices of $\bar{v}$ that fall into the set W. For each subsequence $\bar{w}^i$ of $\bar{v}$, for $0 \leqslant i < c$, let $\gamma_i(\bar{v})$ denote the fraction of vertices of $\bar{w}^i$ that are in W. Observe that $\gamma(\bar{v}) = (1/c)\sum_{i=0}^{c} \gamma_i(\bar{v})$.

By our assumption, with probability p, we choose a t-vertex walk $\bar{v}$ in G such that $\gamma(\bar{v}) \geqslant \gamma$. Condition on choosing such a $\bar{v}$. For each fixed $0 \leqslant i < c$, the conditional success probability of our experiment (given i) is $(1-q)^{(1-\gamma_i(\bar{v}))t'}$. Thus the average (over i) success probability (conditioned on $\bar{v}$) is $\mathbf{Exp}_i[(1-q)^{(1-\gamma_i(\bar{v}))t'}] \geqslant (1-q)^{(1-\gamma(\bar{v}))t'} \geqslant (1-q)^{(1-\gamma)t'}$, where the first inequality is by convexity, and the second one by our assumption that $\gamma(\bar{v}) \geqslant \gamma$. Lifting the conditioning on $\bar{v}$, we get that $\mathbf{Pr}[\text{ success }] \geqslant p \cdot (1-q)^{(1-\gamma)t'}$.

On the other hand, for each fixed $0 \leqslant i < c$, our experiment concerns random t'-vertex walks in the graph G'. By the same argument as in the proof of Theorem 3.1 and by the hitting property of G' noted above, we get that the conditional success probability (given i) is at most $(q\delta + 1 - q)^{t'}$. Since this is true for every $0 \leqslant i < c$, it is also true for the average over i's. Thus we get $\mathbf{Pr}[\text{ success }] \leqslant (q\delta + 1 - q)^{t'}$.

Comparing the upper and lower bounds for $\mathbf{Pr}[\text{ success }]$ obtained above, we get $p \leqslant (q\delta + 1 - q)^{t'}/(1-q)^{(1-\gamma)t'} = (f(q))^{t'} \leqslant e^{-t'D(\gamma\|\delta)}$, for $q = q^*$ that minimizes the function f (cf. Section 2). Recall that $\gamma = \mu + \epsilon$ and $\delta = \mu + \epsilon/2$. Using $D(\sigma + \epsilon \parallel \sigma) \geqslant 2\epsilon^2$, we upperbound p by $e^{-t'\epsilon^2/2}$. Recalling that $t' = t/c$ for $c = (\ln 4/\epsilon)/(\ln 1/\lambda)$, and using the inequality $\ln 1/x \geqslant 1 - x$, we get $p \leqslant e^{-t\epsilon^2/(2c)} \leqslant e^{-\epsilon^2(1-\lambda)t/(2\ln 4/\epsilon)}$, as required. □

Compared to the best Chernoff bounds for expander walks [Gil98, Hea08], we lose a factor $\ln 1/\epsilon$ in the exponent. The advantage of our proof is simplicity. Nonetheless, it is an interesting question if an optimal Chernoff bound can be obtained from the hitting property of expander walks, using a simple direct reduction like ours.

3.5 Chernoff-Hoeffding's concentration bound from small biases

Here we give a simple, combinatorial proof of Unger's theorem [Ung09]. Let $X_1, \dots, X_n$ be Boolean random variables. For any set $S \subseteq [n]$, let $bias(S) = \mathbf{Pr}[\oplus_{i \in S} X_i = 0] - \mathbf{Pr}[\oplus_{i \in S} X_i = 1]$.

Theorem 3.9 ([Ung09])**.** *Let $X_1, \dots, X_n$ be 0-1-valued random variables. Suppose there are $-1 \leqslant \beta_i \leqslant 1$, $1 \leqslant i \leqslant n$, such that, for every $S \subseteq [n]$, $bias(S) \leqslant \prod_{i \in S} \beta_i$. Let $\beta = (1/n) \sum_{i=1}^{n} \beta_i$. Then, for any $\beta \leqslant \nu \leqslant 1$, $\mathbf{Pr}[(X_1, \dots, X_n)$ has at least $(1/2 + \nu/2)n$ zeros$] \leqslant e^{-nD(1/2+\nu/2 \| 1/2+\beta/2)}$.*

By an argument very similar to that of Viola and Wigderson [VW08], we first prove the following.

Theorem 3.10. *Let $X_1, \dots, X_n$ be 0-1-valued random variables. Suppose that there are $-1 \leqslant \beta_i \leqslant 1$, $1 \leqslant i \leqslant n$, such that, for every $S \subseteq [n]$, $bias(S) \leqslant \prod_{i \in S} \beta_i$. Let $\beta = (1/n) \sum_{i=1}^{n} \beta_i$. Then, for every $S \subseteq [n]$, $\mathbf{Pr}[\wedge_{i \in S}(X_i = 0)] \leqslant (1/2 + \beta/2)^{|S|}$.*

Proof. Denote $\delta = (1/2 + \beta/2)$. For the sake of contradiction, suppose that there is some subset S of size $0 < k \leqslant n$ such that $\mathbf{Pr}[\wedge_{i \in S}(X_i = 0)] > \delta^k$. For simplicity of notation, let us assume that $S = [k]$ (i.e., is the set of the first k positions). Consider the following random experiment. Sample a random n-tuple $(x_1, \dots, x_n)$ from the distribution $(X_1, \dots, X_n)$. Let $T \sim Bin(k, 1/2)$. Declare the experiment a success if $\oplus_{i \in T} x_i = 0$.

Define the bias of the experiment as the probability it succeeds minus the probability it fails. With probability greater than δ^k, we sample the all-zero k-tuple 0^k. Conditioned on that, our experiment succeeds with probability 1, and hence has bias 1. If the sampled k-tuple has at least one non-zero position, then the conditional success probability of the experiment is exactly 1/2, and hence the bias is 0. Thus, the overall bias of the experiment is greater than δ^k.

We next upperbound the bias of the random experiment. We first consider the case of equal β_i's, and then show how the general case follows by convexity.

Case of equal β_i's: For every fixed subset T of size m (for $0 \leqslant m \leqslant k$), the bias of the random experiment on this T is at most β^m. For each $0 \leqslant m \leqslant k$, there are $\binom{k}{m}$ subsets $T \subseteq [k]$ of size m, each selected with probability 2^{-k}. Hence, the overall bias of our experiment is at most $2^{-k} \sum_{m=0}^{k} \binom{k}{m} \beta^m = 2^{-k}(1+\beta)^k = 2^{-k} 2^k (1/2 + \beta/2)^k = \delta^k$, which is a contradiction.

Case of possibly distinct β_i's: Let $(z_1, \dots, z_k) \in \{0,1\}^k$ be the characteristic vector of a set T chosen in the random experiment above. That is, each $z_i = 1$ with probability 1/2; all z_i's are independent. The bias of our random experiment is then at most $\mathbf{Exp}_{z_1,\dots,z_k}\left[\prod_{i=1}^{k} \beta_i^{z_i}\right] = \prod_{i=1}^{k} \mathbf{Exp}_{z_i}[\beta_i^{z_i}] = \prod_{i=1}^{k}(\beta_i/2 + 1/2) \leqslant (\beta/2 + 1/2)^k$, where the first equality is by the independence of z_i's, and the last inequality by convexity (cf. the proof of Theorem 3.1). The rest of the proof is as in the previous case. □

Proof of Theorem 3.9. The proof is immediate from Theorems 3.1 and 3.10. □

4 A constructive version of the Chernoff-Hoeffding theorem

We assume here that we are given an oracle access to the distribution $X_1, \dots, X_n$ of random variables, and so we can efficiently sample from that distribution. Suppose that these variables fail to satisfy the concentration bound of Theorem 1.1 by a noticeable amount, then we can efficiently, probabilistically, find a subset $S \subseteq [n]$ such that $\mathbf{Pr}[\wedge_{i \in S} X_i = 1]$ is greater than $\delta^{|S|}$ by a noticeable amount. Our proof of Theorem 1.1 suggests an obvious algorithm for doing this: sample random subsets $S \sim Bin(n, q)$ (for an appropriate parameter $0 < q < 1$), and check if S is a good set.

More precisely, we have the following.

Theorem 4.1. *There is a randomized algorithm $\mathcal{A}$ such that the following holds. Let $X_1, \dots, X_n$ 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 $X_1, \dots, X_n$, runs in time $\mathrm{poly}(\alpha^{-1/((\gamma-\delta)\delta)}, n)$ and outputs a set $S \subseteq [n]$ such that, with probability at least $1 - o(1)$, $\mathbf{Pr}[\wedge_{i \in S} X_i = 1] > \delta^{|S|} + \Omega(\alpha^{4/((\gamma-\delta)\delta)})$.*

Proof. Recall the random experiment from the proof of Theorem 3.1. Sample a random n-tuple $(x_1, \dots, x_n)$ from the distribution $(X_1, \dots, X_n)$. Pick $S \sim Bin(n, q)$ for a parameter $0 < q < 1$ to be determined. Declare success if $\wedge_{i \in S} x_i = 1$.

As in the proof of Theorem 3.1, the overall success probability of our random experiment is greater than $p(1-q)^{(1-\gamma)n}$. For any subset $S \subseteq [n]$, let $a(S) = \mathbf{Pr}[\wedge_{i \in S} X_i = 1]$. The success probability of our experiment is $\mathbf{Exp}_S[a(S)]$. By the linearity of expectation, we have: $\mathbf{Exp}_S[a(S) - \delta^{|S|}] = \mathbf{Exp}_S[a(S)] - \mathbf{Exp}_S[\delta^{|S|}]$. As in the proof of Theorem 3.1, we have $\mathbf{Exp}_S[\delta^{|S|}] = (1 - q(1-\delta))^n$. So we get

$$\mathbf{Exp}_S[a(S) - \delta^{|S|}] > p(1-q)^{(1-\gamma)n} - (1 - q(1-\delta))^n = (1-q)^{(1-\gamma)n}(p - (f(q))^n), \tag{6}$$

where $f(q) = (1 - q(1-\delta))/(1-q)^{1-\gamma}$.

We consider two cases depending on how big α is.

Case 1: $\alpha \geqslant e^{-n(\gamma-\delta)^2/4}$. We choose $0 < q_0 \leqslant 1/2$ so that $(f(q_0))^n \leqslant \alpha$. Using the inequalities $1 - x \leqslant e^{-x}$ (valid for all x) and $1 - x \geqslant e^{-x-x^2}$ (valid for $0 < x \leqslant 1/2$, as can be seen by considering the Taylor expansion $\ln(1-x) = -\sum_{j \geqslant 1} x^j/j$), we upperbound $f(q)$ by $e^{-q(1-\delta)}e^{(1-\gamma)(q+q^2)} = e^{-q(1-\delta-(1-\gamma)(1+q))} \leqslant e^{-q(\gamma-\delta-q)}$. Assuming $q \leqslant (\gamma-\delta)/2$ (which we will ensure by our choice of q), we can upperbound $f(q) \leqslant e^{-q(\gamma-\delta)/2}$. To make the latter expression at most $\alpha^{1/n}$, we solve for q, getting $q_0 = (2\ln 1/\alpha)/(n(\gamma-\delta))$. Note that the condition $q_0 \leqslant (\gamma-\delta)/2$ is equivalent to $(\gamma-\delta)^2 \geqslant (4\ln 1/\alpha)/n$, which follows from the assumed lower bound on α. Observe also that $q_0 \leqslant 1/2$, as required.

Using the chosen q_0, we lowerbound the last expression in Eq. (6) by $(1-q_0)^{(1-\gamma)n}\alpha$ (since $p > 2\alpha$ and $(f(q_0))^n \leqslant \alpha$). Using the formula $1 - x \geqslant e^{-2x}$ (valid for $0 < x \leqslant 1/2$), we get $(1-q_0)^{(1-\gamma)n} \geqslant e^{-2q_0(1-\gamma)n} = \alpha^{4(1-\gamma)/(\gamma-\delta)}$. Thus, the expression in Eq. (6) is at least $\alpha^{4(1-\gamma)/(\gamma-\delta)+1} \geqslant \alpha^{4/(\gamma-\delta)}$.

Case 2: $e^{-nD(\gamma\|\delta)} \leqslant \alpha < e^{-n(\gamma-\delta)^2/4}$. In this case, we choose $q = q^*$ so that $f(q^*) = e^{-D(\gamma\|\delta)}$, where $q^* = (\gamma-\delta)/(\gamma(1-\delta))$ (cf. Section 2). For this choice of q, the right-hand side of Eq. (6) is $(1-q^*)^{(1-\gamma)n}(p - ((f(q^*))^n) \geqslant (1-q^*)^{(1-\gamma)n}((f(q^*))^n = (1-q^*(1-\delta))^n = (\delta/\gamma)^n$, where the first inequality is by the assumption that $p > 2(f(q^*))^n$, and then we used the definitions of f and q^*. Using $1 + x \leqslant e^x$, we get $\gamma/\delta = 1 + (\gamma-\delta)/\delta \leqslant e^{(\gamma-\delta)/\delta}$. Thus we get $\mathbf{Exp}_S[a(S) - \delta^{|S|}] > e^{-n(\gamma-\delta)/\delta} > \alpha^{4/((\gamma-\delta)\delta)}$, where the last inequality is by our assumption that $\alpha < e^{-n(\gamma-\delta)^2/4}$.

Define $\mu = \alpha^{4/((\gamma-\delta)\delta)}$. Observe that for both cases considered above, $\mathbf{Exp}_S[a(S) - \delta^{|S|}] > \mu$. Our algorithm $\mathcal{A}$ is: "Randomly sample sets $S \sim Bin(n, q)$ (where q is q_0 or q^*, depending on α), and output the first S with $a(S) - \delta^{|S|} \geqslant \mu/2$, where the probability $a(S)$ is estimated by random sampling. If no such S is found within the allotted time, then output $\varnothing$." It is easy to see that, with very high probability, $\mathcal{A}$ finds a required set S within time $\mathrm{poly}(1/\mu, n)$. □

In a similar fashion, the proof of Theorem 3.10 yields a simple randomized algorithm for finding a set T where $\oplus_{i \in T} X_i$ has a large bias, when there is a subset S where $\mathbf{Pr}[\wedge_{i \in S} X_i = 1]$ is large. For completeness, we state this below.

Theorem 4.2. *There is a randomized algorithm $\mathcal{A}$ satisfying the following. Let $X_1, \dots, X_n$ be* 0-1*-valued random variables. Let $S \subseteq [n]$ be a subset such that $\mathbf{Pr}[\wedge_{i \in S}(X_i = 0)] > (1/2 + \beta/2)^{|S|} + \alpha$, for some $-1 \leqslant \beta \leqslant 1$ and $\alpha > 0$. Then the algorithm $\mathcal{A}$, on inputs n, S, α, β, using oracle access* *to* $X_1, \dots, X_n$*, runs in time* $\mathrm{poly}(1/\alpha, n)$*, and outputs a subset* $T \subseteq S$ *such that, with probability at least* $1 - o(1)$*, bias*$(T) > \beta^{|T|} + \Omega(\alpha)$*.*

Proof. Without loss of generality, assume that $S = [k]$, for some $1 \leqslant k \leqslant n$. Let $\delta = 1/2 + \beta/2$, and let $\rho = \delta^k + \alpha$. We consider the same random experiment as in the proof of Theorem 3.10. Let $T \sim Bin(k, 1/2)$. Declare success if $\oplus_{i \in T} X_i = 0$.

As in the proof of Theorem 3.10, we get that the bias of our random experiment is greater than the probability that $\wedge_{i \in S}(X_i = 0)$, which is ρ. On the other hand, the bias of our experiment is equal to $\mathbf{Exp}_{T \subseteq [k]}[bias(T)]$, where the expectation is over uniformly chosen subsets T of $[k]$. By linearity of expectation, $\mathbf{Exp}_T[bias(T) - \beta^{|T|}] = \mathbf{Exp}_T[bias(T)] - \mathbf{Exp}_T[\beta^{|T|}]$. As in the proof of Theorem 3.10, we get that $\mathbf{Exp}_T[\beta^{|T|}] = \delta^k$. Thus, $\mathbf{Exp}_T[bias(T) - \beta^{|T|}] > \rho - \delta^k = \alpha$.

The algorithm $\mathcal{A}$ randomly samples $\mathrm{poly}(1/\alpha, n)$ sets $T \subset [k]$ and outputs a T with large $bias(T)$ (estimated by sampling). By Chernoff bounds, we get with high probability a set T such that $bias(T) > \beta^{|T|} + \Omega(\alpha)$. □

Combining Theorems 4.2 and 4.1 immediately yields an algorithmic version of Unger's theorem (Theorem 3.9), with parameters essentially the same as those of Theorem 4.1.

4.1 The near-optimality of Theorem 4.1

Theorem 4.1 shows that if given random variables $X_1, \dots, X_n$ fail the Chernoff-Hoeffding concentration bound by a noticeable amount α, then there is subset S of indices where $\Pr[\wedge_{i \in S} X_i = 1] \geqslant \delta^{|S|} + \alpha^{O(1/((\gamma - \delta)\delta)}$. Note that in the latter expression, the exponent of α depends on γ and δ, and so, in general, may not be constant. In the remainder of this section, we will argue that *the exponent of* α *must have dependence on* γ *and* δ. That is, the parameters of Theorem 4.1 are essentially optimal.

Lemma 4.3. *There are Boolean random variables* $X_1, \dots, X_n$*, 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* $S \subseteq [n]$*,*

$$\mathbf{Pr}[\wedge_{i \in S} X_i = 1] - \delta^{|S|} \leqslant (4\alpha)^{\delta(\ln 1/\delta)/(\gamma - \delta)}.$$

Proof. Let $0 < \delta < \gamma < 1$ and let $n \in \mathbb{N}$ be such that γn is an integer. Let $p > 4 \cdot \exp(-nD(\gamma \parallel \delta))$. Consider the following distribution $X_1, \dots, X_n$. Flip a random coin which is "heads" with probability p. If the coin falls "heads", then, independently for each $1 \leqslant i \leqslant n$, assign $X_i = 1$ with probability γ and 0 with probability $1 - \gamma$. Otherwise (if the coin falls "tails"), assign all $X_i = 0$.

Note that in the first case, with probability p, we get an expected number γn of 1's. Conditioned on the event that our first coin is "heads", the probability we get at least the expected γn of 1's is at least $1/2$, by Lemma 2.1. Thus, the random variables $X_1, \dots, X_n$ contain at least γn of 1's, with the overall probability at least $p/2$. Let $\alpha = p/4$. Then the defined probability distribution satisfies the assumption of Theorem 4.1.

We know that there must exist a subset $S \subseteq [n]$ of indices such that $\lambda = \Pr[\wedge_{i \in S} X_i = 1] - \delta^{|S|} > 0$. We will upperbound this quantity λ next. We have $\Pr[\wedge_{i \in S} X_i = 1] \leqslant p\gamma^{|S|}$. Thus, the maximum value of λ is at most the maximum of the function $g(x) = p \cdot \gamma^x - \delta^x$, for $0 < x \leqslant n$. Taking the derivative of g, we see that it achieves its global maximum at $x^* = (\ln 1/p + \ln(\ln \delta / \ln \gamma))/(\ln \gamma/\delta)$. Note that, for $x_0 = (\ln 1/p)/(\ln \gamma/\delta)$, we have $g(x_0) = 0$, i.e., $p\gamma^{x_0} = \delta^{x_0}$. Writing $x^* = x_0 + \epsilon$ (for $\epsilon = x^* - x_0$), we get $p\gamma^{x^*} - \delta^{x^*} = p\gamma^{x_0}\gamma^\epsilon - \delta^{x_0}\delta^\epsilon = \delta^{x_0}(\gamma^\epsilon - \delta^\epsilon)$. The latter expression is at most $\delta^{x_0} = p^{(\ln 1/\delta)/(\ln \gamma/\delta)}$. Recalling that $p = 4\alpha$, we get that the maximum value of λ can be at most $(4\alpha)^t$ for $t \geqslant \delta(\ln 1/\delta)/(\gamma - \delta)$ (where we used the inequality $(\ln \gamma/\delta) \leqslant (\gamma - \delta)/\delta$). □

The dependence on $1/(\gamma-\delta)$ in the exponent of α, as stated n Theorem 4.1, is indeed necessary. A way to circumvent this bad dependence on $\gamma-\delta$ is suggested in the next subsection.

4.2 Better efficiency through conditioning

We are given n Boolean-valued random variables $X_1,\dots,X_n$ with $\mathbf{Exp}[X_i] \leqslant \delta$ for every $i \in [n]$, and some extra condition on the X_i's (e.g., negative correlation), and we want to show that $\mathbf{Pr}[\sum_{i=1}^n X_i \geqslant \gamma n] \leqslant p$ for some $\delta < \gamma \leqslant 1$ and some $p = \exp((\gamma-\delta)^2 n)$. Moreover, we would like a constructive proof that would say that if $\mathbf{Pr}\left[\sum_{i=1}^n X_i \geqslant \gamma n\right] > \alpha$ for some $\alpha \gg p$, then one can efficiently find a subset $S \subseteq [n]$ such that $\mathbf{Pr}\left[\wedge_{i\in S} X_i = 1\right] > \delta^{|S|} + g(\alpha)$, for some function g, ideally a polynomial.

In Theorem 4.1, we argued such a constructive version for the function $g(\alpha) \geqslant \alpha^{4/(\gamma-\delta)\delta}$. This function g is not polynomial in α, and we have showed that, in general, g cannot be polynomial in α (see Section 4.1). Here we suggest a way to avoid this limitation on the function g. To achieve this, we will assume efficient sampleability from the *conditional* distributions $X_1,\dots,X_n \mid \mathcal{E}$, where $\mathcal{E}$ is a random event of the form $\sum_{j\in T} X_j \geqslant t$, for some $T \subseteq [n]$ and integer $0 \leqslant t \leqslant n$. (This assumption is usually true in the computational setting; we give an example in Section 5.2.)

For this new framework, we get the following.

Theorem 4.4. *There is a randomized algorithm $\mathcal{A}$ satisfying the following. Let $X_1,\dots,X_n$ 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 $(X_1,\dots,X_n \mid \sum_{j\in S} X_j \geqslant \gamma n/2)$, 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 $1 - o(1)$, $\mathbf{Pr}\left[X_{i_0} = 1 \mid \sum_{j\in S} X_j \geqslant \gamma n/2\right] > \delta + (\gamma-\delta)/16$.*

First we give the intuition behind the proof. Suppose we are given $X_1,\dots,X_n$ such that $\mathbf{Pr}[\sum_{i=1}^n X_i \geqslant \gamma n] > \alpha$ for α. Randomly partition the set $[n]$ into two disjoint subsets S and $\bar{S}$ (of size $n/2$ each), and consider the random variable $\sum_{i\in\bar{S}} X_i$ *conditioned on* the event $\mathcal{E}$ that $\sum_{j\in S} X_j \geqslant \gamma|S|$. We will argue that, with constant probability over the random choice of $S \subset [n]$,

1. the event $\mathcal{E}$ has probability at least $\alpha/2$, and
2. $\mathbf{Pr}\left[\sum_{i\in\bar{S}} X_i \geqslant \gamma'|\bar{S}| \mid \mathcal{E}\right] \geqslant 1 - (\gamma'-\delta)/2$, where $\gamma' \in [\delta, \gamma]$ (e.g., $\gamma' = (\gamma+\delta)/2$).

Observe what has happened. We started with n random variables $X_1,\dots,X_n$ whose sum exceeds a fractional threshold γ with some noticeable (but possibly exponentially small) probability α. Then, by conditioning on an event $\mathcal{E}$ (of non-negligible probability), we get $n/2$ random variables X_i, for $i \in \bar{S}$, such that the sum of these variables exceeds a (slightly smaller) fractional threshold γ' with the conditional probability that is close to one!

In this new conditioned sample space, we can then randomly sample a single coordinate $i \in \bar{S}$, and get an i such that $\mathbf{Pr}\left[X_i = 1 \mid \mathcal{E}\right] \geqslant \gamma'(1-(\gamma'-\delta)/2) > \delta + (\gamma'-\delta)/2$, where the first inequality follows since, with probability at least $1-(\gamma'-\delta)/2$, we get a tuple with at least γ' fraction of 1s. Thus we are likely to get a singleton set $\{i\}$ with $\mathbf{Pr}[X_i = 1] > \delta$ by a noticeable amount. This reduction can be used to derive a TDPT from DPT in concrete computational settings, as we will illustrate in the next section.

Next we will give the formal proofs.

Lemma 4.5. *Let $X_1, \ldots, X_n$ be Boolean-valued random variables, and let $0 \leqslant \delta < \gamma' < \gamma \leqslant 1$. Suppose that $\mathbf{Pr}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geqslant \gamma\right] > \alpha$, for some $\alpha > 0$. Let $S \subset [n]$ be a random subset of size $n/2$, and let $\bar{S} = [n] - S$ be its complement. We have*

$$\mathbf{Pr}\left[\frac{1}{|\bar{S}|}\sum_{i \in \bar{S}} X_i \geqslant \gamma' \mid \frac{1}{|S|}\sum_{j \in S} X_j \geqslant \gamma\right] \geqslant 1 - \epsilon, \tag{7}$$

for $\epsilon \leqslant 4 \cdot e^{-(\gamma-\gamma')^2 n/16}/\alpha$, where the probability is over $X_1, \ldots, X_n$ and a random choice of S.

Proof. As before, let us denote by $\mathcal{E}$ the event that $\frac{1}{|S|}\sum_{j \in S} X_j \geqslant \gamma$. First we show that $\mathcal{E}$ has probability $\Omega(\alpha)$.

Claim 4.6. $\mathbf{Pr}[\mathcal{E}] \geqslant \alpha/2$.

Proof of Claim 4.6. By our assumption, with probability at least α, we get a tuple $\bar{x} = (x_1, \ldots, x_n)$ from $X_1, \ldots, X_n$ such that $\bar{x}$ has at least γn ones. Conditioned on $\bar{x}$ having at least that many 1s, a random $n/2$-size subset $S \subset [n]$ is expected to contain $\gamma|S|$ positions with 1s. By the "mean is median" theorem for hypergeometric distributions, Lemma 2.2, we get that at least $1/2$ fraction of sets S will contain at least $\lfloor \gamma|S| \rfloor$ positions with 1s. Lifting the conditioning, we get that $\mathcal{E}$ happens with probability at least $\alpha/2$. □

The desired inequality in Eq. (7) can be equivalently written as: $\mathbf{Pr}\left[\frac{1}{|\bar{S}|}\sum_{i \in \bar{S}} X_i < \gamma' \mid \mathcal{E}\right] \leqslant \epsilon$. Hence, we need to upperbound $\mathbf{Pr}\left[\frac{1}{|\bar{S}|}\sum_{i \in \bar{S}} X_i < \gamma' \;\&\; \mathcal{E}\right] / \mathbf{Pr}[\mathcal{E}\,]$. We can use Claim 4.6 to lowerbound the denominator of the expression above. To upperbound the numerator, we use the fact that both S and its complement $\bar{S}$ are likely to give very accurate estimates of the fraction of 1s in a given n-tuple $\bar{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$. More precisely, let $\bar{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$ be any string with μn ones. Let $S \subseteq [n]$ be a random s-size subset (for some $1 \leqslant s \leqslant n$; think of $s = n/2$). Then, by the Hoeffding bound for hypergeometric distributions [Hoe63] (see, e.g., [JLR00] for the proof of the variant of the bound given below), we have

$$\mathbf{Pr}_S\left[\left|\frac{1}{s}\sum_{i \in S} x_i - \mu\right| \geqslant \nu\right] \leqslant 2e^{-\nu^2 s/2}. \tag{8}$$

In our case, both S and its complement $\bar{S}$ are random subsets of size $s = n/2$, and so each of them, individually, satisfies the above concentration bound. We get the following claim.

Claim 4.7. *For every $\nu > 0$ and every n-tuple $\bar{x}$, $\mathbf{Pr}_S\left[\left|\frac{1}{|S|}\sum_{i \in S} x_i - \frac{1}{|\bar{S}|}\sum_{j \in \bar{S}} x_j\right| > 2\nu\right] \leqslant 2e^{-\nu^2 n/4}$.*

Proof of Claim 4.7. Fix any $\bar{x}$. Let μ be the fraction of 1s in $\bar{x}$. If both S and $\bar{S}$ estimate the fraction of 1s in $\bar{x}$ to within an additive error ν, then their estimates are within 2ν of each other. Thus, the probability of the event in the claim is at most the probability that either S or $\bar{S}$ errs in the estimate of μ by more than additive ν. Applying the Hoeffding concentration bound of Eq. (8) completes the proof. □

Using Claim 4.7 with $\nu = (\gamma - \gamma')/2$, we get $\mathbf{Pr}\left[\frac{1}{|\bar{S}|}\sum_{i \in \bar{S}} X_i < \gamma' \;\&\; \frac{1}{|S|}\sum_{j \in S} X_i \geqslant \gamma\right] \leqslant 2e^{-(\gamma-\gamma')^2 n/16}$. Together with $\mathbf{Pr}[\mathcal{E}] \geqslant \alpha/2$ (of Claim 4.6), this concludes the proof. □

Proof of Theorem 4.4. Let $\gamma' \in [\delta, \gamma]$ such that $(\gamma - \delta)/2 = \gamma - \gamma' = \gamma' - \delta$. By Lemma 4.5, for random $n/2$-size subset $S \subset [n]$, $\mathbf{Pr}\left[\sum_{i \in \bar{S}} X_i \geqslant \gamma'(n/2) \mid \sum_{j \in S} X_j \geqslant \gamma n/2\right] \geqslant 1 - \epsilon$, where $\epsilon \leqslant 4 \cdot e^{-(\gamma-\delta)^2 n/64}/\alpha$. By Markov, with probability at least 1/2 over the choice of S, we have $\mathbf{Pr}\left[\sum_{i \in \bar{S}} X_i \geqslant \gamma'(n/2) \mid \sum_{j \in S} X_j \geqslant \gamma n/2\right] \geqslant 1 - 2\epsilon$, where the probability is over $X_1, \dots, X_n$ only.

By our assumption, $2\epsilon \leqslant (\gamma - \delta)/4$. Thus, in the sample space conditioned on the event $\mathcal{E}$ that $\sum_{j \in S} X_j \geqslant \gamma n/2$, we have that $\mathbf{Pr}[\sum_{i \in \bar{S}} X_i \geqslant \gamma' n/2 \mid \mathcal{E}] \geqslant 1 - (\gamma - \delta)/4$. If we randomly sample an index $i \in \bar{S}$, we get $\mathbf{Pr}[X_i = 1 \mid \mathcal{E}] \geqslant \gamma'(1 - (\gamma - \delta)/4) \geqslant \delta + (\gamma - \delta)/8$.

If we randomly sample $i \in \bar{S}$ and estimate $\mathbf{Pr}[X_i = 1 \mid \mathcal{E}] - \delta$ to within $(\gamma - \delta)/16$, we get, with high probability, an index $i_0 \in \bar{S}$ such that $\mathbf{Pr}[X_{i_0} = 1 \mid \mathcal{E}] \geqslant \delta + (\gamma - \delta)/16$. Our algorithm $\mathcal{A}$ will randomly choose $S \subseteq [n]$ and $i_0 \in \bar{S}$, and check (by sampling) that these are good choices. The running time of the described algorithm for finding S and i_0 is $\text{poly}(n, 1/\alpha, 1/(\gamma - \delta))$ and the time required to sample from the conditional distribution $(X_1, \dots, X_n \mid \sum_{j \in S} X_j \geqslant \gamma n/2)$. □

5 Applications: Uniform TDPTs

5.1 Hardness amplification of CAPTCHA puzzles

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 $(1 - \delta) + \mathit{negl}$.

Theorem 5.1 ([CHS05])**.** *Suppose 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 5.2. *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 $\vec{x} = (x_1, \dots, x_k)$ of instances of P^k where A solves correctly at least γk of the x_i's, is at most $e^{-kD(\gamma \| \rho)} + \mathit{negl}$.*

Proof. For the sake of contradiction, suppose A is a randomized polynomial-time algorithm that violates the conclusion of the theorem. For random strings $\alpha_1, \dots, \alpha_k$, define the 0-1-valued random variables $Z_1, \dots, Z_k$ so that, for each $1 \leqslant i \leqslant k$, $Z_i = 1$ iff the algorithm $A(x_1, \dots, x_k)$ is correct on x_i, where $x_1, \dots, x_k$ are the puzzles determined by the random strings $\alpha_1, \dots, \alpha_k$. Thus, the random variables Z_i's are defined based on the random tapes α_i's and the internal randomness of the algorithm A. Also note that the distribution of $Z_1, \dots, Z_k$ 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 4.1, we can efficiently find (in randomized time $\text{poly}(\eta^{-1/(\nu\rho)})$, which is polynomial for constant ν and ρ, and for a nonnegligible function η) 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. Let $|S| = t$. The above means that we have an efficient algorithm that solves the t-wise direct product P^t with success probability noticeably higher than ρ^t. By Theorem 5.1, this contradicts the assumed $(1-\rho)$-hardness of P. $\square$

Remark 5.3. It is easy to see that the argument in the proof of Theorem 5.2 continues to hold for *any* cryptographic interactive protocol as long as the protocol can be *efficiently simulated*; the latter is needed to ensure efficient sampleability of the distribution $Z_1, \dots, Z_k$ as defined in the proof above. As a corollary, we get that 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 5.2) are assumed to be constants.

Our Theorem 5.2 provides an optimal concentration bound, but it needs 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 [IJK09b, Jut10] do not make such an assumption, but they are not optimal. Using conditioning in the reductions (cf. Section 4.2), we can remove the said limitation on γ and δ, albeit at the expense of losing the tightness of the probability bound. Nonetheless, the resulting proof is conceptually simpler than, e.g., the earlier proof of [IJK09b]. For completeness, we present a version of the argument in the following subsection.

5.2 Using conditioning

To simplify the notation, we show a uniform TDPT for Boolean circuits, rather than for CAPTCHAs. The results for CAPTCHAs follow by essentially the same reasoning.

Let $F\colon U \to R$ be a function. Let $0 \leqslant \delta \leqslant \gamma \leqslant 1$. Suppose we are given a circuit $C\colon U^n \to R^n$ such that, for at least $\alpha > 0$ fraction of inputs $u \in U^n$, the circuit $C(u)$ correctly computes $F^n(u)$ in at least γn positions. We will show how to compute F on at least δ fraction of inputs in U, assuming we can generate the pairs $(u, F(u))$ for random $u \in U$. The ability to simulate a function (protocol) on random inputs of our choice is the only assumption on the model. Intuitively, this assumption is used to argue that the following distribution (as well as its conditioned versions) is efficiently sampleable: $X_1, \dots, X_n$, where $X_i = 1$ iff C is correct in the ith position, for a random input in U^n. In turn, this sampleability allows us to apply Theorem 4.4.

Theorem 5.4. *There is a randomized algorithm $\mathcal{A}$ satisfying the following. let $F\colon U \to R$ be any function. For parameters $0 \leqslant \delta \leqslant \gamma \leqslant 1$, let $C\colon U^n \to R^n$ be a circuit such that, for at least $\alpha > 0$ fraction of inputs $u \in U^n$, the circuit $C(u)$ correctly computes $F^n(u)$ in at least γn positions, where $\alpha > (32/(\gamma-\delta)) \cdot \exp(-(\gamma-\delta)^2 n/1024)$. Then the algorithm $\mathcal{A}$, on inputs $n, \alpha, \gamma, \delta, C$, runs in time $\mathrm{poly}(|C|, 1/\alpha, (\ln 1/(\gamma-\delta)))$, and outputs a circuit A such that, with probability at least $1-o(1)$, A computes F on at least $\delta + (\gamma-\delta)/20$ fraction of inputs in U.*

Our proof of Theorem 5.4 will proceed in two stages. In the **first stage**, we give a simple probabilistic construction of a new circuit $C'\colon U^{n/2} \to R^{n/2}$ such that, with constant probability, C' is a "zero-error version" of the algorithm C in the following sense: either C' outputs $\perp$ (which we interpret as "don't know"), or C' outputs a tuple which is correct in at least γ' fraction of positions. More precisely, we show that, conditioned on C' outputting an answer other than $\perp$, it outputs a tuple with γ' fraction of correct positions with probability close to 1. Moreover, C' outputs a non-$\perp$ answer on at least about α fraction of inputs.

Note the difference between C and C'. The circuit C gives a good answer on at least α fraction of inputs, but may be arbitrarily bad on other inputs. In contrast, C' almost never gives a bad answer, but may just say "don't know" for some inputs. Moreover, C' will give a good answer on at least about α fraction of inputs. Thus, in some sense, C' filters out the bad answers of C and keeps only good answers. We should point out that the new circuit C' computes the direct product of F of half the size, i.e., $F^{n/2}$. So we trade the size of the direct product for the "zero-error" property. (Our reduction from C to C' may be viewed as a simple alternative to the "Trust Halving Strategy" of [IW97, BIN97].)

In the **second stage**, we use our "zero-error" circuit C' to compute F. The algorithm A is simple: "Given an input $x \in U$, we randomly embed it in a $n/2$-size tuple $v \in U^{n/2}$, and run $C'(v)$. If $C'(v)$ produces a definite answer, we output $C'(v)_x$. Otherwise, we randomly embed our x into a new tuple v', and repeat the above for this new v'. We continue sampling tuples v for at most $\text{poly}(1/\alpha, (\ln 1/(\gamma - \delta))$ iterations. If still no answer is produced, we output $\perp$."

The analysis of the above algorithm is fairly standard. It is based on a sampling lemma, due to Raz [Raz98], showing that the distribution obtained by sampling a random m-tuple $(x_1, \dots, x_m)$ from a subset $T \subseteq U^m$ of measure μ, and then outputting x_i for a uniformly random position $i \in [m]$ is statistically close to the uniform distribution on U. Using this sampling lemma, we can show that the success probability of our algorithm described above is essentially the same as when we first sample a random tuple v on which C' outputs a definite answer, and then pick a random position $i \in [m]$ and output $C'(v)_i$. With probability close to 1, our picked tuple v is such that $C'(v)$ is correct in at least γ' fraction of positions. Conditioned on picking such a tuple, the probability of producing a correct output is at least γ'. Overall, we get that the success probability of our algorithm is close to $\gamma' > \delta$.

The two stages are described in more detail and analyzed in Sections 5.2.1 and 5.2.2 below. Combining them, we get the proof of Theorem 5.4 at the end of Section 5.2.2.

Remark 5.5. The two-stage approach described above was used in [IJK09b] to give intuition why their algorithm works. However, the actual proof argument in [IJK09b] was considerably more complicated. Here we use our Theorem 4.4 to show that this intuitive and simple two-stage approach can indeed be implemented.

5.2.1 Trading DP size for "zero error"

We now define and analyze our reduction from C to C'. As usual, we define the Boolean random variables $X_1, \dots, X_n$, dependent on F, C, and a random n-tuple $u = (u_1, \dots, u_n) \in U^n$, so that, for each $1 \leqslant i \leqslant n$, $X_i = 1$ iff $C(u)_i = F(u_i)$ (i.e., X_i indicates whether C "succeeded" on input u_i).

For α satisfying the assumption of Theorem 4.4, we can apply Theorem 4.4 to these $X_1, \dots, X_n$. Let γ' be the midpoint in the interval $[\delta, \gamma]$. We get that, with probability at least 1/2 over random subsets $S \subseteq [n]$ of size $n/2$,

$$\mathbf{Pr}\left[(2/n)\sum_{i \in \bar{S}} X_i \geqslant \gamma' \mid (2/n)\sum_{j \in S} X_j \geqslant \gamma\right] \geqslant 1 - (\gamma - \delta)/4. \tag{9}$$

For simplicity of notation, let us assume that $\bar{S} = [n/2]$ (and so $S = \{n/2 + 1, \dots, n\}$). Also, let $m = n/2$.

Consider the following **algorithm C':**

> "On input $v = (v_1, \dots, v_m) \in U^m$, randomly sample an m-tuple $w = (w_1, \dots, w_m)$ and test if $C(v, w)_w = F^m(w)$ in at least γ fraction of positions. If the test is satisfied, then output $C(v, w)_v$, and halt. Otherwise, re-sample w. If no acceptable w is sampled within $t_1 = (4/\alpha) \cdot \ln 8/\alpha$ trials, then output some default (say, $\perp$) answer, and halt."

The following lemma summarizes the properties of C'.

Lemma 5.6. *With probability at least* $1/2$ *over the choice of random* $S \subset [n]$*, the algorithm* C' *is such that*

1. $\mathbf{Pr}[C'(v) = F^m(v)$ *in at least* γ' *fraction of positions* $\mid C'(v) \neq \bot] \geqslant 1 - (\gamma - \delta)/4$,
2. $\mathbf{Pr}[C'(v) \neq \bot] \geqslant \alpha/8$*;*

where, in both cases, the probability is over a uniformly random $v \in U^m$ *and internal randomness of* C'.

Proof. (1): To simplify the notation, let us introduce the events $\mathcal{E}$: $(2/n)\sum_{j\in S} X_j \geqslant \gamma$ and $\mathcal{D}$: $(2/n)\sum_{i\in\bar{S}} X_i \geqslant \gamma'$. Let $\mathcal{E}(v)$ be the event event $\mathcal{E}$ conditioned on v (i.e., conditioned on the fixed randomness of the variables X_i for $i \in \bar{S}$).

For a given input v, let us say that $C'(v)$ *succeeds* if $C'(v)$ agrees with $F^m(v)$ in at least γ' fraction of positions. Conditioned on $C'(v) \neq \bot$, the output of $C'(v)$ is distributed exactly like $C(v,w)_v$ for a uniformly random w satisfying $\mathcal{E}(v)$. That is, for a fixed v, we have

$$\mathbf{Pr}[C'(v) \text{ succeeds } \mid C'(v) \neq \bot] = \mathbf{Pr}_w[C(v,w)_v \text{ is correct in at least } \gamma' \text{ positions } \mid \mathcal{E}(v)].$$

Thus, over a random input v, we get $\mathbf{Pr}_v[C'(v) \text{ succeeds } \mid C'(v) \neq \bot] = \mathbf{Pr}[\mathcal{D} \mid \mathcal{E}]$, which is at least $1 - (\gamma-\delta)/4$ by Eq. (9).

(2): Define the subset

$$G = \{v \in U^m \mid \mathbf{Pr}_{w\in U^m}[C(v,w)_w = F^m(w) \text{ in at least } \gamma \text{ fraction of positions }] \geqslant \alpha/4\}.$$

Note that, for $v \in G$, the algorithm $C'(v)$ is unlikely to timeout. By our choice of $t_1 = (4/\alpha)\cdot \ln 8/\alpha$, we ensure that the probability of timeout on an input $v \in G$ is at most $(1-\alpha/4)^{t_1} \leqslant e^{-t_1\alpha/4} \leqslant \alpha/8$.

It remains to lowerbound the size of G. We have $\mathbf{Pr}[\mathcal{E}] = \mathbf{Exp}_v[\mathbf{Pr}_w[\mathcal{E}(v)]]$. Note that $\mathbf{Pr}[\mathcal{E}] \geqslant \alpha/2$ by Claim 4.6. Hence, by Markov, we get that for at least $\alpha/4$ of v's, $\mathbf{Pr}_w[\mathcal{E}(v)] \geqslant \alpha/4$. Thus, $\mathbf{Pr}_v[v \in G] \geqslant \alpha/4$.

Finally, we get that $\mathbf{Pr}_v[C'(v) \neq \bot] \geqslant \mathbf{Pr}_v[v \in G \ \& \ C'(v) \neq \bot] = \mathbf{Pr}_v[C'(v) \neq \bot \mid v \in G] \cdot \mathbf{Pr}_v[v \in G] \geqslant (1-\alpha/8)\cdot \alpha/4 \geqslant \alpha/8$. □

5.2.2 Using the "zero error" DP algorithm

Here we show how to compute the function F, given a "zero-error" DP circuit C' from the previous subsection. Recall that the statistical distance between two probability distributions D_1 and D_2 over the same finite universe U is defined as half the ℓ_1-norm of the difference $D_1 - D_2$, i.e., $(1/2)\cdot\sum_{x\in U}|D_1(x) - D_2(x)|$, where $D_i(x)$ is the probability of x under the distribution D_i, for $i = 1, 2$. We will use the following sampling lemma is implicit in [Raz98] (see, e.g., [IJK09a] for the proof).

Lemma 5.7 (Sampling Lemma [Raz98])**.** *Let* $G \subseteq U^m$ *be any set of measure at least* ϵ*. Let* D *be the distribution on* U *defined as follows: Pick a uniformly random* m*-tuple* $(x_1, \ldots, x_m) \in G$*, then pick a uniformly random index* $i \in [m]$*, and output* x_i*. Then the statistical distance between* D *and the uniform distribution over* U *is at most* $\sqrt{2(\ln 1/\epsilon)/m}$.

Consider the following **algorithm** A:

"On input $x \in U$, pick a random $i \in [m]$, pick a random m-tuple $v \in U^m$ containing x in position i, and run the circuit $C'(v)$. If $C'(v) \neq \bot$, then output $C'(v)_i$, and halt. Otherwise, re-sample i and v. If no output is produced within $t_2 = (128/\alpha) \cdot \ln 120/(\gamma - \delta)$ iterations, then output $\bot$, and halt."

We will show that A computes F well on average.

Lemma 5.8. $\mathbf{Pr}_x[A(x) = F(x)] \geqslant \delta + (\gamma - \delta)/20$.

Proof. We have

$$\mathbf{Pr}_x[A(x) = F(x)] = \mathbf{Pr}_x[A(x) = F(x) \mid A(x) \text{ doesn't time out}] \cdot \mathbf{Pr}_x[A(x) \text{ doesn't time out}].$$

In what follows, we lowerbound both probabilities on the right-hand side of the equation above.

For a fixed $x \in U$, $\mathbf{Pr}[A(x) = F(x) \mid A(x) \text{ doesn't time out}] = \mathbf{Pr}_{i \in [m], v \in U^m}[C'(v)_i = F(v_i) \mid C'(v) \neq \bot]$. Thus,

$$\mathbf{Pr}_{x \in U}[A(x) = F(x) \mid A(x) \text{ doesn't time out}] = \sum_{x \in U} \frac{1}{|U|} \mathbf{Pr}_{i \in [m], v \in U^m}[C'(v)_i = F(v_i) \mid C'(v) \neq \bot].$$

Now define the distribution D on U as follows: Sample a uniformly random $i \in [m]$, a uniformly random m-tuple $v \in U^m$ such that $C'(v) \neq \bot$, and output v_i. By Lemma 5.6, the measure of the set $\{v \in U^m \mid C'(v) \neq \bot\}$ is at least $\alpha/8$. By Lemma 5.7, D is close to the uniform distribution over U, where the statistical distance between the two is at most $\sqrt{2(\ln 8/\alpha)/m}$. By the assumption, $\alpha > (32/(\gamma - \delta)) \cdot e^{-(\gamma-\delta)^2 n/1024}$. Then the statistical distance between D and the uniform distribution over U is at most $\sqrt{2(\gamma - \delta)^2 (2m)/(1024m)} = (\gamma - \delta)/16$.

We get that $\sum_{x \in U}(1/|U|) \cdot \mathbf{Pr}_{i \in [m], v \in U^m}[C'(v)_i = F(v_i) \mid C'(v) \neq \bot]$ is at least

$$\sum_{x \in U} D(x) \cdot \mathbf{Pr}_{i \in [m], v \in U^m}\left[C'(v)_i = F(v_i) \mid C'(v) \neq \bot\right] - \sum_{x \in U} \left|1/|U| - D(x)\right|.$$

By the above, the second term is at most $(\gamma - \delta)/8$. By the definition of D, the first term is exactly equal to $\mathbf{Pr}_{i \in [m], v \in U^m}[C'(v)_i = F(v_i) \mid C'(v) \neq \bot]$, which is at least $\gamma'(1 - (\gamma - \delta)/4)$ by Lemma 5.6. Overall, we get

$$\mathbf{Pr}_{x \in U}[A(x) = F(x) \mid A(x) \text{ doesn't time out}] \geqslant \gamma'(1 - (\gamma - \delta)/4) - (\gamma - \delta)/8 \geqslant \delta + (\gamma - \delta)/8, \quad (10)$$

where we used the definition of $\gamma' = \delta + (\gamma - \delta)/2$.

Finally, we upperbound the probability that A times out. To this end, we define the set $H = \{x \in U \mid D(x) \leqslant 1/(16|U|)\}$. Since the statistical distance between D and the uniform distribution is at most $(\gamma - \delta)/16$, we get that the $|H|/|U| \leqslant (\gamma - \delta)/15$.

For each $x \in U$, let $p_x = \mathbf{Pr}_{i \in [m], v \in U^m}[C'(v) \neq \bot \mid v_i = x]$. We have $\mathbf{Exp}_x[p_x] = \mathbf{Pr}_{v \in U^m}[C'(v) \neq \bot] \geqslant \alpha/8$. On the other hand, $D(x) = \mathbf{Pr}_{i \in [m], v \in U^m}[v_i = x \mid C'(v) \neq \bot]$. It is not hard to see that $D(x) = p_x/(\sum_{x'} p_{x'})$. Hence, $p_x = D(x) \cdot |U| \cdot \mathbf{Pr}_{v \in U^m}[C'(v) \neq \bot]$. It follows that, for every $x \notin H$, we have $p_x \geqslant \alpha/128$.

For each $x \notin H$, the timeout probability is then at most $(1 - \alpha/128)^{t_2} \leqslant e^{-t_2 \alpha/128}$, where t_2 is the number of iterations of our algorithm A. By our choice of $t_2 = (128/\alpha) \cdot \ln 120/(\gamma - \delta)$, we ensure that the timeout probability is at most $(\gamma - \delta)/120$.

Overall, we get

$$\mathbf{Pr}_x[A(x) \text{ times out}] \leqslant \mathbf{Pr}_x[A(x) \text{ times out} \mid x \notin H] + \mathbf{Pr}_x[x \in H] \leqslant (\gamma - \delta)/120 + (\gamma - \delta)/15. \quad (11)$$

By Eqs. (10) and (11), we get $\mathbf{Pr}_x[A(x) = F(x)] \geqslant \delta + (\gamma - \delta)(1/8 - 1/120 - 1/15) = \delta + (\gamma - \delta)/20$. □

Now we can finish the proof of Theorem 5.4.

Proof of Theorem 5.4. We get our algorithm $\mathcal{A}$ by combining the algorithms of the two stages described above. Putting together Lemmas 5.6 and 5.8, we get with constant probability (at least 1/2) an algorithm A such that $\mathbf{Pr}_{x \in U}[A(x) = F(x)] \geqslant \delta + (\gamma - \delta)/20$.

Since we assume that we have an efficient way to generate pairs $(x, F(x))$ for uniformly random $x \in U$, we can test if our produced algorithm A is good for F by estimating its success probability (through random sampling). If A is not good, then we sample a different A. After a small number of trials, we get a good algorithm A with very high probability. □

6 Summary

Here we summarize some of the results mentioned in the paper. Let $X_1, \dots, X_n$ be Boolean random variables such that, for some $0 \leqslant \delta \leqslant 1$, $\mathbf{Pr}[X_i = 0] \leqslant \delta$, for all $1 \leqslant i \leqslant n$. Note that $bias(X_i) = \mathbf{Pr}[X_i = 0] - \mathbf{Pr}[X_i = 1] \leqslant \beta = 2\delta - 1$, for $1 \leqslant i \leqslant n$. Consider the following statements.

1. $X_1, \dots, X_n$ 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. $\forall 0 \leqslant \delta \leqslant \gamma \leqslant 1$, $\mathbf{Pr}[X_1, \dots, X_n \text{ has } \geqslant \gamma n \text{ zeros }] \leqslant e^{-n \cdot D(\gamma \| \delta)}$.

Theorem 6.1. (1) ⇒ (2) ⇒ (3) ⇔ (4).

Proof. (1) ⇒ (2) is trivial. For (2) ⇒ (3), see Theorem 3.10. For (3) ⇒ (4), see Theorem 3.1 (the implication (4) ⇒ (3) is trivial). □

An analogous statement for direct product theorems is: optimal XOR Theorems ⇒ optimal DPTs ⇔ optimal TDTPs. 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.

[IJK09a] R. Impagliazzo, R. Jaiswal, and V. Kabanets. Approximately list-decoding direct product codes and uniform hardness amplification. *SIAM Journal on Computing*, 39(2):564–605, 2009.

[IJK09b] 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.

[IW97] R. Impagliazzo and A. Wigderson. P=BPP if E requires exponential circuits: Derandomizing the XOR Lemma. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, pages 220–229, 1997.

[JLR00] S. Janson, T. Luczak, and A. Rucinski. *Random Graphs.* John Wiley & Sons, Inc., New York, 2000.

[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.

[Sie01] A. Siegel. Median bounds and their application. *Journal of Algorithms*, 38:184–236, 2001.

[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.

ECCC ISSN 1433-8092
http://eccc.hpi-web.de