On Yao's XOR-Lemma*

Oded Goldreich† Noam Nisan‡ Avi Wigderson§

March 1995 (corrected Dec. 1995, March and Nov. 1998)

Abstract

A fundamental Lemma of Yao states that computational weak-unpredictability of functions gets amplified if the results of several independent instances are XOR together. We survey two known proofs of Yao's Lemma and present a third alternative proof. The third proof proceeds by first proving that a function constructed by *concatenating* the values of the function on several independent instances is much more unpredictable, with respect to specified complexity bounds, than the original function. This statement turns out to be easier to prove than the XOR-Lemma. Using a result of Goldreich and Levin and some elementary observation, we derive the XOR-Lemma.

*Work done in part while the authors were visiting BRICS, Basic Research in Computer Science, Center of the Danish National Research Foundation.

†Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. Partially supported by grant No. 92-00226 from the United States - Israel Binational Science Foundation (BSF), Jerusalem, Israel.

‡Institute for Computer Science, Hebrew University, Jerusalem, Israel.

§Institute for Computer Science, Hebrew University, Jerusalem, Israel.

1 Introduction

A fundamental Lemma of Yao states that computational weak-unpredictability of predicates gets amplified if the results of several independent instances are XOR together, analogously to the information theoretic wire-tape channel Theorem (cf., Wyner). By weak-unpredictability we mean that any efficient algorithm will fail to predict the predicate with probability beyond a stated bound, where the probability is taken over all possible inputs (say with uniform probability distribution). In particular, the lemma known as Yao's XOR Lemma asserts that if the predicate f is weakly-unpredictable (within some complexity bound) then $F(x_1, ..., x_t) \stackrel{\rm def}{=} \oplus_{i=1}^t f(x_i)$, for sufficiently large t, is almost unpredictable within a related complexity bound (i.e., algorithms of this complexity cannot do substantially better than flip a coin for the answer).

Yao stated the XOR Lemma in the context of one-way functions, where the predicate f is the composition of an easy to compute Boolean predicate and the inverse of the one-way function (i.e., $f(x) = b(g^{-1}(x))$, where g is a 1-1 one-way function and b is an easy to compute predicate). Clearly, this is a special case of the setting described above. Yet, the XOR Lemma is sometimes used within the more general setting described above (under the false assumption that proofs for this setting have appeared in the literature). Furthermore, in contrary to common beliefs, the lemma itself has not appeared in Yao's original paper "Theory and Applications of Trapdoor Functions" [9] (but rather in oral presentations of his work).

A proof of Yao's XOR Lemma has first appeared in Levin's paper [6]. Levin's proof is for the context of one-way functions and is carried through in a uniform model of complexity. The presentation of this proof in [6] is very succinct and does not decouple the basic approach from difficulties arising from the uniform-complexity model. In Section 3, we show that Levin's basic approach suffices for the general case (mentioned above) provided it is stated in terms of non-uniform complexity. The proof also extends to a uniform-complexity setting, provided that some sampling condition (which is satisfied in the context of one-way functions) holds. We do not know whether the XOR Lemma holds in the uniform-complexity model in case this sampling condition is not satisfied.

Recently, Impagliazzo has shown that, in the non-uniform model, any weakly-unpredictable predicate has a "hard-core"[1] on which it is almost unpredictable [5]. Using this result, Impagliazzo has presented an alternative proof for the general case of the XOR-Lemma within the non-uniform model. We present this proof in Section 4.

A third proof for the general case of the XOR-Lemma is presented in Section 5. This proof proceeds by first proving that a function constructed by *concatenating* the values of the predicate on several independent instances is much more unpredictable, with respect to specified complexity bounds, than the original predicate. Loosely speaking, it is hard

[1] Here the term 'hard-core' means a subset of the predicate's domain. This meaning is certainly different from the usage of the term 'hard-core' in [3], where it means a strongly-unpredicatable predicate associated with a one-way function.

to predict the value of the function with probability substantially higher than δ^t, where δ is a bound on the probability of predicting the predicate and t is the number of instances concatenated. Not surprisingly, this statement turns out to be easier to prove than the XOR-Lemma. Using a result of Goldreich and Levin [3] and some elementary observation, we derive the XOR-Lemma.

We remark that Levin's proof yields a stronger quantitative statement of the XOR Lemma than the other two proofs. In fact, the quantitative statement provided by Levin's proof is almost optimal. Both Levin's proof and ours can be transformed to the uniform-complexity provided some natural sampling condition holds. We do not know how to transform Impagliazzo's proof to the uniform-complexity setting, even under this condition.

A different perspective on the concatenating problem considered above is presented in Section 6 where we consider the conditional entropy of the function's value given the result of a computation (rather than the probability that the two agree).

2 Formal Setting

The basic framework consists of a Boolean predicate $f:\{0,1\}^* \mapsto \{0,1\}$ and a non-uniform complexity class such as $\mathcal{P}/\text{poly}$. Specifically, we consider all families of polynomial-size circuits and for each family, $\{C_n\}$, we consider the probability that it correctly computes f, where the probability is taken over all n-bit inputs with uniform probability distribution. Alternatively, one may consider the most successful n-bit input circuit among all circuits of a given size. This way we obtain a bound on unpredictability of f with respect to a specific complexity class.

In the sequel, it will be more convenient to redefine f as mapping bit string into $\{\pm 1\}$ and to consider the correlation of a circuit (outputting a value in $\{\pm 1\}$) with the function (i.e., redefine $f(x) \stackrel{\text{def}}{=} (-1)^{f(x)}$). [2] Also, we generalize the treatment to arbitrary distributions over the set of n-bit long inputs (rather than uniform ones) and to "probabilistic" predicates (or processes) that on input x return some distribution on $\{\pm 1\}$ (i.e., for a fixed x, $f(x)$ is a random variable distributed over $\{\pm 1\}$ rather than a fixed value). One motivation for this generalization is that it allows us to treat as a special case 'hard predicates' of one-way functions, when the functions are not necessarily 1-1.

Definition 1 (algorithmic correlation): *Let P be a randomized process/algorithm that maps bit strings into values in $\{\pm 1\}$ and let $\mathbf{X} \stackrel{\text{def}}{=} \{X_n\}$ be a probability ensemble so that, for each n, the random variable X_n is distributed over $\{0,1\}^n$. The* correlation *of a circuit family $\mathbf{C} = \{C_n\}$ with P over $\mathbf{X}$ is defined as $c:\mathbb{N} \mapsto \mathbf{R}$ so that*

$$c(n) \stackrel{\text{def}}{=} \mathrm{E}[C_n(X_n) \cdot P(X_n)]$$

[2] This suggestion, of replacing the standard $\{0,1\}$ by $\{\pm 1\}$ and using correlations rather than probabilities, is due to Levin. It is indeed amazing how this simple change of notation simplifies both the statements and the proofs.

where the expectation is taken over the random variable X_n (and the process P). We say that a complexity class (i.e., set of circuit families) *has* correlation *at most $c(\cdot)$ with P over $\mathbf{X}$ if, for every circuit family $\mathbf{C}$ in this class, the correlation of $\mathbf{C}$ with P over $\mathbf{X}$ is bounded by $c(\cdot)$.*

The above definition may be used to discuss both uniform and non-uniform complexity classes. In the next subsection we relate the above definition to the standard treatment of unpredictability within the context of one-way functions.

The context of one-way functions

For sake of simplicity, we consider only length-preserving functions (i.e., functions $f : \{0,1\}^* \mapsto \{0,1\}^*$ satisfying $|f(x)| = |x|$ for all x). A *one-way function* $f : \{0,1\}^* \mapsto \{0,1\}^*$ is a function that is easy to compute but hard to invert. Namely, there exists a polynomial-time algorithm for computing f, but for any probabilistic polynomial-time[3] algorithm A, the probability that $A(f(x))$ is a preimage of $f(x)$ is negligible (i.e., smaller than $1/p(|x|)$ for any positive polynomial p), where the probability is taken uniformly over all $x \in \{0,1\}^n$ and all possible internal coin tosses of algorithm A.

Let $\delta : \mathbb{N} \mapsto \mathbf{R}$. The predicate $b : \{0,1\}^* \mapsto \{\pm 1\}$ is said to be *at most δ-correlated to f in polynomial-time* if b is easy to compute (i.e., there exists a polynomial-time algorithm for computing b), but $b(x)$ is hard to predict from the value of the function (i.e., from $f(x)$). Namely, for any probabilistic polynomial-time algorithm G, the expected correlation of $G(f(x))$ and $b(x)$, is at most $\delta(n)$ (for all but finitely many n's). (Again, the probability space is uniform over all $x \in \{0,1\}^n$ and all possible internal coin tosses of the algorithm.)

Suppose, first, that f is 1-1 and that b is an easy to compute predicate. Then, saying that b is at most δ-correlated to f in polynomial-time is equivalent to saying that the class of (probabilistic) polynomial-time algorithms has correlation at most $\delta(\cdot)$ with the predicate $P(x) \stackrel{\text{def}}{=} b(f^{-1}(x))$, over the uniform distribution. Note that in this case, although P is assumed not to be polynomial-time computable, it is easy to generate randomly pairs $(y, P(y))$ for randomly distributed y's. This is done, by uniformly selecting $r \in \{0,1\}^n$ and outputting the pair $(f(r), b(r)) = (f(r), P(f(r)))$.

The treatment is extended to one-way functions which are not necessarily 1-1 as follows. Let f be such a function and b a predicate which is at most δ-correlated to f (by polynomial-time algorithms). Define the probability ensemble $\mathbf{X} = \{X_n\}$ by letting $X_n = f(r)$, where r is uniformly selected in $\{0,1\}^n$ and define the randomized process $P(x)$ by uniformly selecting $r \in f^{-1}(x)$ and outputting $b(r)$. Now, it follows that the class of (probabilistic) polynomial-time algorithms has correlation at most $\delta(\cdot)$ with the predicate P over $\mathbf{X}$. Again, although P is not polynomial-time computable, it is easy to generate randomly pairs $(x, P(x))$, with distribution identical to $(X_n, P(X_n))$, where $n = |x|$.

[3]Here we adopt the standard definition of one-way function; however, our treatment applies also to the general definition where inverting is infeasible with respect to a specified time bound and success probability.

Getting random examples

An important issue regarding the general setting, is whether it is possible to get random examples of the distribution $(X_n, P(X_n))$. As mentioned in the previous subsection, in the context of one-way functions such random examples can be generated by a (uniform) probabilistic polynomial-time algorithm. On the other hand, the effect of such random examples can be easily simulated by non-uniform polynomial-size circuits (i.e., random/typical examples can be hard-wired into the circuit). Random examples are needed in all known proofs of the XOR Lemma (i.e., they are used in the algorithms deriving a contradiction to the difficulty of correlating the basic predicate). Thus, we can prove the XOR Lemma both in the general non-uniform complexity setting and in the (uniform-complexity) context of one-way functions.

Three (non-uniform) forms of the XOR Lemma

Following the description in the introduction (and Yao's expositions), the basic form of the XOR Lemma states that the tractable correlation of the XOR-predicate $P^{(t)}(x_1, ..., x_t) \stackrel{\text{def}}{=} \prod_{i=1}^{t} P(x_i)$ decays exponentially with t (upto a negligible fraction). Namely,

Lemma 1 (XOR Lemma – Yao's version): *Let P and $\mathbf{X} = \{X_n\}$ be as in Definition 1, $s : \mathbb{N} \mapsto \mathbb{N}$ be a size function, and $\delta : \mathbb{N} \mapsto [-1, +1]$ be a function that is* bounded-away-from-1 (i.e., $|\delta(n)| < 1 - \frac{1}{p(n)}$, for some polynomial p and all sufficiently large n's). *For every function $t : \mathbb{N} \mapsto \mathbb{N}$, define the predicate*

$$P^{(t)}(x_1, ..., x_{t(n)}) \stackrel{\text{def}}{=} \prod_{i=1}^{t(n)} P(x_i) \ ,$$

where $x_1, ..., x_{t(n)} \in \{0,1\}^n$, and let $\mathbf{X}^{(t)} \stackrel{\text{def}}{=} \{X_n^{(t)}\}$ be a probability ensemble such that $X_n^{(t)}$ consists of $t(n)$ independent copies of X_n.

(hypothesis) *Suppose that δ is a bound on the correlation of families of $s(\cdot)$-size circuits with P over $\mathbf{X}$.*

(conclusion) *Then, there exists a bounded-away-from-1 function δ' and a polynomial p such that, for every function $\epsilon : \mathbb{N} \mapsto [0, +1]$, the function*

$$\delta^{(t)}(n) \stackrel{\text{def}}{=} p(n) \cdot \delta'(n)^{t(n)} + \epsilon(n)$$

is a bound on the correlation of families of $s'(\cdot)$-size circuits with $P^{(t)}$ over $\mathbf{X}^{(t)}$, where

$$s'(t(n) \cdot n) \stackrel{\text{def}}{=} \text{poly}\left(\frac{\epsilon(n)}{n}\right) \cdot s(n) - \text{poly}(n \cdot t(n))$$

All three proofs presented below establish Lemma 1. The latter two proofs do so for various values of δ' and p (i.e., in Impagliazzo's proof $\delta'(n) = \frac{1+\delta(n)}{2} + o(1-\delta(n))$ and $p(n) = 2$, whereas in our proof $\delta'(n) = \sqrt[3]{\frac{1+\delta(n)}{2}}$ and $p(n) = o(n)$). Levin's proof does even better; it establishes the following

Lemma 2 (XOR Lemma – Levin's version): *Yao's version holds with $\delta' = \delta$ and $p = 1$.*

Lemma 2 still contains some slackness; specifically, the closest one wants to get to the "obvious" bound of $\delta^{(t)}(n) = \delta'(n)^{t(n)}$, the more one losses in terms of the complexity bounds (i.e., bounds on circuit size). In particular, if one insists on having $s'(t(n) \cdot n) = \frac{s(n)}{\text{poly}(n)}$ then one only gets a result only for $\epsilon(n) = 1/\text{poly}(n)$ (i.e., $\delta^{(t)}(n) = \delta'(n)^{t(n)} + 1/p(n)$, for any polynomial p). We do not know how to remove this slackness. We even do not know if it can be weaken "a little" as follows.

Lemma 3 (XOR Lemma – dream version – a conjecture): *For some fixed negligible function μ (e.g., $\mu(n) \stackrel{\text{def}}{=} 2^{-n}$ or even $\mu(n) \stackrel{\text{def}}{=} 2^{-(\log_2 n)^2}$), Yao's version holds with $\delta^{(t)}(n) = \delta'(n)^{t(n)} + \mu(n)$, and $s'(t(n) \cdot n) = \frac{s(n)}{\text{poly}(n)}$.*

Steven Rudich has observed that the Dream Version does not hold in a relativized world. Specifically, his argument proceeds as follows. Fix μ as in the Dream Version and set t so that $\delta^{(t)} < \mu(n)$. Consider an oracle that for every $(x_1, ..., x_{t(n)}) \in (\{0,1\}^n)^{t(n)}$ and for a $2\mu(n)$ fraction of the r's in $\{0,1\}^n$, answers the query $(x_1, ..., x_{t(n)}, r)$ with $(P(x_1), ..., P(x_t))$, otherwise the oracle answers with a special symbol. These r's may be selected at random (thus constructing a random oracle). The hypothesis of the lemma may hold relative to this oracle, but the conclusion cannot possibly hold. Put differently, one can argue that a (polynomial-time) "black-box" reduction of the task of correlating P (by $\geq \delta$) to the task of correlating $P^{(t)}$ (by $\geq 2\mu$) cannot possibly work. The reason being that the polynomial-time machine (effecting this reduction) cannot distinguish a black-box of negligible correlation (i.e., correlation 2μ) from a black-box of zero correlation.

Uniform forms of the XOR Lemma

Above, we have stated three forms of the XOR Lemma in terms of non-uniform complexity. Analogous statements in terms of uniform complexity can be made as well. These statements relate to the time required to construct the circuits in the hypothesis and those in the conclusion. For example, one may refer to circuit families, $\{C_n\}$, for which, given n, the circuit C_n can be constructed in poly($|C_n|$)-time. In addition, all functions referred to in the statement of the lemma (i.e., $s, t : \mathbb{N} \mapsto \mathbb{N}$, $\delta : \mathbb{N} \mapsto [-1, +1]$ and $\epsilon : \mathbb{N} \mapsto [-1, +1]$) need to be computable within corresponding time bounds. Analogues of the two first versions can be proven, provided that one can construct random examples of the distribution $(X_n, P(X_n))$ within the stated (uniform) complexity bounds (and in particular in polynomial-time). See comments in the subsequent sections.

3 Levin's Proof

The key ingredient in Levin's proof is the following lemma which provides an accurate account of the decrease of the computational correlation in case two predicates are xor-ed together. It should be stressed that the statement of the lemma is not symmetric with respect to the two predicates.

Lemma 4 (Isolation Lemma): *Let P_1 and P_2 be two predicates, $l : \mathbb{N} \mapsto \mathbb{N}$ be a length function, and $P(x) \stackrel{\rm def}{=} P_1(y) \cdot P_2(z)$ where $x = yz$ and $|y| = l(|x|)$. Let $\mathbf{X} = \{X_n\}$ be a probability ensemble so that the first $l(n)$ bits of X_n are statistically independent of the rest, and let $\mathbf{Y} = \{Y_{l(n)}\}$ (resp., $\mathbf{Z} = \{Z_{n-l(n)}\}$) denote the projection of $\mathbf{X}$ on the first $l(\cdot)$ bits (resp., last $n - l(n)$ bits).*

(hypothesis) *Suppose that $\delta_1(\cdot)$ is a bound on the correlation of families of $s_1(\cdot)$-size circuits with P_1 over $\mathbf{Y}$, and that $\delta_2(\cdot)$ is a bound on the correlation of families of $s_2(\cdot)$-size circuits with P_2 over $\mathbf{Z}$.*

(conclusion) *Then, for every function $\epsilon : \mathbb{N} \mapsto \mathbf{R}$, the function*

$$\delta(n) \stackrel{\rm def}{=} \delta_1(l(n)) \cdot \delta_2(n - l(n)) + \epsilon(n)$$

is a bound on the correlation of families of $s(\cdot)$-size circuits with P over $\mathbf{X}$, where

$$s(n) \stackrel{\rm def}{=} \min\left\{ \frac{s_1(l(n))}{\mathrm{poly}(n/\epsilon(n))} \;,\; s_2(n - l(n)) - n \right\}$$

The lemma is asymmetric with respect to the dependency of $s(\cdot)$ on the s_i's. The fact that $s(\cdot)$ maybe almost equal to $s_2(\cdot)$ plays a central role in deriving the XOR Lemma from the Isolation Lemma.

Proof: Assume to the contradiction that a circuit family $\mathbf{C}$ (of size $s(\cdot)$) has correlation greater than $\delta(\cdot)$ with P over $\mathbf{X}$. Thus, denoting by Y_l (resp., Z_m) the projection of X_n on the first $l \stackrel{\rm def}{=} l(n)$ bits (resp., last $m \stackrel{\rm def}{=} n - l(n)$ bits), we get

$$\begin{aligned} \delta(n) &< \mathrm{E}[C_n(X_n) \cdot P(X_n)] \\ &= \mathrm{E}[C_n(Y_l, Z_m) \cdot P_1(Y_l) \cdot P_2(Z_m)] \\ &= \mathrm{E}[P_1(Y_l) \cdot \mathrm{E}[C_n(Y_l, Z_m) \cdot P_2(Z_m)]] \end{aligned}$$

where, in the last expression, the outer expectation is over Y_l and the inner one is over Z_m. For every fixed $y \in \{0,1\}^l$, let

$$T(y) \stackrel{\rm def}{=} \mathrm{E}[C_n(y, Z_m) \cdot P_2(Z_m)]$$

Then, by the above,

$$\mathrm{E}[T(Y_l) \cdot P_1(Y_l)] > \delta(n) \tag{1}$$

We shall see that Eq. (1) either contradicts the hypothesis concerning P_2 (see Claim 4.1) or contradicts the hypothesis concerning P_1 (by a slightly more involved argument).

claim 4.1: For all but finitely many n's and every $y \in \{0,1\}^l$

$$|T(y)| \leq \delta_2(m)$$

proof: Otherwise, fixing a y contradicting the claim, we get a circuit $C'_m(z) \stackrel{\text{def}}{=} C_n(y,z)$ of size $s(n)+l < s_2(m)$, having greater correlation with P_2 than that allowed by the Lemma's hypothesis. □

By Claim 4.1, the value $T(y)/\delta_2(m)$ lies in the interval $[-1,+1]$; and, on the other hand (by Eq. (1)), it has good correlation with P_1. In the rest of the argument we "transform" the function T into a circuit which contradicts the hypothesis concerning P_1. Suppose for a moment, that one could compute $T(y)$, on input y. Then, one would get an algorithm with output in $[-1,+1]$ that has correlation $\delta(n)/\delta_2(m) > \delta_1(l)$ with P_1 over Y_l, which is almost[4] in contradiction to the hypothesis of the lemma. The same holds if one can approximate $T(y)$ "well enough" using circuits of size $s_1(l)$. Indeed, the lemma follows by observing that such an approximation is possible. Namely,

claim 4.2: For every n, $l = l(n)$, $m = n - l$, $q = \text{poly}(n/\epsilon(n))$ and $y \in \{0,1\}^l$, let

$$\tilde{T}(y) \stackrel{\text{def}}{=} \frac{1}{q}\sum_{i=1}^{q} C_n(y,z_i)\cdot\sigma_i$$

where $(z_1,\sigma_1), ..., (z_q,\sigma_q)$ is a sequence of q independent samples from the distribution $(Z_m, P_2(Z_m))$. Then,

$$\text{Prob}[|T(y) - \tilde{T}(y)| > \epsilon(n)] < 2^{-l(n)}$$

proof: immediate by the definition of $T(y)$ and application of Chernoff bound. □

The above claim suggests an approximation algorithm (for the function T), which is given as auxiliary input a sequence of samples from the distribution $(Z_m, P_2(Z_m))$. (The algorithm merely computes the average of $C_n(y,z_i)\cdot\sigma_i$ over the sample sequence $(z_1,\sigma_1), ..., (z_q,\sigma_q)$.) If such a sample sequence can be generated efficiently, by a uniform algorithm (as in the context of one-way functions), then we are done. Otherwise, we use non-uniformity to obtain a fixed sequence which is good for all possible y's. (Such a sequence does exist since with positive probability, a randomly selected sequence, from the above distribution, is good for all $2^{l(n)}$ possible y's.) Thus, there *exists* a circuit of size $\text{poly}(n/\epsilon(n)) \cdot s(n)$ that, on input $y \in \{0,1\}^{l(n)}$, outputs $(T(y) \pm \epsilon(n))/\delta_2(m)$. We note that this output is at least $\frac{\delta(n)}{\delta_2(m)} - \frac{\epsilon(n)}{\delta_2(m)} = \delta_1(l)$ correlated with P_1 which almost contradicts the hypothesis of the Lemma. The only problem is that the resulting circuit has output in the interval $[-1,+1]$

[4]See discussion below; what is "wrong" is that the output is in the $[-1,+1]$ interval rather than being a binary value in $\{\pm 1\}$.

instead of a binary output in $\{\pm 1\}$. This problem is easily corrected by modifying the circuit so that on output $r \in [-1,+1]$ it outputs $+1$ with probability $(1+r)/2$ and -1 otherwise. Noting that this modification preserves the correlation of the circuit, we derive a contradiction to the hypothesis concerning P_1. ■

The stronger version of the XOR Lemma (i.e., Lemma 2) follows by a (careful) successive application of the Isolation Lemma. Loosely speaking, we write $P^{(t)}(x_1, x_2, ..., x_{t(n)}) = P(x_1) \cdot P^{(t-1)}(x_2, ..., x_{t(n)})$, assume that $P^{(t-1)}$ is hard to correlate as claimed, and apply the Isolation Lemma to $P \cdot P^{(t-1)}$. This way, the lower bound on circuits correlating $P^{(t)}$ is related to the lower bound assumed for circuits correlating the original P and is almost the bound derived for $P^{(t-1)}$ (losing only an additive term!).

Remarks concerning the uniform complexity setting

A uniform-complexity analogue of Lemma 2 can be proven provided that one can construct random examples of the distribution $(X_n, P(X_n))$ within the stated (uniform) complexity bounds. To this end, one should state and prove a uniform-complexity version of the Isolation Lemma which also assumes that example from both distributions (i.e., $(Y_l, P_1(Y_l))$ and $(Z_m, P_2(Z_m))$) [5] can be generated within the relevant time complexity; certainly, sampleability in probabilistic polynomial-time suffices. Furthermore, in order to derive the XOR Lemma it is important to prove a strong statement regarding the relationship between the time required to construct the circuits referred to in the lemma. Namely,

Lemma 5 (Isolation Lemma – uniform complexity version): *Let $P_1, P_2, l, P, \mathbf{X}, \mathbf{Y}$ and $\mathbf{Z}$ be as in Lemma 4.*

(hypothesis) *Suppose that $\delta_1(\cdot)$ (resp., δ_2) is a bound on the correlation of $t_1(\cdot)$-time-constructible families of $s_1(\cdot)$-size (resp., $t_2(\cdot)$-time-constructible families of $s_2(\cdot)$-size) circuits with P_1 over $\mathbf{Y}$ (resp., P_2 over $\mathbf{Z}$). Furthermore, suppose that one can generate in polynomial-time a random sample from the distribution $(Y_l, Z_m, P_2(Z_m))$.*

(conclusion) *Then, for every function $\epsilon : \mathbb{N} \mapsto \mathbf{R}$, the function*

$$\delta(n) \stackrel{\text{def}}{=} \delta_1(l(n)) \cdot \delta_2(n - l(n)) + \epsilon(n)$$

is a bound on the correlation of $t(\cdot)$-time-constructible families of $s(\cdot)$-size circuits with P over $\mathbf{X}$, where

$$s(n) \stackrel{\text{def}}{=} \min\left\{\frac{s_1(l(n))}{\text{poly}(n/\epsilon(n))} \;,\; s_2(n - l(n)) - n\right\}$$

and

$$t(n) \stackrel{\text{def}}{=} \min\{t_1(l(n)) \;,\; t_2(n - l(n))\} - \text{poly}(n/\epsilon(n)) \cdot s(n)$$

[5] Actually, it suffices to be able to sample the distributions Y_l and $(Z_m, P_2(Z_m))$.

The uniform-complexity version of the Isolation Lemma is proven by adapting the above proof as follows. First, a weaker version of Claim 4.1 is stated, asserting that (for all but finitely many n's)

$$\text{Prob}[|T(Y_l)| > \delta_2(m) + \epsilon'(n)] < \epsilon'(n)$$

where $\epsilon'(n) \stackrel{\text{def}}{=} \epsilon(n)/3$. The new claim is valid, since otherwise, one can find in poly$(n/\epsilon(n))$-time a y violating it; to this end we need to sample Y_l and, for each sample y, approximate the value of $T(y)$ (by using poly$(n/\epsilon(n))$ samples of $(Z_m, P_2(Z_m))$). Once a good y is found, we incorporate it in the construction of C_n, resulting in a circuit which contradicts the hypothesis concerning P_2. (We stress that we have presented an efficient algorithm for constructing a circuit for P_2, given an algorithm that constructs the circuit C_n. Furthermore, the running time of our algorithm is the sum of the time required to construct C_n and the time required for sampling $(Z_m, P_2(Z_m))$ sufficiently many times and for evaluating C_n on sufficiently many instances.)

Clearly, Claim 4.2 remains unchanged (except for the replacing $\epsilon(n)$ by ϵ'). Using the hypothesis that samples from $(Z_m, P_2(Z_m))$ can be efficiently generated, we can construct a circuit for correlating P_1 within time $t(n)+\text{poly}(n/\epsilon(n))\cdot(n+s(n))$. This circuit is merely an approximater of the function T which operates by averaging (as in Claim 4.2); this circuit is constructed by first constructing C_n, generating poly$(n/\epsilon(n))$ samples of $(Z_m, P_2(Z_m))$ and incorporating them in corresponding copies of C_n – thus justifying the above time and size bounds. However, unlike in the non-uniform case, we are not guaranteed that $|T(y)|$ is bounded above (by $\delta_2(m) + \epsilon'(n)$) for all y's. Yet, if we modify our circuit to do nothing whenever its estimate violates the bound, we loss at most $\epsilon'(n)$ of the correlation and we can proceed as in the non-uniform case.

As in the non-uniform case, the (strong form of the) XOR Lemma follows by a (careful) successive application of the Isolation Lemma. Again, we write $P^{(\tau)}(x_1, x_2, ..., x_{\tau(n)}) = P(x_1) \cdot P^{(\tau-1)}(x_1, ..., x_{\tau(n)-1})$, assume that $P^{(\tau-1)}$ is hard to correlate as claimed, and apply the Isolation Lemma to $P \cdot P^{(\tau-1)}$. This way, the lower bounds on circuits correlating $P^{(\tau)}$ is related to the lower bound assumed for circuits correlating the original P and is almost the bound derived for $P^{(\tau-1)}$ (losing only an additive terms!). This almost concludes the proof, except that we have implicitly assumed that we know the value of τ for which the XOR Lemma first fails; this value is needed in order to construct the circuit violating the hypothesis for the original P. In the non-uniform case this value of τ can be incorporated into the circuit, but in the uniform-complexity case we need to find it. This is not a big problem as they are only polynomially many possible values and we can test each of them within the allowed time complexity.

4 Impagliazzo's Proof

The key ingredient in Impagliazzo's proof is the notion of a hard-core of a weakly-unpredictable predicate and a lemma that asserts that every weakly-unpredictable predicate has a hard-core of substantial size.

Definition 2 (hard-core of a predicate): *Let $f:\{0,1\}^* \mapsto \{0,1\}$ be a Boolean predicate, $s:\mathbb{N}\mapsto\mathbb{N}$ be a size function, and $\epsilon:\mathbb{N}\mapsto[0,1]$ be a function.*

- *We say that a sequence of sets, $\mathbf{S} = \{S_n \subseteq \{0,1\}^n\}$, is a* hard-core *of f with respect to $s(\cdot)$-size circuits families and advantage $\epsilon(\cdot)$ if for every n and every circuit C_n of size at most $s(n)$,*

$$\text{Prob}[C_n(X_n)=f(X_n)] \leq \frac{1}{2} + \epsilon(n)$$

where X_n is a random variable uniformly distributed on S_n.

- *We say that f has a* hard-core *of density $\rho(\cdot)$ with respect to $s(\cdot)$-size circuits families and advantage $\epsilon(\cdot)$ if there exists a sequence of sets $\mathbf{S} = \{S_n \subseteq \{0,1\}^n\}$ so that $\mathbf{S}$ is a hard-core of f with respect to the above and $|S_n| \geq \rho(n)\cdot 2^n$.*

We stress that the usage of the term 'hard-core' in the above definition (and in the rest of this section) is different from the usage of this term in [3]. Observe that every strongly-unpredictable predicate has a hard-core of density 1 (i.e., the entire domain itself). Impagliazzo proves that also weakly-unpredicatabe predicates have hard-core sets, but these have density related to the amount of unpredictability. Namely,

Lemma 6 (existence of hard-core for unpredictable predicates): *Let $f:\{0,1\}^* \mapsto \{0,1\}$ be a Boolean predicate, $s:\mathbb{N}\mapsto\mathbb{N}$ be a size function, and $\rho:\mathbb{N}\mapsto[0,1]$ be a non-negligible function* (i.e., $\rho(n) > 1/\text{poly}(n)$), *so that for every n and every circuit C_n of size at most $s(n)$ –*

$$\text{Prob}[C_n(U_n)=f(U_n)] \leq 1 - \rho(n)$$

where U_n is a random variable uniformly distributed on $\{0,1\}^n$. Then, for every function $\epsilon:\mathbb{N}\mapsto[0,1]$, the function f has a hard-core of density $\rho'(\cdot)$ with respect to $s'(\cdot)$-size circuits families and advantage $\epsilon(\cdot)$, where $\rho'(n) \stackrel{\text{def}}{=} (1-o(1))\cdot\rho(n)$ and $s'(n) \stackrel{\text{def}}{=} s(n)/\text{poly}(n/\epsilon(n))$.

The proof of Lemma 6 is given in the Appendix. Using Lemma 6, we derive a proof of the XOR-Lemma, for the special case of uniform distribution, as follows –

Suppose that $\delta(\cdot)$ is a bound on the correlation of $s(\cdot)$-circuits with f over the uniform distribution. Then, it follows that such circuits cannot guess the value of f better than with probability $p(n) \stackrel{\text{def}}{=} \frac{1+\delta(n)}{2}$ and the existence of a hard-core $\mathbf{S} = \{S_n\}$ (w.r.t. $s'(n)$-circuits and $\epsilon(n)$-advantage) with density $\rho'(n) \stackrel{\text{def}}{=} (1-o(1))\cdot(1-p(n))$ follows. Clearly,

$$\rho'(n) = (1-o(1))\cdot\frac{1-\delta(n)}{2} > \frac{1}{3}\cdot(1-\delta(n))$$

Now, suppose that in contradiction to the XOR Lemma, the predicate $F^{(t)}(x_1,...,x_t) \stackrel{\text{def}}{=} \oplus_i f(x_i)$ can be correlated by "small" circuits with correlation greater than $c'(n) \stackrel{\text{def}}{=} 2\,\cdot$

$(\frac{2+\delta(n)}{3})^t + \epsilon(n)$. In other words, such circuits can guess $F^{(t)}$ with success probability $\geq \frac{1}{2} + \frac{1}{2} \cdot c'(n)$. With probability at most $(1-\rho'(n))^t$ none of the t arguments to $F^{(t)}$ falls in the hard-core. Thus, conditioned on the event that at least one argument falls in the hard-core $\mathbf{S}$, the circuit guess $F^{(t)}$ correctly with probability at least

$$\frac{1}{2} + \frac{1}{2} \cdot c'(n) - (1-\rho'(n))^t > \frac{1}{2} + \frac{\epsilon(n)}{2}$$

For every non-empty $I \subseteq \{1, ..., t\}$, we consider the event, denoted E_I, that the arguments to $F^{(t)}$ which fall in the hard-core are eaxctly those with index in I. We have just shown that, conditioned on the union of these events, the circuit guesses the predicate $F^{(t)}$ correctly with probability at least $\frac{1}{2} + \frac{\epsilon(n)}{2}$. Thus, there exists an (non-empty) I so that, conditioned on E_I, the circuit guesses $F^{(t)}$ correctly with probability at least $\frac{1}{2} + \frac{\epsilon(n)}{2}$. Let $i \in I$ be arbitrary. By another averaging argument, we fix all inputs to the circuit except the i^{th} input and obtain a circuit which guesses f correctly with probability at least $\frac{1}{2} + \frac{\epsilon(n)}{2}$. (For these fixed x_j's, $j \neq i$, the circuit incorporates also the value of $\oplus_{j \neq i} f(x_j)$.) This contradicts the hypothesis that $\mathbf{S}$ is a hard-core.

We have just established the validity of the Lemma 1 for the case of the uniform probability ensemble and parameters $p(n) = 2$ and $\delta'(n) = \frac{2+\delta(n)}{3}$. The bound for δ' can be improved to $\delta'(n) = \frac{1+\delta(n)}{2} + o(1-\delta(n))$. The argument extends to arbitrary probability ensembles. To this end one needs to properly generalize Definition 2 and prove a generalization of Lemma 6 as done in the Appendix.

5 Going through the direct product problem

The third proof of the XOR Lemma proceeds in two steps. First it is shown that the success probability, of feasible algorithms which try to predict the values of a predicate on several unrelated arguments, decreases exponentially with the number of arguments. This statement is a generalization of another theorem due to Yao [9], hereafter called the *Concatenation Lemma*. Invoking a result of Goldreich and Levin [3], the XOR-Lemma follows.

The Concatenation Lemma

Lemma 7 (concatenation lemma): *Let P, $\mathbf{X} = \{X_n\}$, $s:\mathbb{N} \mapsto \mathbb{N}$, and $\delta:\mathbb{N} \mapsto [-1,+1]$ be as in Lemma 1. For every function $t:\mathbb{N} \mapsto \mathbb{N}$, define the function $F^{(t)}(x_1, ..., x_{t(n)}) \stackrel{\text{def}}{=} (P(x_1), ..., P(x_{t(n)}))$, where $x_1, ..., x_{t(n)} \in \{0,1\}^n$, and the probability ensemble $\mathbf{X}^{(t)} = \{X_n^{(t)}\}$, where $X_n^{(t)}$ consists of $t(n)$ independent copies of X_n.*

(hypothesis) *Suppose that δ is a bound on the correlation of families of $s(\cdot)$-size circuits with P over $\mathbf{X}$. Namely, suppose that for every n and for every $s(n)$-size circuit C:*

$$\mathrm{Prob}[C(X_n) = P(X_n)] \leq p(n) \stackrel{\text{def}}{=} \frac{1+\delta(n)}{2}$$

(conclusion) *Then, for every function* $\epsilon:\mathbb{N}\mapsto[0,+1]$, *for every* n *and for every* $\mathrm{poly}(\frac{\epsilon(n)}{n})\cdot s(n)$*-size circuit* C'*:*

$$\mathrm{Prob}[C'(X_n^{(t)})=F^{(t)}(X_n^{(t)})] \leq p(n)^{t(n)} + \epsilon(n)$$

Remark: Nisan et. al. [8] have used the XOR-Lemma in order to derive the Concatenation Lemma. Our feeling is that the Concatenation Lemma is more "basic" than the XOR Lemma, and thus that their strategy is not very natural. In fact, this feeling was our motivation for trying to find a "direct" proof for the Concatenation Lemma. Extrapolating from the situation regarding the two original lemmata of Yao (i.e., the XOR Lemma and the Concatenation Lemma w.r.t. one-way functions)[6], we believed that such a proof (for the Concatenation Lemma) should be easy to find. Indeed, we consider the following proof of Concatenation Lemma much simpler than the proofs of the XOR Lemma (given in previous sections).

Lemma 7 is derived from Lemma 8 (below) analogously to the way Lemma 2 was derived from Lemma 4; that is, we write $F^{(t)}(x_1, x_2, ..., x_{t(n)}) = (P(x_1), F^{(t-1)}(x_2, ..., x_{t(n)}))$, assume that $F^{(t-1)}$ is hard to guess as claimed, and apply the Concatenation Lemma to $(P, F^{(t-1)})$. This way, the lower bound on circuits guessing $F^{(t)}$ is related to the lower bound assumed for circuits guessing the original P and is almost the bound derived for $F^{(t-1)}$ (losing only an additive term!). It is thus left to prove the following:

Lemma 8 (two argument version of concatenation lemma): *Let* F_1 *and* F_2 *be two functions,* $l : \mathbb{N} \mapsto \mathbb{N}$ *be a length function, and* $F(x) \stackrel{\rm def}{=} (F_1(y), F_2(z))$ *where* $x = yz$ *and* $|y| = l(|x|)$. *Let* $\mathbf{X} = \{X_n\}$, $\mathbf{Y} = \{Y_{l(n)}\}$ *and* $\mathbf{Z} = \{Z_{n-l(n)}\}$ *be probability ensembles as in Lemma 4 (i.e.,* $X_n = (Y_{l(n)}, Z_{n-l(n)})$*).*

(hypothesis) *Suppose that* $p_1(\cdot)$ *is a bound on the probability that families of* $s_1(\cdot)$*-size circuits guess* F_1 *over* $\mathbf{Y}$. *Namely, for every such circuit family* $\mathbf{C} = \{C_l\}$

$$\mathrm{Prob}[C_l(Y_l)=F_1(Y_l)] \leq p_1(l)$$

Likewise, suppose that $p_2(\cdot)$ *is a bound on the probability that families of* $s_2(\cdot)$*-size circuits guess* F_2 *over* $\mathbf{Z}$.

[6] Yao's original XOR Lemma (resp., Concatenation Lemma) refers to the setting of one-way functions. In this setting, the basic predicate P is a composition of an easy to compute predicate b and the inverse of a 1-1 one-way function f; i.e., $P(x) \stackrel{\rm def}{=} b(f^{-1}(x))$. For years, the first author has considered the proof of the XOR Lemma (even for this setting) too complicated to be presented in class; whereas, a proof of the Concatenation Lemma (for this setting) has appeared in his classnotes [1] (see also [2]).

(conclusion) *Then, for every function* $\epsilon : \mathbb{N} \mapsto \mathbf{R}$*, the function* $p(n) \stackrel{\rm def}{=} p_1(l(n)) \cdot p_2(n - l(n)) + \epsilon(n)$ *is a bound on the probability that families of* $s(\cdot)$*-size circuits guess* F *over* $\mathbf{X}$*, where*

$$s(n) \stackrel{\rm def}{=} \min\left\{\frac{s_1(l(n))}{\mathrm{poly}(n/\epsilon(n))}\ ,\ s_2(n - l(n)) - n\right\}$$

Proof: Let $\mathbf{C} = \{C_n\}$ be a family of $s(\cdot)$-size circuits. Fix an arbitrary n, and write $C = C_n$, $\epsilon = \epsilon(n)$, $l = l(n)$, $m = n - l(n)$, $Y = Y_l$ and $Z = Z_m$. Abusing notation, we let $C_1(x,y)$ denote the first component of $C(x,y)$ (i.e., the guess for $F_1(x)$) and likewise $C_2(x,y)$ is C's guess for $F_2(y)$. It is instructive to write the success probability of C as follows:

$$\begin{aligned}\mathrm{Prob}[C(Y,Z)=F(Y,Z)] &= \mathrm{Prob}[C_2(Y,Z)=F_2(Z)] && (2)\\ &\quad \cdot \mathrm{Prob}[C_1(Y,Z)=F_1(Y) \mid C_2(Y,Z)=F_2(Z)] && (3)\end{aligned}$$

The basic idea is that using the hypothesis regarding F_2 allows to bound the first factor by $p_2(m)$, whereas the hypothesis regarding F_1 allows to bound the second factor by approximately $p_1(l)$. The basic idea for the latter step is that a sufficiently large sample of $(Z, F_2(Z))$, which may be hard-wired into the circuit, allows to use the conditional probability space (in such a circuit), provided the condition holds with noticeable probability. The last caveat motivates a separate treatment for y's with noticeable $\mathrm{Prob}[C_2(y,Z)=F_2(Z)]$ and for the rest.

We call y *good* if $\mathrm{Prob}[C_2(y,Z)=F_2(Z)] \geq \epsilon/2$ and *bad* otherwise. Let G be the set of good y's. Then, we bound the success probability of C as follows

$$\begin{aligned}\mathrm{Prob}[C(Y,Z)=F(Y,Z)] &= \mathrm{Prob}[C(Y,Z)=F(Y,Z)\ \&\ Y \in G] && (4)\\ &\quad + \mathrm{Prob}[C(Y,Z)=F(Y,Z)\ \&\ Y \notin G] && (5)\\ &< \mathrm{Prob}[C(Y,Z)=F(Y,Z)\ \&\ Y \in G] + \frac{\epsilon}{2} && (6)\end{aligned}$$

where the inequality follows by observing that for any $y \notin G$:

$$\mathrm{Prob}[C(Y,Z)=F(Y,Z)] \ \leq\ \mathrm{Prob}[C_2(Y,Z)=F_2(Z)] \ <\ \frac{\epsilon}{2}.$$

Thus, using $p(n) = p_1(l)p_2(m) + \epsilon$, it remains to prove that

$$\mathrm{Prob}[C(Y,Z)=F(Y,Z)\ \&\ Y \in G] \leq p_1(l) \cdot p_2(m) + \epsilon/2 \qquad (7)$$

We proceed according to the above outline. We first show that $\mathrm{Prob}[C_2(Y,Z)=F_2(Z)]$ cannot be too large, as otherwise the hypothesis concerning F_2 is violate. Actually, we prove claim 8.1: For every y,

$$\text{Prob}[C_2(y,Z)=F_2(Z)] \leq p_2(m)$$

proof: Otherwise, using any $y \in \{0,1\}^l$ so that $\text{Prob}[C_2(y,Z)=F_2(Z)] > p_2(m)$, we get a circuit $C'(z) \stackrel{\text{def}}{=} C_2(y,z)$ that contradicts the lemma's hypothesis concerning F_2. □

Next, we use claim 8.1 in order to relate the success probability of C to the success probability of small circuits for F_1.

claim 8.2: There exists a circuit C' of size $s_1(l)$ so that

$$\text{Prob}[C'(Y)=F_1(Y)] \;\geq\; \frac{\text{Prob}[C(Y,Z)=F(Y,Z)\ \&\ Y\in G]}{p_2(m)} - \frac{\epsilon}{2}$$

proof: The circuit C' is constructed as suggested in the outline above. Specifically, we take a poly(n/ϵ)-large sample, denoted S, from the distribution $(Z,F_2(Z))$ and let $C'(y) \stackrel{\text{def}}{=} C_1(y,z)$, where (z,β) is a uniformly selected among the elements of S for which $C_2(y,z)=\beta$ holds. Details follow.

Let S be a sequence of $t \stackrel{\text{def}}{=} \text{poly}(n/\epsilon)$ pairs, generated by taking t independent samples from the distribution $(Z,F_2(Z))$. We stress that we do not assume here that such a sample can be produced by an efficient (uniform) algorithm (but, jumping ahead, we remark that such a sequence can be fixed non-uniformly). For each $y \in G \subseteq \{0,1\}^l$, we denote by S_y the set of pairs $(z,\beta) \in S$ for which $C_2(y,z)=\beta$. Note that S_y is a random sample for the residual probability space defined by $(Z,F_2(Z))$ conditioned on $C_2(y,Z)=F_2(Z)$. Also, with overwhelmingly high probability, $|S_y| = \Omega(l/em^2)$ (since $y \in G$ implies $\text{Prob}[C_2(y,Z)=F_2(Z)] \geq \epsilon/2$). Thus, with overwhelming probability (i.e., probability greater than $1-2^{-l}$), taken over the choices of S, the sample S_y provides a good approximation to the conditional probability space, and in particular

$$\frac{|\{(z,\beta)\in S_y : C_1(y,z)=F_1(y)\}|}{|S_y|} \geq \text{Prob}[C_1(y,Z)=F_1(y) \mid C_2(y,Z)=F_2(Z)] - \frac{\epsilon}{2} \qquad (8)$$

Thus, with positive probability, Eq. (8) holds for all $y \in G \subseteq \{0,1\}^l$. The circuit C' guessing F_1 is now defined as follows. A set $S=\{z_i,\beta_i\}$ satisfying Eq. (8) for all good y's is "hard-wired" into the circuit C'. (In particular, S_y is not empty for any good y.) On input y, the circuit C' first determines the set S_y, by running C for t times and checking, for each $i=1,...,t$, whether $C_2(y,z_i)=\beta_i$. In case S_y is empty, the circuit returns an arbitrary value. Otherwise, the circuit selects uniformly a pair $(z,\beta)\in S_y$ and outputs $C_1(y,z)$. (This latter random choice can be eliminated by a standard averaging argument.) Using the definition of C', Eq. (8), claim 8.1, and some trivialities, we have:

$$\begin{aligned}\text{Prob}[C'(Y)=F_1(Y)] &\geq \sum_{y\in G} \text{Prob}[Y=y]\cdot\text{Prob}[C'(y)=F_1(y)] \\ &= \sum_{y\in G} \text{Prob}[Y=y]\cdot\frac{|\{(z,\beta)\in S_y : C_1(y,z)=F_1(y)\}|}{|S_y|}\end{aligned}$$

$$\begin{aligned}
&\geq \sum_{y\in G} \mathrm{Prob}[Y=y] \cdot \left(\mathrm{Prob}[C_1(y,Z)=F_1(y) \mid C_2(y,Z)=F_2(Z)] - \frac{\epsilon}{2}\right) \\
&\geq \left(\sum_{y\in G} \mathrm{Prob}[Y=y] \cdot \frac{\mathrm{Prob}[C_1(y,Z)=F_1(y) \;\&\; C_2(y,Z)=F_2(Z)]}{\mathrm{Prob}[C_2(y,Z)=F_2(Z)]}\right) - \frac{\epsilon}{2} \\
&\geq \left(\sum_{y\in G} \mathrm{Prob}[Y=y] \cdot \frac{\mathrm{Prob}[C(y,Z)=F(y,Z)]}{p_2(m)}\right) - \frac{\epsilon}{2}
\end{aligned}$$

where the last inequality is due to claim 8.1. The claim follows. $\square$

Now, by the lemma's hypothesis concerning F_1, we have $\mathrm{Prob}[C'(Y)=F_1(Y)] \leq p_1(l)$, and so using claim 8.2 we get

$$\begin{aligned}
\mathrm{Prob}[Y\in G \;\&\; C(Y,Z)=F(Y,Z)] &\leq (p_1(l)+\epsilon/2)\cdot p_2(m) \\
&\leq p_1(l)\cdot p_2(m)+\epsilon/2
\end{aligned}$$

This proves Eq. (7) and the lemma follows. ■

Deriving the XOR Lemma from the Concatenation Lemma

Using the techniques of [3], we obtain the following

Lemma 9 (Goldreich and Levin): *Let $F:\{0,1\}^* \mapsto \{0,1\}^*$, $p:\mathbb{N}\mapsto[0,1]$, and $s:\mathbb{N}\mapsto\mathbb{N}$, and let $\mathbf{X}=\{X_n\}$ be as in Lemma 1. For $\alpha,\beta\in\{0,1\}^\ell$, we denote by $IP_2(\alpha,\beta)$ the inner-product mod 2 of α and β viewed as binary vectors of length ℓ.*

(hypothesis) *Suppose that for every n and for every $s(n)$-size circuit C:*

$$\mathrm{Prob}[C(X_n)=F(X_n)] \leq p(n)$$

(conclusion) *Then, for some constant $c>0$, for every n and for every* $\mathrm{poly}(\frac{p(n)}{n})\cdot s(n)$-*size circuit C':*

$$\mathrm{Prob}[C'(X_n,U_\ell)=IP_2(F(X_n),U_\ell)] \leq \frac{1}{2} + c\cdot\sqrt[3]{n^2\cdot p(n)}$$

where U_ℓ denotes the uniform distribution over $\{0,1\}^\ell$, *with* $\ell \stackrel{\mathrm{def}}{=} |F(X_n)|$.
(That is, C' has correlation at most $2c\sqrt[3]{n^2 p(n)}$ with IP_2 over $(F(X_n),U_\ell)$.)

Proof Sketch: Let $q(n) \stackrel{\mathrm{def}}{=} c\sqrt[3]{n^2\,p(n)}$. Suppose that C' contradicts the conclusion of the lemma. Then, there exists a set S so that $\mathrm{Prob}[X_n\in S]\geq q(n)$ and for every $x\in S$ the probability that $C'(x,U_\ell)=\mathrm{IP}_2(F(x),U_\ell)$ is at least $\frac{1}{2}+\frac{q(n)}{2}$, where the probability is taken over U_ℓ (while x is fixed). Employing the techniques of [3], we obtain a circuit C (of size at most a poly$(n/p(n))$ factor larger than C') which, for every $x \in S$, outputs $F(x)$ with probability at least $c' \cdot (q(n)/n)^2$ (where the constant $c' > 0$ is determined in the proof of [3] according to Chebishev's Inequality).[7] Thus, C satisfies

$$\begin{aligned}
\mathrm{Prob}[C(X_n)=F(X_n)] &\geq \mathrm{Prob}[C(X_n)=F(X_n) \wedge X_n \in S] \\
&= \mathrm{Prob}[X_n \in S] \cdot \mathrm{Prob}[C(X_n)=F(X_n) | X_n \in S] \\
&\geq q(n) \cdot \left(c' \cdot (q(n)/n)^2\right) = p(n)
\end{aligned}$$

in contradiction to the hypothesis. The lemma follows. ■

Combining the Concatenation Lemma (Lemma 7) with Lemma 9 we establish the validity of Lemma 1 for the third time; this time with respect to the parameters $p(n) = cn^{2/3} = o(n)$ and $\delta'(n) = \sqrt[3]{\frac{1+\delta(n)}{2}}$. Details follow.

Starting with a predicate for which δ is a correlation bound and using Lemma 7, we get a function that is hard to guess with probability substantially higher than $(\frac{1+\delta(n)}{2})^{t(n)}$. Applying Lemma 9 establishes that given $(x_1, ..., x_{t(n)})$ and a uniformly chosen subset $S \subseteq \{1, 2, ..., t(n)\}$ it is hard to correlate $\oplus_{i \in S} P(x_i)$ better than with correlation

$$O\left(\sqrt[3]{n^2 \cdot \left(\frac{1+\delta(n)}{2}\right)^{t(n)}}\right) = o(n) \cdot \left(\sqrt[3]{\frac{1+\delta(n)}{2}}\right)^{t(n)}$$

This is almost what we need, but not quite (what we need is a statement concerning $S = \{1, ..., t(n)\}$). The gap is easily bridged by some standard "padding" trick. For example, by using a sequence of fixed pairs (z_i, σ_i), such that $\sigma_i = P(z_i)$, we reduce the computation of $\oplus_{i \in S} P(x_i)$ to the computation of $\oplus_{i=1}^{t(n)} P(y_i)$ (by setting $y_i = x_i$ if $i \in S$ and $y_i = z_i$ otherwise). Thus, Lemma 1 follows (with the stated parameters).

Remarks concerning the uniform complexity setting

A uniform-complexity analogue of the above proof can be carried out provided that one can construct random examples of the distribution $(X_n, P(X_n))$ within the stated (uniform) complexity bounds (and in particular in polynomial-time). Actually, this condition is required only for the proof of the Concatenation Lemma. Thus we confine ourselves to presenting a uniform-complexity version of the Concatenation Lemma.

Lemma 10 (Concatenation Lemma – uniform complexity version): *Let $P, \mathbf{X}, s, \delta, t$ and $F^{(t)}$ be as in Lemma 7.*

[7]The algorithm in [3] will actually retrieve all values $\alpha \in \{0,1\}^\ell$ for which the correlation of $C'(x, U_\ell)$ and $\mathrm{IP}_2(\alpha, U_\ell)$ is at least $q(n)$. With overwhelming probability it outputs a list of $O((n/q(n))^2)$ strings containing all the values just mentioned and thus uniformly selecting one of the values in the list yields $F(x)$ with probability at least $1/O((n/q(n))^2)$.

(hypothesis) *Suppose that $\delta(\cdot)$ is a bound on the correlation of $T(\cdot)$-time-constructible families of $s(\cdot)$-size circuits with P over $\mathbf{X}$. Furthermore, suppose that one can generate in polynomial-time a random sample from the distribution $(X_n, P(X_n))$.*

(conclusion) *Then, for every function $\epsilon : \mathbb{N} \mapsto [0, +1]$, the function $q(n) \stackrel{\text{def}}{=} p(n)^{t(n)} + \epsilon(n)$ is a bound on the correlation of $T'(\cdot)$-time-constructible families of $s'(\cdot)$-size circuits with F over $\mathbf{X}^{(t)}$, where $T'(t(n) \cdot n) = \text{poly}(\epsilon(n)/n) \cdot T(n)$ and $s'(t(n) \cdot n) = \text{poly}(\epsilon(n)/n) \cdot s(n)$.*

The uniform-complexity version of the Concatenation Lemma is proven by adapting the above proof as follows. Firstly, we observe that it suffices to prove an appropriate (uniform-complexity) version of Lemma 8. This is done by first proving a weaker version of Claim 8.1 which asserts that for all but at most an $\epsilon(n)/8$ measure of the y's (under Y),

$$\text{Prob}[C_2(y, Z) = F_2(Z)] \leq p_2(m) + \epsilon(n)/8$$

This holds since otherwise one may sample Y to find a y for which $\text{Prob}[C_2(y, Z) = F_2(Z)] > p_2(m)$ holds, and then use this y to construct (uniformly!) a circuit that contradicts the hypothesis concerning F_2. Next, we prove a weaker version of Claim 8.2 by observing that, for a uniformly selected pair sequence S, with overwhelmingly high probability (and not only with positive probability), Eq. (8) holds for all good $y \in \{0,1\}^l$. Thus, if we generate S by taking random samples from the distribution $(Z_m, F_2(Z_m))$, then with overwhelmingly high probability we end-up with a circuit as required by the modified claim. (The modified claim has $p_2(m) + \epsilon/8$ in the denominator (rather than $p_2(m)$) as well as an extra additive term of $\epsilon/8$.) Using the hypothesis concerning F_1, we are done as in the non-uniform case.

6 A Different Perspective: the Entropy Angle

The XOR Lemma and the Concatenation Lemma are special cases of the so-called "direct sum conjecture" asserting that computational difficulty increases when many independent instances of the problem are to be solved. In both cases the "direct sum conjecture" is postulated by considering insufficient resources and bounding the probability that these tasks can be performed within these resources, as a function of the number of instances. In this section we suggest an analogous analysis based on entropy rather than probability. Specifically, we consider the amount of information remaining in the task (e.g., of computing $f(x)$) when given the result of a computation (e.g., $C(x)$). This analysis turns out to be much easier.

Proposition 11 *Let f be a predicate, X be a random variable and $\mathcal{C}$ be a class of circuits so that for every circuit $C \in \mathcal{C}$*

$$H(f(X)|C(X)) \geq \epsilon$$

Furthermore, suppose that, for every circuit $C \in \mathcal{C}$, fixing any of the inputs of C yields a circuit also in $\mathcal{C}$. Then, for every circuit $C \in \mathcal{C}$

$$H(f(X^{(1)}),...,f(X^{(t)})|C(X^{(1)},...,X^{(t)})) \geq t \cdot \epsilon$$

where the $X^{(i)}$'s are independently distributed copies of X.

We stress that the class $\mathcal{C}$ in the above Proposition may contain circuits with several Boolean outputs. Furthermore, for a meaningful conclusion, the class $\mathcal{C}$ must contain circuits with t outputs (otherwise, for a circuit C with much fewer outputs, the conditional entropy $H(f(x_1),...,f(x_t)|C(x_1,...,x_t))$ is large merely due to information theoretical reasons). On the other hand, the more outputs the circuits in $\mathcal{C}$ have, the stronger the hypothesis of the Proposition is. In particular, the number of outputs must be smaller that $|X|$ otherwise the circuit $C(x) = x$ determines $f(x)$ (i.e., $H(f(x)|x) = 0$). Thus, a natural instantiation of the Proposition is for a family of small (e.g., poly-size) circuits each having t outputs.

Proof: By definition of conditional entropy, we have for every $C \in \mathcal{C}$,

$$\begin{aligned} H(f(X^{(1)}),...,f(X^{(t)})|C(X^{(1)},...,X^{(t)})) &= \sum_{i=1}^{t} H(f(X^{(i)})|C(X^{(1)},...,X^{(t)}),f(X^{(1)}),...,f(X^{(i-1)})) \\ &\geq \sum_{i=1}^{t} H(f(X^{(i)})|C(X^{(1)},...,X^{(t)}),X^{(1)},...,X^{(i-1)}) \end{aligned}$$

Now, for each i, we show that

$$H(f(X^{(i)})|C(X^{(1)},...,X^{(t)}),X^{(1)},...,X^{(i-1)}) \geq \epsilon$$

We consider all possible settings of all variables, except $X^{(i)}$, and bound the conditional entropy under this setting (which does not effect $X^{(i)}$). The fixed $X^{(j)} = x_j$ can be eliminated from the entropy condition and incorporated into the circuit. However, fixing some of the inputs in the circuit C yields a circuit also in $\mathcal{C}$ and so we can apply the proposition's hypothesis and get

$$H(f(X^{(i)})|C(x_1,...,x_{i-1},X^{(i)},x_{i+1},...,x_t)) \geq \epsilon$$

The proposition follows. ■

Additional Comments: Let $\mathcal{C}$ be a family of small (e.g., poly(n)-size) circuits each having t outputs and X be distributed over $\{0,1\}^n$.

- For $t = O(\log n)$, the hypothesis in the Proposition is related to the hypotheses used in all Lemmas above (and in particular in the Concatenation Lemma). Clearly, an entropy lower bound (on a single bit) translates to some unpredictability bound on this bit. (This does not hold for many bits as can be seen below.) The other direction (i.e., unpredictability implies a lower bound on the conditional entropy) is obvious for $t = 1$, but we need to consider the conditional entropy with respect to circuits which have $t > 1$ outputs. However, for each possible value of the circuit, there exists a value for f which is more likely. These 2^t values can be incorporated into the circuit, since $t = O(\log n)$, and so a small circuit offering small conditional entropy can be translated into a small circuit which predicts the value of f better than postulated.

- As for the conclusion of the Proposition, it is significantly weaker than the conclusion of Concatenation Lemma. In particular, it is possible that $C(x_1, ..., x_t)$ determines all the f values correctly with probability $1 - \epsilon$, and yields no information (e.g., outputs a special fail symbol) otherwise. Thus, although C may predict f simultaneously on many instances with probability $1 - \epsilon$, the conditional entropy is $(1 - \epsilon) \cdot 0 + \epsilon \cdot t$.

Acknowledgement

We wish to thank Mike Saks for useful discussions regarding Levin's proof of the XOR Lemma. We also thank Salil Vadhan and Ronen Shaltied for pointing out errors in previous versions, and for suggesting ways to fix these errors.

References

[1] O. Goldreich. *Foundation of Cryptography – Class Notes*, Spring 1989, Computer Science Department, Technion, Haifa, Israel.

[2] O. Goldreich. *Foundation of Cryptography – Fragments of a Book*, February 1995. Available from the *Electronic Colloquium on Computational Complexity* (*ECCC*), `http://www.eccc.uni-trier.de/eccc/`.

[3] O. Goldreich and L.A. Levin. "A Hard-Core Predicate for all One-Way Functions", in *ACM Symp. on Theory of Computing*, pp. 25–32, 1989.

[4] J. Hastad, R. Impagliazzo, L.A. Levin and M. Luby, "Construction of Pseudorandom Generator from any One-Way Function", manuscript, 1993. See preliminary versions by Impagliazzo et. al. in *21st STOC* and Hastad in *22nd STOC*.

[5] R. Impagliazzo, manuscript 1994.

[6] L.A. Levin, "One-Way Functions and Pseudorandom Generators", *Combinatorica*, Vol. 7, No. 4, 1987, pp. 357–363.

[7] L.A. Levin, "Average Case Complete Problems", *SICOMP*, Vol. 15, 1986, pp. 285–286.

[8] N. Nisan, S. Rudich, and M. Saks. in *35th FOCS*, 1994.

[9] A.C. Yao, "Theory and Application of Trapdoor Functions", in *23st FOCS*, pages 80–91, 1982.

A Proof of a Generalization of Lemma 6

We first generalize Impagliazzo's treatment to the case of non-uniform distributions; Impagliazzo's treatment is regained by letting $\mathbf{X}$ be the uniform probability ensemble.

Definition 3 (hard-core of a predicate relative to a distribution): *Let $f:\{0,1\}^* \mapsto \{0,1\}$ be a Boolean predicate, $s:\mathbb{N} \mapsto \mathbb{N}$ be a size function, $\epsilon:\mathbb{N} \mapsto [0,1]$ be a function, and $\mathbf{X}=\{X_n\}$ be a probability ensemble.*

- *We say that a sequence of sets, $\mathbf{S}=\{S_n \subseteq \{0,1\}^n\}$, is a* hard-core of f relative to $\mathbf{X}$ *with respect to $s(\cdot)$-size circuits families and advantage $\epsilon(\cdot)$ if for every n and every circuit C_n of size at most $s(n)$,*

$$\mathrm{Prob}[C_n(X_n)=f(X_n)|X_n\in S_n] \leq \frac{1}{2}+\epsilon(n)$$

- *We say that f has a* hard-core *of density $\rho(\cdot)$ relative to $\mathbf{X}$ with respect to $s(\cdot)$-size circuits families and advantage $\epsilon(\cdot)$ if there exists a sequence of sets $\mathbf{S}=\{S_n \subseteq \{0,1\}^n\}$ so that $\mathbf{S}$ is a hard-core of f relative to $\mathbf{X}$ with respect to the above and $\mathrm{Prob}[X_n\in S_n] \geq \rho(n)$.*

Lemma 12 (generalization of Lemma 6): *Let $f:\{0,1\}^* \mapsto \{0,1\}$ be a Boolean predicate, $s:\mathbb{N} \mapsto \mathbb{N}$ be a size function, $\mathbf{X}=\{X_n\}$ be a probability ensemble, and $\rho:\mathbb{N} \mapsto [0,1]$ be a non-negligible function, so that for every n and every circuit C_n of size at most $s(n)$,*

$$\mathrm{Prob}[C_n(X_n)=f(X_n)] \leq 1-\rho(n)$$

Then, for every function $\epsilon:\mathbb{N} \mapsto [0,1]$, the function f has a hard-core of density $\rho'(\cdot)$ relative to $\mathbf{X}$ with respect to $s'(\cdot)$-size circuits families and advantage $\epsilon(\cdot)$, where $\rho'(n) \stackrel{\text{def}}{=} (1-o(1))\cdot\rho(n)$ and $s'(n) \stackrel{\text{def}}{=} s(n)/\mathrm{poly}(n/\epsilon(n))$.

proof: We start by proving a weaker statement; namely, that $\mathbf{X}$ "dominates" an ensemble $\mathbf{Y}$ under which the function f is strongly unpredictable. Our notion of domination originates in a different work of Levin [7]. Fixing the function ρ we define domination as assigning probability mass which is at least a ρ fraction of the mass assigned by the dominated ensemble; namely

definition: We say that the ensemble $\mathbf{X}=\{X_n\}$ dominates the ensemble $\mathbf{Y}=\{Y_n\}$ if for every string α,

$$\mathrm{Prob}[X_n=\alpha] \geq \rho(|\alpha|)\cdot\mathrm{Prob}[Y_n=\alpha]$$

In this case we also say that $\mathbf{Y}$ is dominated by $\mathbf{X}$. We say that $\mathbf{Y}$ is critically dominated by $\mathbf{X}$ if for every string α either $\mathrm{Prob}[Y_n=\alpha]=(1/\rho(|\alpha|))\cdot\mathrm{Prob}[X_n=\alpha]$ or $\mathrm{Prob}[Y_n=\alpha]=0$.

(Actually, we allow at most one string $\alpha \in \{0,1\}^n$ to satisfy $0 < \text{Prob}[Y_n = \alpha] < (1/\rho(|\alpha|)) \cdot \text{Prob}[X_n = \alpha]$.)

The notion of domination and critical domination play a central role in the proof which consists of two parts. In the first part (cf., claim 12.1), we prove the existence of a ensemble dominated by $\mathbf{X}$ so that f is strongly unpredictable under this ensemble. In the second part (cf., claims 12.2 and 12.3), we essentially prove that the existence of such a dominated ensemble implies the existence of an ensemble which is *critically* dominated by $\mathbf{X}$ so that f is strongly unpredictable under this ensemble. However, such a critically dominated ensemble defines a hard-core of f relative to $\mathbf{X}$ and the lemma follows. Before starting, we make the following simplifying assumptions (used in claim 12.3).

simplifying assumptions: Without loss of generality

- $\text{Prob}[X_n = x] < \text{poly}(n)/s(n)$, for all x's.
  (Since x's violating this condition cannot contribute to the hardness of f with respect to X_n – as one can incorporate all these $s(n)/poly(n)$ many violating x's with their corresponding $f(x)$'s into the circuit).

- $\text{poly}(\epsilon(n)) > \text{poly}(n)/s(n)$.
  (Since otherwise the claim of the lemma holds vacuasly – as $s'(n) = \frac{s(n)}{\text{poly}(n/\epsilon(n))} < 1$).

claim 12.1: Let $T(n) = 1/\rho(n)$. Under the hypothesis of the lemma it holds that there exists a probability ensemble $\mathbf{Y} = \{Y_n\}$ so that $\mathbf{Y}$ is dominated by $\mathbf{X}$ and so that, for every $s'(n)$-circuit C_n, it holds

$$\text{Prob}[C_n(Y_n) = f(Y_n)] \leq \frac{1}{2} + \epsilon(n) \tag{9}$$

motivation: Suppose that for every $\mathbf{Y}$ dominated by $\mathbf{X}$, Eq. (9) does not hold (i.e., for every such Y_n there exists a small circuit C_n so that $\text{Prob}[C_n(Y_n) = f(Y_n)] > \frac{1}{2} + \epsilon(n)$). Suppose that we could have applied the min-max principle (which we can't since we do not have a bound on the behaviour of C_n's under all distributions – we only have a bound for $\mathbf{X}$-dominated distributions). We would have inferred that there is a randomized circuit R_n so that $\text{Prob}[R_n(x) = f(x)] > \frac{1}{2} + \epsilon(n)$, for every $x \in \{0,1\}^n$. By amplifying the advantage of the circuit R_n (using repeated trials) we would have obtained a circuit that for every $x \in \{0,1\}^n$, guesses correctly $f(x)$, with overwhelmingly high probability. This stands in contradiction to the hypothesis of the lemma. However, we cannot apply the min-max principle directly to $\mathbf{X}$-dominated distributions; instead, we consider an arbitrary superposition of $\mathbf{X}$-dominated distributions and apply the min-max principle in that context.

proof: Consider the (finite) set, denoted D, of all distributions that are critically dominated by X_n. Recall, that Y is *critically dominated* by X_n if for every α either $\text{Prob}[Y = \alpha] = T(n) \cdot \text{Prob}[X_n = \alpha]$ or $\text{Prob}[Y = \alpha] = 0$. (Actually, this is inaccurate and we should allow a single string α for which $0 < \text{Prob}[Y = \alpha] < T(n) \cdot \text{Prob}[X_n = \alpha]$.) Let $Y_1, Y_2, ..., Y_k$ be an enumeration of all the elements in D (i.e., enumeration of all the critically $\mathbf{X}$-dominated distributions) and let $\overline{p} = (p_i : i \in D)$ be an arbitrary probability distribution on the set D.

Let C be an arbitrary deterministic circuit. For every critically dominated distribution Y_i, denote by c_i the probability that C guesses f correctly on input Y_i. We now consider the expected value of c_i when i is selected according to the probability space $\overline{p}$. That is, we consider the average

$$\begin{aligned} \sum_i p_i \cdot c_i &= \sum_i p_i \cdot \text{Prob}[C(Y_i) = f(Y_i)] & (10) \\ &= \sum_i p_i \cdot \sum_y \text{Prob}[Y_i = y] \cdot \chi(C(y) = f(y)) & (11) \end{aligned}$$

where $\chi(B)$ is the indicator function of the Boolean expression B (i.e., $\chi(B) = 1$ if B holds and $\chi(B) = 0$ otherwise). Now, consider the r.h.s. of Eq. (11). Each y appears in the sum with weight $\sum_i p_i \cdot \text{Prob}[Y_i = y]$. Now, since all Y_i's are critically dominated by $\mathbf{X}$ it follows that, for each y and Y_i, the probability $\text{Prob}[Y_i = y]$ is either $T(n) \cdot \text{Prob}[X_n = \alpha]$ or zero. Denoting by D_y the subset of critically dominated distributions in which y appears with non-zero probability, we infer that each y appears in the sum with weight

$$\begin{aligned} \sum_i p_i \cdot \text{Prob}[Y_i = y] &= \sum_{i \in D_y} p_i \cdot T(n) \cdot \text{Prob}[X_n = y] \\ &= T(n) \cdot \text{Prob}[X_n = y] \cdot \sum_{i \in D_y} p_i \\ &\leq T(n) \cdot \text{Prob}[X_n = y] \end{aligned}$$

Thus, any probability distribution on the set of critically $\mathbf{X}$-dominated distributions induces an $\mathbf{X}$-dominated distribution. Therefore, assuming to the contrary of the claim that for every $\mathbf{X}$-dominated distribution there exists a s'-size circuit family $\{C_n\}$ violating Eq. (9), it follows that for every distribution $\overline{p}$ on D there exists a circuit C so that

$$\sum_i p_i \cdot \text{Prob}[C(Y_i) = f(Y_i)] > \frac{1}{2} + \epsilon(n) \qquad (12)$$

Now, we can apply the min-max principle (to Eq. (12)) and obtain a "randomized circuit" R (actually a distribution of circuits) which satisfies

$$\text{Prob}[R(Y) = f(Y)] > \frac{1}{2} + \epsilon(n) \qquad (13)$$

for every critically dominated distribution Y. We will use the "randomized circuit" R to derive a contradiction to the hypothesis of the lemma. We first denote by B the set of instances on which R performs badly; namely

$$B \stackrel{\text{def}}{=} \left\{ x : \text{Prob}[R(x) = f(x)] \leq \frac{1}{2} + \epsilon(n) \right\}$$

Now, $\text{Prob}[X_n \in B] < \frac{1}{T(n)}$ since otherwise we can define a critically dominated distribution Y for which Eq. (13) does not hold; suppose for simplicity that $\text{Prob}[X_n \in B] = \frac{1}{T(n)}$ then Y is defined by letting $\text{Prob}[Y = x] = T(n) \cdot \text{Prob}[X_n = x]$ if $x \in B$ and $\text{Prob}[Y = x] = 0$ otherwise. Now, using standard amplification techniques we derive a "randomized circuit" R' (actually a distribution over larger circuits) satisfying for each $x \notin B$

$$\text{Prob}[R'(x) = f(x)] > 1 - 2^{-n}$$

(Here we need to increase the size of the circuit by a factor of $\text{poly}(n/\epsilon(n))$.) Thus, with positive probability, a circuit selected by the distribution defining R' is correct on all x's in $\{0,1\}^n - B$, and so there exists such a circuit, denoted C'. This contradicts the hypothesis of the lemma since

$$\begin{aligned}
\text{Prob}[C'(X_n) = f(X_n)] &\geq \text{Prob}[X_n \notin B] \cdot \text{Prob}[C'(X_n) = f(X_n) | X_n \notin B] \\
&> \left(1 - \frac{1}{T(n)}\right)
\end{aligned}$$

Thus, the claim follows. $\Box$

In the rest of the proof, we fix an arbitrary ensemble, denoted $\mathbf{Y} = \{Y_n\}$ satisfying Claim 12.1. Using this ensemble, which is dominated by $\mathbf{X}$, we prove the validity of the lemma by a probabilistic argument. Specifically, we consider the following probabilistic construction.

probabilistic construction: We define a random set $R_n \subseteq \{0,1\}^n$ by selecting each string $x \in \{0,1\}^n$ to be in R_n with probability

$$p(x) \stackrel{\text{def}}{=} \frac{\rho(n) \cdot \text{Prob}[Y_n = x]}{\text{Prob}[X_n = x]} \leq 1 \tag{14}$$

independently of the choices made for all other strings. The inequality is due to the domination condition.

First we show that R_n is likely to be hit by X_n with the desired probability (i.e., $\text{Prob}[X_n \in R_n] \approx \rho(n)$).

claim 12.2: Let $\alpha > 0$ and suppose that $\text{Prob}[X_n = x] \leq \rho(n) \cdot \alpha^2/\text{poly}(n)$, for every x. Then,

$$|\text{Prob}[X_n \in R_n] - \rho(n)| < \alpha \cdot \rho(n)$$

for all but a $2^{-\text{poly}(n)}$ measure of the choices of R_n.

proof: Let $w_x \stackrel{\text{def}}{=} \text{Prob}[X_n = x]$, for every $x \in \{0,1\}^n$. We define random variables $\zeta_x = \zeta_x(R_n)$, over the probability space defined by the random choices of R_n, so that ζ_x indicate whether $x \in R_n$. Namely, the ζ_x's are independent of one another, $\text{Prob}[\zeta_x = 1] = p(x)$ and $\zeta_x = 0$ otherwise. Thus, for every choice of R_n we have

$$\text{Prob}[X_n \in R_n] = \sum_x \zeta_x(R_n) \cdot w_x$$

and consequently we are interested in the behaviour of the sum $\sum_x w_x \zeta_x$ as a random variable (over the probability space of all possible choices of R_n). Taking expactation over the possible choices of R_n, we get

$$\begin{aligned}
\mathrm{E}[\sum_x w_x \zeta_x] &= \sum_x p(x) \cdot w_x \\
&= \sum_x \frac{\rho(n) \cdot \mathrm{Prob}[Y_n = x]}{\mathrm{Prob}[X_n = x]} \cdot \mathrm{Prob}[X_n = x] \\
&= \rho(n)
\end{aligned}$$

Now, using Chernoff bound, we get

$$\mathrm{Prob}\left[\left|\sum_x w_x \zeta_x - \rho(n)\right| > \alpha \cdot \rho(n)\right] < \exp\left(-\Omega\left(\frac{\alpha^2 \rho(n)}{\max_x\{w_x\}}\right)\right)$$

Now, using the claim's hypotheses $w_x \leq \alpha^2 \cdot \rho(n)/\mathrm{poly}(n)$ (for all x's), the latter expression is bounded by $\exp(-\mathrm{poly}(n))$ and the claim follows. $\square$

Next we show that R_n is likely to be a hard-core of f realtive to $\mathbf{X}$ (w.r.t. sufficiently small circuits).

claim 12.3: Let C_n be a circuit of size $s'(n)$. Then,

$$\mathrm{Prob}[C_n(X_n) = f(X_n) | X_n \in R_n] < \frac{1}{2} + 2\epsilon$$

for all but a $2^{-(s'(n)^2+1)}$ measure of the choices of R_n.

proof: We define the same random variables $\zeta_x = \zeta_x(R_n)$ as in the proof of the previous claim; $\zeta_x(R_n) = 1$ if $x \in R_n$ and $\zeta_x(R_n) = 0$ otherwise. Also, as before, $w_x \stackrel{\mathrm{def}}{=} \mathrm{Prob}[X_n = x]$, for every $x \in \{0,1\}^n$. Let C be the set of inputs on which C_n correctly computes f; namely,

$$C \stackrel{\mathrm{def}}{=} \{x : C_n(x) = f(x)\}$$

For every choice of R_n, we are interested in the probability

$$\mathrm{Prob}[X_n \in C | X_n \in R_n] = \frac{\mathrm{Prob}[X_n \in C \wedge X_n \in R_n]}{\mathrm{Prob}[X_n \in R_n]} \tag{15}$$

We first determine the expected value of the numerator of Eq. (15), where the expactation is taken over the possible choices of R_n. We rewrite the numerator as $\sum_{x \in C} \zeta_x(R_n) \cdot w_x$, and bound it as follows

$$\begin{aligned}
\mathrm{E}[\sum_{x \in C} \zeta_x \cdot w_x] &= \sum_{x \in C} p(x) \cdot w_x \\
&= \sum_{x \in C} \frac{\rho(n) \cdot \mathrm{Prob}[Y_n = x]}{\mathrm{Prob}[X_n = x]} \cdot \mathrm{Prob}[X_n = x] \\
&= \rho(n) \cdot \mathrm{Prob}[Y_n \in C] \\
&\leq \rho(n) \cdot \left(\frac{1}{2} + \epsilon\right)
\end{aligned}$$

where the last inequality is due to the hypothesis regarding Y_n. Next, we use Chernoff bound and get

$$\mathrm{Prob}[\sum_{x\in C} w_x \zeta_x > (\frac{1}{2}+\frac{4\epsilon}{3})\cdot\rho(n)] < \exp\left(-\Omega\left(\frac{\epsilon^2\rho(n)}{\max_x\{w_x\}}\right)\right)$$

Now, using the simplifying assumptions regarding the w_x's and ϵ, the latter expression is bounded by $\exp(-\sqrt{s(n)/\mathrm{poly}(n)})$. Thus, for all but a $\exp(-s'(n)^2+2)$ measure of the R_n's the numerator of Eq. (15) is bounded above by $(\frac{1}{2}+\frac{4\epsilon}{3})\cdot\rho(n)$. Using the previous claim, we conclude that for a similar measure of R_n's the denumerator of Eq. (15) is bounded below by $(1-\frac{\epsilon}{3})\cdot\rho(n)$. The claim follows. □

The lemma now follows by combining the above three claims. Claim 12.1 provides us with a suitable $\mathbf{Y}$ for which we apply the probabilistic construction, whereas Claims 12.2 and 12.3 establish the existence of a set R_n such that both

$$\mathrm{Prob}[X_n \in R_n] > (1-o(1))\cdot\rho(n)$$

and

$$\mathrm{Prob}[C_n(X_n) = f(X_n) | X_n \in R_n] < \frac{1}{2}+2\epsilon$$

for all $2^{s'(n)^2}$ possible circuits, C_n, of size $s'(n)$. The lemma follows. ■