Electronic Colloquium on Computational Complexity, Report No. 105 (2019)

A note on the relation between XOR and Selective XOR Lemmas*

Ragesh Jaiswal**

Department of Computer Science and Engineering,
Indian Institute of Technology Delhi.

Abstract. Given an unpredictable Boolean function $f : \{0,1\}^n \to \{0,1\}$, the standard Yao's XOR lemma is a statement about the unpredictability of computing $\oplus_{i\in[k]} f(x_i)$ given $x_1, ..., x_k \in \{0,1\}^n$, whereas the Selective XOR lemma is a statement about the unpredictability of computing $\oplus_{i\in S} f(x_i)$ given $x_1, ..., x_k \in \{0,1\}^n$ and $S \subseteq \{1, ..., k\}$. We give a reduction from the Selective XOR lemma to the standard XOR lemma. Our reduction gives better quantitative bounds for certain choice of parameters and does not require the assumption of being able to sample $(x, f(x))$ pairs.

1 Introduction

A boolean function $f : \{0,1\}^* \to \{0,1\}$ is said to be $\delta(n)$-unpredictable for family of $s(n)$ size circuits iff for any any n and any circuit C of size $s(n)$, $\mathbf{Pr}_{x\leftarrow\{0,1\}^n}[C(x) \neq f(x)] \geq \delta(n)$, where $\delta : \mathbb{N} \to \mathbb{R}$ and $s : \mathbb{N} \to \mathbb{N}$ are functions over positive integers. Note that in the previous statement, the probability is over the uniform distribution on $\{0,1\}^n$. The unpredictability can be defined more generally on a probability ensemble. However, in this work we work with the uniform distribution since the ideas developed here can easily be generalised. Note that in case the circuit C above is a randomised circuit, the probability is also over the internal randomness of the circuit. The classical *Yao's XOR lemma* is a statement about unpredictability amplification by defining the XOR function that computes the xor of function values evaluated simultaneously at multiple inputs from $\{0,1\}^n$. More specifically, consider the function $F^{\oplus k}(x_1, ..., x_k) \equiv \oplus_i f(x_i)$, where k is a positive integer and $x_1, ..., x_k \in \{0,1\}^n$. It can be shown that the function F is $\varepsilon(n)$-unpredictable for family of $s'(n)$ size circuits, where $\varepsilon : \mathbb{N} \to \mathbb{R}$ and $s' : \mathbb{N} \to \mathbb{N}$ are functions such that $\varepsilon(n)$ is much smaller than $\delta(n)$ and $s'(n)$ is upper bounded by $s(n)$. A theorem that is closely related to Yao's XOR lemma is the *direct product lemma*, that is a statement about the unpredictability of the direct product function $F^{|k}(x_1, ..., x_k) \equiv (f(x_1), ..., f(x_k))$. The XOR and direct product lemmas are connected and reductions between them are known [2, 5, 6, 4, 3]. One of the main ingredients in the reduction from the direct product to the XOR lemma is the *Selective XOR Lemma*. The selective XOR lemma is a statement about the unpredictability of the function F^k defined over strings of size $(nk + k)$ as $F^k(x_1, ..., x_k, r) \equiv \oplus_i (f(x_i) \cdot r_i)$, where $x_1, ..., x_k \in \{0,1\}^n$ and $r \in \{0,1\}^k$. In this work, we give a reduction from the selective XOR lemma to the standard XOR lemma. Such reductions are already known. Goldreich [1] gave a reduction that uses pairs $(x, f(x))$ for random $x \in \{0,1\}^n$. Note that such pairs may be considered as non-uniformity in a non-uniform computational setting. Impagliazzo *et al.* [3] gave a reduction that does not require $(x, f(x))$ pairs. We give a reduction that does not require $(x, f(x))$ pairs as in the reduction of Goldreich and our reduction gives better quantitative bounds on unpredictability than that of Impagliazzo *et al.* [3] for certain choice of parameters. We discuss these in the related work subsection. We now give the formal statement of our reduction.

* A preliminary version of this work was posted on arxiv in 2014 (*arXiv:1404.5169*). This is a significantly simplified version highlighting the main result.

** Part of the work was done while the author was visiting University of California San Diego.
Email: rjaiswal@cse.iitd.ac.in

ISSN 1433-8092

Theorem 1 (Main Theorem). *If there is a (randomised) circuit C of size s such that*

$$\mathbf{Pr}\left[C(x_1,...,x_k) = F^{\oplus k}(x_1,...,x_k)\right] \geq \frac{1}{2} + \varepsilon,$$

then there is a randomised circuit C' of size $\Omega(s)$ such that

$$\mathbf{Pr}[C'(x_1,...,x_k,r) = F^k(x_1,...,x_k,r)] \geq \frac{1}{2} + \varepsilon^2.$$

The probability in the above inequalities is over the internal randomness of the circuits and uniform choice of inputs (i.e., $x_1,...,x_k \in \{0,1\}^n$ and $r \in \{0,1\}^k$).

We give the proof of this theorem in the next section. First, we look at the related work that gives reduction of the form given above.

1.1 Related work

Goldreich gave a reduction (Ex. 7.17 in [1]) that uses a circuit C for computing $F^{\oplus k}$ to construct a circuit C' for F^k in the following manner: Given an input $(x_1,...,x_k,r)$ for F^k, the circuit C' outputs $(C(x_1,...,x_k) \oplus \oplus_{i:r_i=0} f(z_i))$, where $z_1,...,z_k$ are randomly chosen strings from $\{0,1\}^n$. Note that for this reduction to work, we need $(z, f(z))$ pairs for randomly chosen $z \in \{0,1\}^n$. In the non-uniform computational setting, this can be considered the *non-uniform advice* that is *hard-wired* to the circuit. Impagliazzo *et al.* [3] gave a reduction in the uniform computational model where availability of $(z, f(z))$ pairs is not required. They get around the $(z, f(z))$ pairs requirement by using the circuit C that predicts $F^{\oplus k}$ to construct a circuit C' that predicts F^{2k} (instead of F^k as in the previous reduction). On input $(x_1,...,x_{2k},r)$ the circuit C' checks if $|\{i : r_i = 1\}|$ is equal to k. If this is not the case, then the circuit C' outputs a random bit. Otherwise, it constructs the input $(y_1,...,y_k)$ using the k strings in the set $\{x_i : r_i = 1\}$, and outputs $C(y_1,...,y_k)$. Since $\Omega(\frac{1}{\sqrt{k}})$ fraction of the $2k$ bit strings have exactly k 1's, it can be argued that if $\mathbf{Pr}[C(x_1,...,x_k) = F^{\oplus k}(x_1,...,x_k)] \geq \frac{1}{2} + \varepsilon$, then $\mathbf{Pr}[C'(x_1,...,x_{2k},r) = F^{2k}(x_1,...,x_{2k},r)] \geq \frac{1}{2} + \Omega(\frac{\varepsilon}{\sqrt{k}})$. Note that unlike the reduction of Goldreich, there is a factor of $\frac{1}{\sqrt{k}}$ in the advantage of the circuit C' over C. In our reduction, there is a factor of ε instead of $\frac{1}{\sqrt{k}}$ as in Impagliazzo *et al.*'s reduction. So, for certain parameters ($\varepsilon \geq \frac{1}{\sqrt{k}}$) our reduction gives better quantitative bounds. Moreover, we give a reduction from the selective k-XOR lemma to the standard k-XOR lemma (instead of selective $2k$-XOR to standard k-XOR lemma).

2 Proof of Theorem 1

The inputs to the functions $F^{\oplus k}$ and F^k are assumed to be sets $\{x_1,...,x_k\} \subseteq \{0,1\}^n$ and not k-tuples. This makes sense in the context of XOR lemmas since the order of inputs in a tuple is not important when taking XOR's. However, we mention this explicitly since past literature use both k-tuples and k-sets when discussing XOR lemmas [1]. In the discussion below, we will use $w(r)$ to denote the number of 1's in the string $r \in \{0,1\}^k$. Also, for a set $\{x_1,...,x_k\}$, we use $\{x_1,...,x_k\}_{|r}$ to denote the subset $\{x_i : r_i = 1\}$. Given a circuit C such that $\mathbf{Pr}[C(\{x_1,...,x_k\}) = F^{\oplus k}(\{x_1,...,x_k\})] \geq \frac{1}{2} + \varepsilon$, we construct the following circuit C' for predicting F^k:

[1] Ideas from results using one input representation can easily be extended to the other. So, it is sufficient to discuss in terms of one input representation.

$C'(\{x_1, ..., x_k\}, r)$
(1) If $(w(r)$ is odd) return a random bit
(2) Randomly partition $\{x_1, ..., x_k\}_{|r}$ into sets Y and Z each containing $\frac{w(r)}{2}$ strings
(3) Pick a random subset $S \subseteq \{0,1\}^n \setminus (Y \cup Z)$ of size $(k-i)$.
(4) Return $C(Y \cup S) \oplus C(Z \cup S)$

First, we show a lower bound on the conditional probability of the circuit C' being correct given that $w(r) = 2i$ for any i.

Lemma 1. *For any* $1 \leq i \leq \lfloor \frac{k}{2} \rfloor$, $\mathbf{Pr}[C'(\{x_1, ..., x_k\}, r) \mid w(r) = 2i] \geq \frac{1}{2} + 2\varepsilon^2$.

Proof. When given an input such that $w(r) = 2i$, the circuit partitions the set $\{x_1, ..., x_k\}_{|r}$ into Y and Z, appends both sets with a randomly chosen set S and then returns $C(Y \cup S) \oplus C(Z \cup S)$. The probability of the success of C' on such inputs can be analysed using a bipartite graph described next. Consider a bipartite graph $G = (L, R, E)$ where the vertices in L correspond to subsets of $\{0,1\}^n$ of size $(k-i)$ and vertices in R correspond to subsets of size k. There is an edge from a subset S of size $(k-i)$ in L to a subset T in R iff $S \subset T$. An edge (S, T) is coloured green if $C(T) = F^{\oplus k}(T)$ and is coloured red otherwise (i.e., green indicates that C gives correct answer on T). For any vertex S on the left, let γ_S denote the fraction of edges incident on it that are green. We know that $\mathbf{E}[\gamma_S] \geq \frac{1}{2} + \varepsilon$ since this is the probability with which circuit C succeeds. The probability of success of circuit C' is the probability that for a randomly chosen vertex on the left and two of its random edges, either both the edges are green or both are red. This probability is given by:

$$\mathbf{E}\left[\gamma_S^2 + (1-\gamma_S)^2\right] = 1 - 2\mathbf{E}[\gamma_S] + 2 \cdot \mathbf{E}[\gamma_S^2] \geq \frac{1}{2} + 2\varepsilon^2.$$

The last inequality follows from Cauchy-Schwarz. □

Analysing the success probability of C' is now simple using the above lemma. The circuit C' outputs a random bit when $w(r)$ is odd. So the conditional probability of success given that $w(r)$ is odd is $\frac{1}{2}$. The above lemma suggests that the conditional probability of success of C' given that $w(r)$ is even is $\geq \frac{1}{2} + 2 \cdot \varepsilon^2$. Combining these, we get that the probability of success of C' is at least $\frac{1}{2} + \varepsilon^2$. This concludes the proof of Theorem 1.

Acknowledgements The author thanks Oded Goldreich for useful comments on a earlier version.

References

1. Oded Goldreich. *Computational Complexity: A Conceptual Perspective*. Cambridge University Press, New York, NY, USA, 1 edition, 2008.
2. Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR lemma. Technical report, Electronic Colloquium on Computational Complexity.
3. Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: Simplified, optimized, and derandomized. *SIAM J. Comput.*, 39(4):1637–1665, January 2010.
4. Ragesh Jaiswal. *New Proofs of (New) Direct Product Theorems*. PhD thesis, University of California San Diego, 2008.
5. Falk Unger. A probabilistic inequality with applications to threshold direct-product theorems. In *Foundations of Computer Science, 2009. FOCS '09. 50th Annual IEEE Symposium on*, pages 221–229, Oct 2009.
6. Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(7):137–168, 2008.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il