

# PAPER PRESENTATION FOR COMPLEXITY II

VIPUL NAIK

ABSTRACT. These are rough notes for a presentation of a paper titled “Hardcore distributions for somewhat hard problems” by Russell Impagliazzo (the original paper is available at <http://www-cse.ucsd.edu/~russell/hardcore.ps>).

## 1. BACKGROUND DEFINITIONS

**Definition.** • A **measure**<sub>(defined)</sub> on strings of length  $n$  is a function  $M$  that maps each string to some point in the interval  $[0, 1]$ . A measure can be viewed as a *fuzzy set* where the value on a string is the *extent* to which it is present.

Any convex combination of measures is a measure.

Given any subset of the set  $\{0, 1\}^n$ , the *characteristic function* on that subset is the measure that takes the value 1 at all strings in the subset and the value 0 at all other points.

- The **relative size**<sub>(defined)</sub> of a measure is the arithmetic mean of the values that the measure takes on all strings. The **absolute size**<sub>(defined)</sub> is the sum of the values taken on all strings.

$$\text{Relative size of } M = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} M(x)$$

$$\text{Absolute size of } M = \sum_{x \in \{0,1\}^n} M(x)$$

The absolute size of the characteristic function on a subset equals the cardinality of the subset. The relative size of the characteristic function on a subset equals the ratio of the cardinality of the subset to the cardinality of the set.

- Let  $1/2 > \delta > 0$  and let  $n \leq g \leq 2^n/n$ . We say  $f$  is  $\delta$  hard on  $D$  for size  $g$  if for any circuit  $C$  with at most  $g$  gates, and any probability distribution  $D$ , we have:

$$\text{Prob}_{x \text{ from } D} [f(x) = C(x)] \leq 1 - \delta$$

The condition  $n \leq g$  is because any circuit on  $n$  inputs must have at least  $g$  gates, while the condition  $g \leq 2^n/n$  is because any Boolean function  $\{0, 1\}^n \rightarrow \{0, 1\}$  can be given a circuit with  $2^n/n$  gates.

- The **distribution induced by a measure**<sub>(defined)</sub>  $M$  is defined as the distribution that at  $x$ , carries the weight  $M(x)/|M|$ .
- For a circuit  $C$  and an input  $x$  define  $R_C(x)$  as 1 if  $f(x) = C(x)$  and 0 otherwise. Then the **advantage**<sub>(defined)</sub> of  $C$  on  $M$  is defined as  $\sum_x M(x)R_C(x)$ . This advantage is denoted as  $\text{Adv}_C(M)$ .
- $f$  is termed  $\epsilon$ -**hardcore**<sub>(defined)</sub> on  $M$  for size  $g$  if for any circuit  $C$  with at most  $g$  gates, the advantage of  $C$  on  $M$  is less than  $\epsilon|M|$ . This is denoted as  $\text{Adv}_C(S)$ . When  $S$  comprises a single element  $x$ , we denote it as  $\text{Adv}_C(x)$ .
- $f$  is termed  $\epsilon$ -**hardcore**<sub>(defined)</sub> on  $S$  for size  $g$  (where  $S$  is a subset of the set of strings) if  $f$  is  $\epsilon$ -hardcore on the measure obtained as the characteristic measure on  $S$ .
- $f$  is termed  $\epsilon$ -**hardcore**<sub>(defined)</sub> for size  $g$  if  $f$  is hardcore on the set of all inputs on the same parameters.

## 2. THE THEOREM STATEMENT

### 2.1. Full statement.

**Theorem 1** (Theorem 1 of text). Let  $f$  be a Boolean function on  $n$ -bit inputs that is  $\delta$ -hard for circuits of size  $g$  on the uniform distribution, and let  $\epsilon > 0$ . Then there is a set  $S \subseteq \{0, 1\}^n$  so that  $|S| \geq \delta 2^n$  and  $f$  is  $\epsilon$ -hard-core on  $S$  for circuits of size  $d\epsilon^2\delta^2g$  where  $d$  is an absolute constant.

**2.2. Parsing of the theorem statement.** The theorem statement:

- Starts with a hardness assumption with respect to a given number  $\delta$ , for the class of circuits of a given size  $g$ .
- Ends to give a hardness result with respect to any arbitrary number  $\epsilon$  for the class of circuits of smaller size, where the smaller size is  $g$  times a constant depending on  $\delta$  and  $\epsilon$ . However, this hardness result is not over the whole collection of inputs, but over an appropriately chosen subset of size at least  $\delta 2^n$ .

Other paraphrasings of the result are:

- We can trade off the hardness (in the sense of the probability of error) against the circuit size, to reach arbitrarily low levels of hardness on large subsets.
- If a function is  $\delta$ -hard for all inputs, then we can locate a *responsible set* containing  $\delta$  times the total number of elements, for which the function is really really hard ( $\epsilon$ -hardcore for any  $\epsilon$ ).

Parametric explanation:

The *given*:

- We begin with hardness  $\delta$ , which could be *potentially large*, for circuits of size  $g$ .
- We name a hardness  $\epsilon$ , which we want should be small.

The *obtained*:

- A subset  $S$  that contains at least a  $\delta$  fraction of all inputs.
- A number  $g'$  that is  $g$  times a polynomial function in  $\delta$  and  $\epsilon$ .
- The subset is  $\epsilon$ -hardcore for circuits of size  $g'$  over the set  $S$ .

**Lemma 1** (Advantage and probability). The following are equivalent:

$$\text{Prob}_{x \in \{0,1\}^n} [f(x) = C(x)] - 1/2 = p$$

‘And: ‘

$$\text{Adv}_C(\{0, 1\}^n) = 2p - 1$$

**2.3. Finding the troublesome hardcore set.** Let’s view a simple probabilistic analogue of the statement. We are saying that no circuit of size  $g$  can really do well on  $f$  except with an advantage of  $\delta$  over random guessing. We are trying to now state that there will be *some* subset whose size is at least a  $\delta$  fraction, and on which the inputs are *much harder*.

### 3. PRELIMINARIES FOR THE MAX-MIN PROOF

**3.1. The von Neumann max-min theorem.** First, some preliminaries:

**Definition.** A **two person zero-sum game**<sub>(defined)</sub> involves the following data: Let  $P_1$  and  $P_2$  be two players.  $P_1$  has a finite set of strategies labelled  $\{1, 2, \dots, n\}$  and  $P_2$  has a finite set of strategies labelled  $\{1, 2, \dots, n\}$ . A payoff function  $A : \{1, 2, \dots, n\} \times \{1, 2, \dots, n\} \rightarrow \mathbb{R}$  is given.  $P_1$  and  $P_2$  both pick strategies from their respective strategy sets, and the payoff function on these is computed. The goal of  $P_1$  is to maximize the payoff while the goal of  $P_2$  is to minimize the payoff. Neither  $P_1$  nor  $P_2$  has any idea of the other person’s move, but they both know the payoff matrix.

For instance, consider the two person zero-sum game given by the matrix:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

2

Here, suppose  $P_1$  chooses strategy 1. Then, if  $P_2$  chooses strategy 1, the payoff is  $A_{11} = 1$ , and if  $P_2$  chooses strategy 2, the payoff is  $A_{22} = 2$ . Clearly,  $P_2$  would prefer to play strategy 1 in order to minimize the payoff.

Suppose  $P_1$  chooses strategy 2. Then, if  $P_2$  chooses strategy 1, the payoff is 3, while if  $P_2$  chooses strategy 2, the payoff is 4. Thus,  $P_2$  should play strategy 1 if  $P_1$  chooses strategy 2.

The upshot:  $P_2$  should choose Strategy 1 whatever strategy  $P_1$  chooses.

Similarly,  $P_1$ , in order to maximize the payoff, should choose Strategy 2 whatever strategy  $P_2$  chooses.

This means that both  $P_1$  and  $P_2$  have *unique optimal strategies* and if they both play their optimal strategies, the value of the game is 3. This discussion motivates a definition:

**Definition.** For any strategy  $i$  of  $P_1$  let the minimum value  $m_i$  denote the minimum possible payoff to  $P_1$  based on  $i$ . The **max-min value**<sub>(defined)</sub> of the game is the maximum over all strategies  $i$  of the minimum values  $m_i$ . Equivalently, the max-min value is the largest possible value that  $P_1$  can guarantee to achieve as payoff.

Similarly, let  $n_j$  denote the maximum possible payoff to  $P_2$  if he chooses strategy  $j$ . The **min-max value**<sub>(defined)</sub> of the game is the minimum over all strategies  $j$  of  $P_2$  of the maximum values  $n_j$ . The min-max value is thus the smallest possible value that  $P_2$  can ensure is not exceeded by the payoff.

Clearly, the max-min value of the game is less than or equal to the min-max value. Hence the max-min value is also termed the **lower value**<sub>(defined)</sub> and the min-max value is also termed the **upper value**<sub>(defined)</sub>.

The game we described above had the property that the max-min value was equal to the min-max value. In general, this is not true if each player is restricted only to picking pure strategies. However, if each player can pick *mixed strategies*, or probability distributions over the space of strategies, then the max-min value and min-max values are equal.

**Theorem 2** (von Neumann max-min theorem). For two-person zero sum games with finite strategy sets for both players, the lower and upper values of the game are equal if both players can choose mixed strategies.

An easy corollary of the theorem:

**Corollary 1.** Given a two-person zero sum game with finite strategy sets, and a real number  $r \in \mathbb{R}$ , there is either a strategy of  $P_1$  that ensures a payoff of  $r$  whatever  $P_2$  does, or there is a strategy of  $P_2$  that ensures the payoff is less than  $r$ , whatever  $P_1$  does. Which case occurs depends on whether  $r \leq v$  or not.

This corollary is the form in which we'll use the von Neumann max-min theorem.

### 3.2. Important Chernoff bounds.

**Fact 1** (Chernoff bounds for 0-1 random variable). Let  $X_i$  be independent 0-1 random variables, with  $\text{Prob}_{X_i} [X_i = 1] = p_i$ . Let  $\mu = \sum_i p_i$ . Then if  $X = \sum_i X_i$ :

$$(1) \quad \text{Prob}_{X_i} [X \leq (1 - d)\mu] \leq e^{-\mu d^2/3}$$

$$(2) \quad \text{Prob}_{X_i} [X \geq (1 + d)\mu] \leq e^{-\mu d^2/3}$$

**3.3. Distributions on the space of circuits.** Earlier on, we defined the advantage  $\text{Adv}_C(M)$  as the weighted linear combination over  $M$  of the  $1, -1$  indicator variables for  $C(x) = f(x)$ . Usually, we are

interested in a *fixed circuit* and measuring how good it is against variable inputs. However, there are also situations where we consider a collection of circuits all of which are supposed to compute the function  $f$ , and look at the advantage of the combination of these circuits on each input.

**3.4. Counting the number of circuits.** First, an easy bound on the number of circuits:

**Lemma 2** (Circuit bound). The number of circuits of size  $g$  is at most  $(2(2n + g))^{2g}$ .

*Proof.* □

We know that  $(2(2n + g))^{2g} \leq 2^{2ng} \leq (1/4) \exp 2^n \epsilon^2 \delta^2 / 2$

#### 4. THE MAX-MIN BASED PROOF

**4.1. Proof steps.** The max-min proof proceeds in three steps:

- The first step is *Lemma 2* of the paper. Define  $F : (0, 1) \times (0, 1) \rightarrow \mathbb{R}$  as follows:

$$F(\epsilon, \delta) := (\epsilon^2/16)(-\log(.5\epsilon\delta))^{-1}$$

Then given a function  $f$  that is  $\delta$ -hard for size  $g$  we have that, for any  $\epsilon > 0$ ,  $f$  is  $\epsilon$ -hard for size  $g' = g/F(\epsilon, \delta)$ .

- The second step is *Lemma 3*. This allows us to move from hard-core measures to hard-core subsets. It states that given an  $\epsilon/2$  hard-core measure for  $f$  of size  $g$  satisfying  $2n < g < (1/8)(2^n/n)(\epsilon\delta)^2$ . Assume that  $\mu(M) \geq \delta$ . Then there is an  $\epsilon$ -hardcore set  $S$  for  $f$  for size  $g$  with  $|S| \leq \delta 2^n$ .

**4.2. The statement of Lemma 2.**

**Lemma 3** (Lemma 2 of the paper). Let  $f$  be  $\delta$ -hard for size  $g$  on the uniform distribution on  $n$ -bit strings, and let  $1 > \epsilon > 0$ . Then there is a measure  $M$  with  $\mu(M) \geq \delta 2^n$  so that  $f$  is  $\epsilon$ -hardcore for size  $g'$  where  $g = g'F(\delta, \epsilon)$  on  $M$ . Here,

$$F(\delta, \epsilon) = 16\epsilon^{-1}(-\log(.5\epsilon\delta))$$

Because lemma 2 of the paper is a little tedious, I have broken it down into easier-to-digest pieces

**4.3. Proof of Lemma 2: applying the max-min theorem.** I begin by proving the following:

**Claim.** Given a number  $g'$  and a function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$ , and  $\epsilon, \delta \in (0, 1)$ . Then, exactly one of the following theorems holds:

- (1) There is a probability distribution over circuits of size  $g'$  such that for every subset  $S$  of size at least  $\delta 2^n$ , the weighted average of the advantages of the circuits over that subset is at least  $\epsilon$ .
- (2) There is a distribution  $M$  of size at least  $\delta 2^n$  such that for circuit of size  $g'$ , the advantage of the circuit on  $M$  is less than or equal to  $\epsilon$ .

*Proof.* Let's begin by defining a two-person zero-sum game. Assume that  $f, n, g, \delta, \epsilon$  are given as above and  $g'$  is as per the above formula. Then, Player 1 seeks a circuit  $C$  of size  $g'$  that purports to compute  $f$ , while Player 2 seeks a subset  $S$  of size at least  $\delta 2^n$ . The payoff for the game is  $Adv_C(M)$ . Player 1 seeks to maximize the advantage by choosing a *good* circuit while Player 2 seeks to minimize the advantage by seeking a *hard* subset.

By the corollary (corollary 1) to the von Neumann max-min theorem (theorem 2) we conclude that exactly one of the two holds:

- (1) Player 1 can choose a probability distribution over circuits of size  $g'$  such that the weighted sum of the advantages over these circuits is greater than  $\epsilon$  for *every* subset  $S$  of size less at least  $\delta 2^n$ . This corresponds exactly to possibility (1) mentioned in the claim.

- (2) Player 2 can choose a probability distribution over subsets of size at least  $\delta 2^n$  such that for any circuit of size  $g'$ , the weighted sum of the advantages of the circuit over these subsets is at most  $\epsilon$ .

A probability distribution over subsets of size at least  $\delta 2^n$  is equivalent to a distribution of size  $\delta 2^n$ , and making this translation we get possibility (2) of the claim.  $\square$

#### 4.4. What's left in Lemma 2.

**Claim.** Given a distribution of circuits of size  $g'$  that solves  $f$  with average advantage at least  $\epsilon$  for every subset of size at least  $\delta 2^n$ . Then, we can construct a single circuit of size  $g = g'F(\delta, \epsilon)$  whose overall error is at most  $\delta$ .

*Proof.* To show this, we again proceed in two steps:

- (1) We first show that the set of really hard inputs for the circuit distribution is not the entire set of inputs. More specifically, if  $S$  is the set of inputs for which the circuit distribution fails to predict correctly with probability more than  $1/2 + \epsilon/4$ , then the size of  $S$  is not more than  $\delta(1 - \epsilon/2)$ .
- (2) We then show, using Chernoff bounds, that by picking  $t$  independent random samples from this distribution of circuits and taking their majority, the probability that this is correct for any  $x \notin S$  is fairly high (around  $\exp -\epsilon^2 t/16$ ).
- (3) We then set  $t = 16\epsilon^2(-\log(.5\epsilon\delta))$ .

Let's first do (1).

**Claim.** If  $S$  is the set of inputs for which the circuit distribution fails to give the correct answer with probability more than  $(1/2 + \epsilon/4)$ , then  $|S| \leq \delta(1 - \epsilon/2)2^n$ .

*Proof.* Note that Player 1 has attempted to choose a good circuit distribution, that is, Player 1 has attempted to choose a circuit distribution that offers an advantage strictly greater than  $\epsilon$  on every subset of size  $\delta 2^n$ .

Suppose now that the set  $S$  has cardinality strictly greater than  $2^n \delta(1 - \epsilon)/2$ . Consider the set  $S'$  comprising the  $\delta 2^n$  inputs with smallest advantage.  $S'$  has strictly less than  $\delta \epsilon 2^{n-1}$  non-members of  $S$ . For these non-members, the total advantage is thus bounded by  $\delta \epsilon 2^{n-1}$  (because each input can give an advantage of at most 1). For the members, the advantage of each member is bounded above by  $\epsilon/2$  from lemma 1 and hence the total advantage from members of  $S$  is at most  $\epsilon \delta 2^{n-1}$ . Adding up, we obtain that the total advantage is less than  $\epsilon \delta 2^n$ . But this contradicts the fact that Player 1's circuit combination ensures an average advantage of at least  $\epsilon$  on every set of size  $\delta 2^n$ .  $\square$

Let's now do (2).

**Claim.** Suppose we pick an  $x \notin S$ . Then, taking  $t$  independent random samples from the circuit, the probability that the output is incorrect is at most  $e^{-\epsilon^2 t/16}$ .

*Proof.* If  $x \notin S$ , then that means that the probability that the circuit distribution yields the correct answer for  $x$  is at least  $1/2 + \epsilon/4$ . Consider the 0-1 random variable experiment that picks a circuit from the distribution, returns 0 if the circuit outputs the wrong answer, and 1 if the circuit outputs the correct answer. If we take  $t$  trials and then look at the majority answer of the trials, we can go wrong only if less than  $1/2$  of the trials are correct. Then:  $\mu = t(1/2 + \epsilon/4)$  and  $d = 1 - \frac{1/2}{1/2 + \epsilon/4}$  which simplifies to  $d = \frac{\epsilon}{2 + \epsilon}$ . We thus get

$$\text{Prob}[X \leq \mu(1 - d)] \leq e^{-(\epsilon/(2+\epsilon))^2 t(1/2 + \epsilon/4)/2}$$

The exponent on the right hand side simplifies to  $e^{-\epsilon^2/(2+\epsilon)(8)t}$ . This gives the answer.  $\square$

Define  $F(\delta, \epsilon) = 16\epsilon^{-1}(-\log(.5\epsilon\delta))$ . We now set  $t = F(\delta, \epsilon) = 16\epsilon^{-2}(-\log(.5\epsilon\delta))$ . Then, the value of  $e^{-\epsilon^2 t/16}$  becomes  $e^{\log(.5\epsilon\delta)}$  which is less than  $.5\epsilon\delta$ .

Thus, we have obtained a circuit of size  $g = g'F(\delta, \epsilon)$  that works as follows: it picks an input  $x$ , performs  $t = F(\delta, \epsilon)$  random experiments on  $x$  from the circuit distribution, and outputs the majority answer. For inputs not in  $S$ , this circuit goes wrong with probability at most  $\epsilon\delta$ . Moreover, the size of  $S$  is itself bounded by  $\delta(1 - \epsilon/2)$ . Thus, we have a circuit of size  $g$  whose overall error is bounded by  $\delta(1 - \epsilon/2) + \delta\epsilon/2$  which is  $\delta$ .  $\square$

**4.5. Piecing together lemma 2.** The two claims we made in the last two subsections show that:

- The first claim shows that if the assumption of Lemma 2 is false, then there exists a distribution of circuits of size  $g'$  that solves the problem with advantage at least  $\epsilon$ .
- The second claim shows that the existence of such a family contradicts  $\delta$ -hardness for size  $g$

**4.6. Getting a hard-core set from a hard-core measure.**

**Lemma 4** (Lemma 3 of paper). Let  $M$  be an  $\epsilon$  hard-core measure for  $f$  of size  $g$  satisfying  $2n < g < (1/8)(2^n/n)(\epsilon\delta)^2$ . Assume that  $\mu(M) \geq \delta$ . Then there is a  $2\epsilon$ -hardcore set  $S$  for  $f$  for size  $g$  with  $|S| \leq \delta 2^n$ .

*Proof.* We are given an  $\epsilon/2$ -hard-core measure for size  $g$  and function  $f$  and we have to obtain an  $\epsilon$ -hard-core subset.

The proof idea is to choose a set randomly according to the measure  $M$ . That is, given the measure  $M$ , choose a random subset  $S$  where, for each  $x$ :

$$\text{Prob}_S [x \in S] = M(x)$$

Let  $M_S$  be the characteristic function for  $S$ . Then the expectation of the random variable  $Adv_C(M_S)$  is precisely  $Adv_C(M)$ . By the assumption of  $\epsilon/2$ -hard-core-ness we get that  $\mathbf{E}[Adv_C(M_S)] \leq \epsilon\mu(M)$ .

Now,  $Adv_C(M_S)$  is the sum of  $2^n$  random variables each in the interval  $[0, 2^{-n}]$ . Hence by Chernoff bounds, we get:

$$\text{Prob}_{C,S} [Adv_C(M_S) \geq 2\epsilon\mu(M)] \leq e^{-2^n \epsilon^2 \delta^2 / 2}$$

Note that the circuit bound stated that the number of circuits of size  $g$  was less than  $1/4e^{2^n \epsilon^2 \delta^2 / 2}$ . Hence, taking the union bound over all circuits of size  $g$ , we get:

$$\text{Prob}_S [\exists C : Adv_C(M_S) \geq 2\epsilon\mu(M)] \leq 1/4$$

Thus, for at least  $3/4$  of the random subsets  $S$ , there is no circuit of size  $g$  that offers an advantage of more than  $2\epsilon\mu(M)$ .

On the other hand, by symmetry of the distribution, the probability that  $|S| \geq \mathbf{E}[|S|]$  is at least  $1/2$ . Since the events  $|S| \geq \mathbf{E}[|S|]$  and  $\forall C : Adv_C(M_S) \leq 2\epsilon\mu(M)$  have probabilities  $1/2$  and  $3/4$ , the probability that at least one of the events occurs is at least  $1/4$ . Hence, there is a set  $S$  for which  $|S| \geq \mu(M)$  and for which no circuit of size  $g$  offers advantage better than  $2\epsilon\mu(M)$ . This translates to an *average* advantage of  $2\epsilon$ .

Thus,  $S$  is a  $2\epsilon$ -hard-core subset for  $f$ .  $\square$

## 5. APPLICATIONS

### 5.1. Statement of the Yao XOR lemma.

**Theorem 3** (Yao XOR lemma). Given a Boolean function that is  $\delta$ -hard for circuits of size  $C$ , the function  $\bar{f}(x_1, x_2, \dots, x_k) = f(x_1) \oplus f(x_2) \dots \oplus f(x_k)$  is  $\epsilon + (1 - \delta)^k$  hard-core for circuits of size  $O(\epsilon^2 C)$ .

5.2. **A little lemma.** We first prove an additional lemma:

**Lemma 5.** If  $f$  is  $\epsilon$ -hard-core on a set of  $\delta 2^n$  inputs for size  $g$  then the function  $\bar{f}(x_1, x_2, \dots, x_k) = \sum_i f(x_i)$  is  $\epsilon + (1 - 2\delta)^k$ -hard-core for size  $g$ .

*Proof.* Suppose  $\bar{f}$  has a circuit of size  $g$  that achieves advantage more than  $\epsilon + (1 - \delta)^k$  for computing the XOR of  $k$  independent instances of the problem. We seek to use  $\bar{f}$  to compute  $f$ . The idea is to hardwire all except the specific  $x$  for which we want to compute  $f$ , compute the value of the circuit at the point, and then subtract the hardwired values of  $f$  at the hardwired points.

Suppose  $S$  is the set of size  $\delta 2^n$  inputs for which  $f$  is  $\epsilon$ -hard-core. Then the probability that none of the inputs is from  $S$  is  $(1 - \delta)^k$ . Let  $A_l$  be the advantage given that exactly  $l$  inputs are from the set  $S$ . Then, there is an  $l \leq k$  for which  $A_l \geq \epsilon$ . Then, consider a circuit  $C'$  that picks  $a_1, a_2, \dots, a_{l-1}$  from the hard-core and  $b_1, b_2, \dots, b_{k-l}$  from outside, and feeds these along with  $x$  to  $C$ . It then subtracts the values of  $f(a_i)$  and  $f(b_j)$  for all the  $a_i$ s and  $b_j$ s (which are hardwired). Consider the probability distribution over such circuits. The average advantage over this probability distribution is  $A_l \geq \epsilon$  and each  $C'$  is of size at most that of  $C$ . This contradicts the assumption that  $S$  was hard-core.  $\square$

5.3. **Putting the proof together.** We use Theorem 1 along with the above lemma to obtain a proof of the Yao XOR lemma. The proof proceeds in two steps:

- We begin by using Theorem 1 to move from  $\delta$ -hardness on circuits of size  $C$  to  $\epsilon$ -hardness of  $f$  for a subset of size  $\delta 2^n$  on circuits of size  $d\epsilon^2\delta^2C$  which is in  $O(\epsilon^2C)$ .
- We then use the above lemma to move from  $\epsilon$ -hardness over a subset of size  $2^n$  to  $\epsilon + (1 - \delta)^k$  hardcoreness for the new size (which is  $O(\epsilon^2C)$ ).