Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 68 (2016)

# On Polynomial Approximations to $AC^0$

Prahladh Harsha* Srikanth Srinivasan†

April 27, 2016

**Abstract**

We make progress on some questions related to polynomial approximations of $AC^0$. It is known, by works of Tarui (*Theoret. Comput. Sci.* 1993) and Beigel, Reingold, and Spielman (*Proc. 6th CCC* 1991), that any $AC^0$ circuit of size $s$ and depth $d$ has an $\varepsilon$-error probabilistic polynomial over the reals of degree $(\log(s/\varepsilon))^{O(d)}$. We improve this upper bound to $(\log s)^{O(d)} \cdot \log(1/\varepsilon)$, which is much better for small values of $\varepsilon$.

We give an application of this result by using it to resolve a question posed by Tal (ECCC 2014): we show that $(\log s)^{O(d)} \cdot \log(1/\varepsilon)$-wise independence fools $AC^0$, improving on Tal's strengthening of Braverman's theorem (*J. ACM* 2010) that $(\log(s/\varepsilon))^{O(d)}$-wise independence fools $AC^0$. Up to the constant implicit in the $O(d)$, our result is tight. As far as we know, this is the first PRG construction for $AC^0$ that achieves optimal dependence on the error $\varepsilon$.

We also prove lower bounds on the best polynomial approximations to $AC^0$. We show that any polynomial approximating the OR function on $n$ bits to a small constant error must have degree at least $\widetilde{\Omega}(\sqrt{\log n})$. This result improves exponentially on a recent lower bound demonstrated by Meka, Nguyen, and Vu (arXiv 2015).

## 1 Motivation and Results

We use $AC^0(s,d)$ to denote $AC^0$ circuits of size $s$ and depth $d$.

**Polynomial approximations to $AC^0$.** In his breakthrough work on proving lower bounds for the class $AC^0[\oplus]$, Razborov [14] studied how well small circuits can be approximated by low-degree polynomials. We recall (an equivalent version of) his notion of polynomial approximation over the reals.

An *$\varepsilon$-error probabilistic polynomial* (over the reals) for a circuit $C(x_1,\ldots,x_n)$ is a random polynomial $\mathbf{P}(x_1,\ldots,x_n) \in \mathbb{R}[x_1,\ldots,x_n]$ such that for any $a \in \{0,1\}^n$, we have $\Pr_{\mathbf{P}}[C(a) \neq \mathbf{P}(a)] \leq \varepsilon$. Further, we say that $\mathbf{P}$ has degree $D$ and $\|\mathbf{P}\|_\infty \leq L$ if $\mathbf{P}$ is supported on polynomials $P$ of degree at most $D$ and $L_\infty$ norm at most $L$ (i.e. polynomials $P$ such that $\max_{a\in\{0,1\}^n} |P(a)| \leq L$). If there is such a $\mathbf{P}$ for $C$, we say that $C$ has $\varepsilon$-error probabilistic degree at most $D$.

It is well-known [19, 18, 2] that any circuit $C \in AC^0(s,d)$ has an $\varepsilon$-error probabilistic polynomial $\mathbf{P}$ of degree $(\log(s/\varepsilon))^{O(d)}$ and satisfying $\|\mathbf{P}\|_\infty < \exp((\log s/\varepsilon)^{O(d)})$. This can be used to

*TIFR, Mumbai, India. prahladh@tifr.res.in

†Department of Mathematics, IIT Bombay, Mumbai, India. srikanth@math.iitb.ac.in



ISSN 1433-8092

prove, for example [16], (a slightly weaker version of) Håstad's theorem [4] that says that Parity does not have subexponential-sized $AC^0$ circuits. It also plays an important role in Braverman's theorem [3] that shows that circuits from $AC^0$ can be fooled by polylog-wise independence.

**Upper bounds for probabilistic polynomials.** We show a general result regarding error reduction of probabilistic polynomials over the reals.

**Theorem 1.** *Suppose $f : \{0,1\}^n \to \{0,1\}$ has a $(\frac{1}{2} - \delta)$-error probabilistic polynomial $\mathbf{P}$ of degree $D$ and $L_\infty$ norm at most $L \geq 2$. Then, for any $\varepsilon > 0$, $f$ has an $\varepsilon$-error probabilistic polynomial of degree at most $O(\frac{D}{\delta^2}\log(1/\varepsilon))$ and $L_\infty$ norm at most $L^{O(\frac{1}{\delta^2}\log\frac{1}{\varepsilon})}$.*

Applying the above result to $(1/10)$-error probabilistic polynomials for $AC^0$ gives us small-error probabilistic polynomials for $AC^0$ with better parameters.

**Theorem 2.** *Let $C$ be any $AC^0$ circuit of size $s$ and depth $d$. Let $\varepsilon > 0$ be any parameter. The circuit $C$ has an $\varepsilon$-error probabilistic polynomial $\mathbf{P}$ of degree $(\log s)^{O(d)} \cdot \log(1/\varepsilon)$ such that $\|\mathbf{P}\|_\infty \leq \exp((\log s)^{O(d)}\log(1/\varepsilon))$.*

Similar results on probabilistic polynomials were obtained over $\mathbb{F}_2$ (for the larger class of $AC^0[\oplus]$ circuits) by Kopparty and Srinivasan [7] and extended to all fixed non-zero characteristics by Oliveira and Santhanam [13]. They have also found applications in the works of Williams [21] — for the purposes of obtaining better algorithms for satisfiability problems — and Oliveira and Santhanam [13], for proving lower bounds on compression by bounded-depth circuits. However, as far as we know, no corresponding results were observed over the reals until now.

The above theorem was motivated by an application to constructing Pseudorandom Generators (PRGs) for $AC^0$. As mentioned above, it was shown by Braverman [3] that $AC^0$ is fooled by polylog-wise independence.The proof of Braverman's theorem proceeds by constructing certain approximating polynomials from $AC^0$, which in turn depends on two previous polynomial approximation results for this circuit class. The first of these is the $L_2$-approximation result of Linial, Mansour and Nisan [8] which is based on the classical Håstad Switching Lemma [4], and the second is the above mentioned result of Tarui [18] and Beigel et al. [2]. Using these constructions, Braverman showed that $AC^0(s,d)$ is $\varepsilon$-fooled by $(\log(s/\varepsilon))^{O(d^2)}$-wise independence.

An example of Mansour appearing in the work of Luby and Veličković [9] demonstrated that $(\log s)^{d-1}\log(1/\varepsilon)$-wise independence is *required* to $\varepsilon$-fool $AC^0(s,d)$. This leads naturally to the question of showing tight bounds for the amount of independence required to fool $AC^0(s,d)$.

Using an improved switching lemma due to Håstad [5] (see also the work of Impagliazzo, Matthews, and Paturi [6]), Tal [17] gave an improved version of the $L_2$-approximation result of Linial et al. [8], and used this to improve the parameters of Braverman's theorem. Specifically, he showed that $(\log(s/\varepsilon))^{O(d)}$-wise independence fools $AC^0$.

Tal asked if the dependence on $\varepsilon$ in this result could be made to match the limit given by Mansour's example. Formally, he asked if $(\log s)^{O(d)} \cdot \log(1/\varepsilon)$-wise independence fools $AC^0(s,d)$. In this work, we are able to answer this question in the affirmative (Corollary 13 below). Up to the constant implicit in the $O(d)$, our result is optimal for all $\varepsilon > 0$.

**Comparison to other PRGs for $AC^0$.** Using standard constructions of $k$-wise independent probability distributions, the above result gives explicit PRGs with seedlength $(\log s)^{O(d)} \cdot \log(1/\varepsilon)$ for fooling circuits from $AC^0(s,d)$. It is easy to see that this seedlength cannot be improved beyond $\Omega(\log(1/\varepsilon))$ and hence that our result is optimal in terms of the error parameter $\varepsilon$.

It is also instructive to see how well this compares to general (i.e. not based on limited independence) PRG constructions for $AC^0$. Using the standard Hardness to Randomness paradigm of Nisan and Wigderson [11] and the best known average case lower bounds for $AC^0$ [6, 5], it is easy to obtain PRGs of seedlength $(\log s)^{O(d)} \cdot (\log(1/\varepsilon))^2$ for $AC^0(s,d)$. Furthermore, the Nisan-Wigderson paradigm cannot yield PRGs of seedlength less than $(\log(1/\varepsilon))^2$ given our current state of knowledge regarding circuit lower bounds (see Appendix A for details). Another recent PRG construction for $AC^0(s,d)$ due to Trevisan and Xue [20] has seedlength $(\log(s/\varepsilon))^{d+O(1)}$.

The reader will note that both constructions are suboptimal in terms of the dependence on $\varepsilon$ (though both are better than ours in terms of dependence on $s$ and $d$). Interestingly, as far as we know, our construction is the first that achieves an optimal dependence on $\varepsilon$.

**Lower bounds for probabilistic polynomials.** We can also ask if our result can be strengthened to yield a seedlength of $(\log s)^{d+O(1)} \cdot \log(1/\varepsilon)$, which would generalize both our current construction and that of Trevisan and Xue [20], and almost match Mansour's lower bound as well. Such a strengthening could conceivably be obtained by improving the polynomial approximation results for $AC^0$ [18, 2]. Razborov [14] observed that to obtain good approximations for $AC^0(s,d)$, it suffices to approximate the OR function on $s$ bits efficiently. Therefore, we study the probabilistic degree of the OR function.

Beigel, Reingold and Spielman [2] and Tarui [18] showed that the OR function on $n$ bits can be $\varepsilon$-approximated by a polynomial of degree $O((\log n) \cdot \log(1/\varepsilon))$. While it is easy to show that the dependence on $\varepsilon$ in this result is tight (in fact for *any field*), for a long time, it was not known if *any* dependence on $n$ is necessary over the reals[1]. Recently, Meka, Nguyen and Vu [10] showed that any *constant error* probabilistic polynomial for the OR function over the reals must have degree $\widetilde{\Omega}(\log\log n)$ and hence the dependence on the parameter $n$ is unavoidable. We further improve the bound of Meka et al. exponentially to $\widetilde{\Omega}(\sqrt{\log n})$, which is only a quadratic factor away from the upper bound.

## 2 Improved probabilistic polynomials and PRGs for $AC^0$

### 2.1 The construction of probabilistic polynomials

**Notation.** Let $P \in \mathbb{R}[x_1, \ldots, x_\ell]$. Given a set $S \subseteq [\ell]$ and a partial assignment $\sigma : S \to \{0,1\}$, we define $P|_\sigma$ to be the polynomial obtained by setting all the bits in $S$ according to $\sigma$. In the case that $\sigma$ sets all the variables in $S$ to a constant $b \in \{0,1\}$, we use $P|_{S \mapsto b}$ instead of $P|_\sigma$. For a function $f : \{0,1\}^\ell \to \{0,1\}$, we define $f|_\sigma$ and $f|_{S \mapsto b}$ similarly.

We define the *weight* of $P$, denoted $w(P)$, to be the sum of the absolute values of all the coefficients of $P$.

**Definition 3.** *Let $P \in \mathbb{R}[x_1, \ldots, x_\ell]$ and say $r$ is a parameter from $[\ell]$. We say that $P$ is an $\ell$-*pseudo-majority *if for $r$ being the least integer greater than $\ell/2$ and any $S \in \binom{[\ell]}{r}$ and $b \in \{0,1\}$, the polynomial $P|_{S \mapsto b}$ is the constant polynomial $b$.*

We show below that the multilinear polynomial representing the Majority function is an $\ell$-pseudo-majority of weight $2^{O(\ell)}$.

[1] In fact, for finite fields of constant size, Razborov [14] showed that the $\varepsilon$-error probabilistic degree of OR is $O(\log(1/\varepsilon))$, independent of the number of input bits.

Before we prove that this construction works, we need a few standard facts about polynomials.

**Fact 4.** *Any Boolean function $f : \{0,1\}^\ell \to \{0,1\}$ can be represented uniquely by a multilinear polynomial $P[x_1, \ldots, x_\ell]$ in the sense that for all $a \in \{0,1\}^n$, we have $P(a) = f(a)$. Furthermore, $w(P) = 2^{O(\ell)}$.*

The uniqueness in the fact above yields the following observation.

**Lemma 5.** *Let $f : \{0,1\}^\ell \to \{0,1\}$ and $P$ be the corresponding unique multilinear polynomial guaranteed by Fact 4. If $\sigma : S \to \{0,1\}$ is a partial assignment such that $f|_\sigma$ is the constant function $b \in \{0,1\}$, then $P|_\sigma$ is formally the constant polynomial $b$.*

*Proof.* Follows from the fact that $P|_\sigma$ is a multilinear polynomial representing the constant function $b$ on the variables not in $S$ and the uniqueness part of Fact 4. □

**Remark 6.** Note that the hypothesis of the lemma above is that $f|_\sigma(a) = b$ for all Boolean assignments $a$ to the remaining variables. However, the conclusion yields a stronger conclusion for the polynomial $P$: namely, we show that $P|_\sigma$ takes value $b$ on *any* assignment $a \in \mathbb{R}^{\ell - |S|}$ to the remaining variables, and not just Boolean assignments. It is this fact that we will use in applications below.

For $\ell \in \mathbb{N}$, define the Boolean function $M_\ell$ to be the Majority function: i.e., $M_\ell(x) = 1$ iff the Hamming weight of $x$ is strictly greater than $\ell/2$. Note that for any $S \subseteq \{x_1, \ldots, x_\ell\}$ of size greater than $\ell/2$ and any $b \in \{0,1\}$, $M_\ell|_{S \mapsto b}$ is the constant function $b$.

Let $P_\ell$ be the multilinear polynomial representing $M_\ell$ guaranteed by Fact 4. Applying Lemma 5 to the pair $M_\ell$ and $P_\ell$, we obtain the following corollary.

**Corollary 7.** *For any $\ell \in \mathbb{N}$, there exist $\ell$-pseudo-majorities of degree $\ell$ and weight $2^{O(\ell)}$.*

We now prove Theorem 1. We will follow the proof of [7, Lemma 10], but some additional justification will be required since we are working over the reals and not over $\mathbb{F}_2$ as in [7].

*Proof of Theorem 1.* We set $\ell = \frac{A}{\delta^2} \log(\frac{1}{\varepsilon})$ for a constant $A > 0$ to be fixed later. Let $\mathbf{P}_1, \ldots, \mathbf{P}_\ell$ be $\ell$ mutually independent copies of the probabilistic polynomial $\mathbf{P}$. Let $r = \lfloor \frac{\ell}{2} \rfloor$. Fix an $\ell$-pseudo-majority $Q$ as guaranteed by Corollary 7. The final probabilistic polynomial is $\mathbf{R} = Q(\mathbf{P}_1, \ldots, \mathbf{P}_\ell)$.

The degree of $\mathbf{R}$ is at most $\deg(Q) \cdot \deg(\mathbf{P}) \leq O(\frac{D}{\delta^2} \log(\frac{1}{\varepsilon}))$. Moreover, it can be seen that the $\|\mathbf{R}\|_\infty \leq w(Q) \cdot L^{\deg(Q)} \leq (2L)^{O(\ell)} \leq L^{O(\ell)}$ since $L \geq 2$.

Finally, we see that for any $a \in \{0,1\}^n$, $\mathbf{R}(a) = f(a)$ unless at least for $r$ many $i \in [\ell]$, we have $\mathbf{P}_i(a) \neq f(a)$. By a Chernoff bound, the probability of this is at most $\varepsilon$ as long as $A$ is chosen to be a suitably large constant. Hence, $\mathbf{R}$ is indeed an $\varepsilon$-error probabilistic polynomial for $f$. □

Theorem 2 immediately follows from the above and standard probabilistic polynomials for $\mathrm{AC}^0$ from [19, 18, 2]. However, for our applications to PRGs for $\mathrm{AC}^0$, we need a slightly stronger statement, which we prove below.

**Definition 8** (Probabilistic polynomial with witness)**.** *An $\varepsilon$-error probabilistic polynomial for circuit $C(x_1, \ldots, x_n)$ with witness ($\varepsilon$-error PPW for short) is a pair $(\mathbf{P}, \boldsymbol{\mathcal{E}})$ of random variables such that $\mathbf{P}$ is a randomized polynomial and $\boldsymbol{\mathcal{E}}$ is a randomized circuit (both on $n$ Boolean variables) such that for any input $a \in \{0,1\}^n$, we have*

- $\Pr_{\boldsymbol{\mathcal{E}}}[\boldsymbol{\mathcal{E}}(a) = 1] \leq \varepsilon$,

- *For any fixing $(P, \mathcal{E})$ of $(\mathbf{P}, \boldsymbol{\mathcal{E}})$, we have $\mathcal{E}(a) = 0 \Rightarrow P(a) = C(a)$.*

*In particular, this implies that $\mathbf{P}$ is an $\varepsilon$-error probabilistic polynomial for $C$.*

*We say that $\boldsymbol{\mathcal{E}}$ belongs to a circuit class $\mathcal{C}$ if it is supported on circuits from class $\mathcal{C}$.*

The above notion was introduced in Braverman [3] who proved the following lemma, building on earlier works of [19, 18, 2].

**Lemma 9** ([3, Lemma 8, Proposition 9])**.** *Fix parameters $s, d \in \mathbb{N}$ and $\varepsilon > 0$. Any $\mathrm{AC}^0$ circuit $C$ of size $s$ and depth $d$ has an $\varepsilon$-error PPW $(\mathbf{P}, \boldsymbol{\mathcal{E}})$ where*

- $\deg(\mathbf{P}) \leq (\log(s/\varepsilon))^{O(d)}$ *and* $\|\mathbf{P}\|_\infty \leq \exp((\log(s/\varepsilon))^{O(d)})$,
- $\boldsymbol{\mathcal{E}} \in \mathrm{AC}^0(\mathrm{poly}(s \log(1/\varepsilon)), d + 3)$.

We show the following variant of the above lemma, which is an improvement in terms of degree and the $L_\infty$ norm of the probabilistic polynomial for small $\varepsilon$.

**Lemma 10.** *Fix parameters $s, d \in \mathbb{N}$ and $\varepsilon > 0$. Any $\mathrm{AC}^0$ circuit $C$ of size $s$ and depth $d$ has an $\varepsilon$-error PPW $(\mathbf{P}, \boldsymbol{\mathcal{E}})$ where*

- $\deg(\mathbf{P}) \leq (\log s)^{O(d)} \cdot \log(1/\varepsilon)$ *and* $\|\mathbf{P}\|_\infty \leq \exp((\log s)^{O(d)} \log(1/\varepsilon))$,
- $\boldsymbol{\mathcal{E}} \in \mathrm{AC}^0(\mathrm{poly}(s \log(1/\varepsilon)), d + O(1))$.

Before we begin the proof, we state one more lemma from the literature. Given an integer parameter $\ell$ and real parameters $\alpha, \beta \in [0, 1]$ with $\alpha < \beta$, we will call a function $f : \{0,1\}^\ell \to \{0,1\}$ an $(\ell, \alpha, \beta)$*-approximate majority* if $f(x) = 0$ for any input of Hamming weight at most $\alpha\ell$ and $f(x) = 1$ for any input of Hamming weight at least $\beta\ell$. The following is a result of Ajtai and Ben-Or [1].

**Lemma 11** (Ajtai and Ben-Or [1])**.** *Fix any constants $\alpha < \beta$. Then, for all $\ell \in \mathbb{N}$, there is an $(\ell, \alpha, \beta)$-approximate majority which has an $\mathrm{AC}^0$ circuit of size $\mathrm{poly}(\ell)$ and depth 3.*

We now prove Lemma 10. The proof is similar to that of Theorem 1 above, but we also need to obtain a witness circuit for our probabilistic polynomial.

*Proof of Lemma 10.* Let $\ell = A \log(1/\varepsilon)$ for a large constant $A$ to be chosen later. W.l.o.g. assume that $\ell$ is even. Let $r = \lceil \ell/2 \rceil + 1$ and let $Q(x_1, \ldots, x_\ell)$ be the $\ell$-pseudo-majority guaranteed by Corollary 7. Let $k = \ell/4$. By Lemma 11, there is an $\mathrm{AC}^0$ circuit $C_1$ of size $\mathrm{poly}(\ell)$ and depth 3 that computes an $(\ell, 1/4, 2/5)$-approximate majority.

Let $(\mathbf{P}_1, \boldsymbol{\mathcal{E}}_1), \ldots, (\mathbf{P}_\ell, \boldsymbol{\mathcal{E}}_\ell)$ be *independent* copies of the $(1/8)$-error PPW guaranteed by Lemma 9. The final PPW is $(\mathbf{P}, \boldsymbol{\mathcal{E}})$ where $\mathbf{P} = Q(\mathbf{P}_1, \ldots, \mathbf{P}_\ell)$ and $\boldsymbol{\mathcal{E}} = C_1(\boldsymbol{\mathcal{E}}_1, \ldots, \boldsymbol{\mathcal{E}}_\ell)$. We show that this PPW has the required properties.

First of all, we know that on any input $a$ to the circuit $C$ and for any $i \in [\ell]$, the probability that $\boldsymbol{\mathcal{E}}_i(a) = 1$ is at most $1/8$. Thus, the expected number of $\boldsymbol{\mathcal{E}}_i$ that output 1 is at most $\ell/8$. However, for $\boldsymbol{\mathcal{E}}(a)$ to be 1, at least $\ell/4$ many $\boldsymbol{\mathcal{E}}_i(a)$ should be 1. By a Chernoff bound, the probability of this event is at most $\exp(-\Omega(\ell)) < \varepsilon$ for a large enough constant $A$.

Now, we need to argue that if $\boldsymbol{\mathcal{E}}(a) = 0$, then $\mathbf{P}(a) = Q(\mathbf{P}_1(a), \ldots, \mathbf{P}_\ell(a)) = C(a)$. Say $C(a) = b \in \{0, 1\}$. If $\boldsymbol{\mathcal{E}}(a) = 0$, then we know that the number of $\boldsymbol{\mathcal{E}}_i(a)$ that are 0 is at least $3\ell/5$; let $I$

denote the set of these $i$. By the definition of PPWs, we know that for each $i \in I$, we have $\mathbf{P}_i(a) = b$ and hence at least $3\ell/5 > r$ many inputs of $Q$ are set to $b$. Since $Q$ is an $\ell$-pseudo-majority, we must have $Q(\mathbf{P}_1(a), \ldots, \mathbf{P}_\ell(a)) = b$. This concludes the proof that $(\mathbf{P}, \boldsymbol{\mathcal{E}})$ is indeed an $\varepsilon$-error PPW for $C$.

Note that $\deg(\mathbf{P}) \leq \deg(Q) \cdot \max_i \deg(\mathbf{P}_i) \leq (\log s)^{O(d)} \log(1/\varepsilon)$. Also, it can be seen that

$$\|\mathbf{P}\|_\infty \leq w(Q) \cdot (\max_{i \in [\ell]} \|\mathbf{P}_i\|_\infty)^{\deg(Q)} \leq \exp((\log s)^{O(d)} \log(1/\varepsilon)).$$

Thus, $\mathbf{P}$ has the required properties. The size and depth properties of $\boldsymbol{\mathcal{E}}$ follow trivially from its definition. This concludes the proof of the lemma. □

### 2.2 Application to PRGs for $\mathrm{AC}^0$

The connection between probabilistic polynomials and PRGs for $\mathrm{AC}^0$ is encapsulated in the following theorem (which is an easy observation from the works of Braverman and Tal):

**Theorem 12** (Braverman [3],Tal [17])**.** *Let $s, d \in \mathbb{N}$ and $\varepsilon > 0$. Suppose that any $\mathrm{AC}^0$ circuit of size $s$ and depth $d$ has an $(\varepsilon/2)$-error PPW $(\mathbf{P}, \boldsymbol{\mathcal{E}})$ such that*

- $\deg(\mathbf{P}) = D$, $\|\mathbf{P}\|_\infty \leq L$,
- $\boldsymbol{\mathcal{E}} \in \mathrm{AC}^0(s_1, d_1)$,

*Then, $\mathrm{AC}^0$ circuits of size $s$ and depth $d$ can be $\varepsilon$-fooled by $k(s, d, \varepsilon)$-wise independence, where*

$$k(s, d, \varepsilon) = O(D) + (\log s_1)^{O(d_1)} \cdot (\log(1/\varepsilon) + \log L)$$

Note that the theorem above is trivial when $\log(1/\varepsilon) > s$ since any $\mathrm{AC}^0$ circuit of size $s$ is trivially fooled by an $s$-wise independent distribution. Hence, the theorem is non-trivial only when $\log(1/\varepsilon) \leq s$. In this case, using Lemma 10 and the theorem above, we immediately get

**Corollary 13.** *Fix parameters $s, d \in \mathbb{N}$ and $\varepsilon > 0$. Any circuit $C \in \mathrm{AC}^0(s, d)$ can be $\varepsilon$-fooled by any distribution that is $(\log s)^{O(d)} \log(1/\varepsilon)$-wise independent.*

**Remark 14.** A close look at the above proof (including the details of Lemma 9 and Theorem 12) shows that the amount of independence required to $\varepsilon$-fool $\mathrm{AC}^0(s, d)$ is $(\log s)^{3d+O(1)} \cdot \log(1/\varepsilon)$. Avishay Tal (personal communication) showed that the above can be further improved to $(\log s)^{2.5d+O(1)} \cdot \log(1/\varepsilon)$-wise independence. It is open if this can be further strengthened to, say, $(\log s)^{d+O(1)} \cdot \log(1/\varepsilon)$ or even $(\log s)^{d-1} \cdot \log(1/\varepsilon)$, matching the lower bound due to Mansour [9].

## 3 The probabilistic degree of OR

**Notation.** For $i \geq 1$ and a set of Boolean variables $X$, let $\mu_i^X$ be the product distribution on $\{0,1\}^X$ defined so that for each $x \in X$, the probability that $x = 1$ is $2^{-i}$. We also use $\mathcal{U}_X$ to denote $\mu_1^X$, the uniform distribution over $\{0,1\}^X$. The OR function on the variables in $X$ is denoted $\mathrm{OR}_X$.

We want to show:

**Theorem 15.** *Assume $|X_0| = n$. The $1/8$-error probabilistic degree of $\mathrm{OR}_{X_0}$ is at least $\frac{\sqrt{\log n}}{(\log \log n)^2}$.*

**Remark 16.** Though the theorem is stated for error $1/8$, it is not hard to see that it holds (with constant factor losses) as long as the error is bounded by $1/2 - \Omega(1)$. One way to see this is to appeal to Theorem 2. Another way is to do a simple error reduction specific to the OR function which we do in the proof of Theorem 15.

In order to prove Theorem 15, we use an anti-concentration lemma due to Meka, Nguyen and Vu [10][2] coupled with a random restriction argument inspired by the work of Razborov and Viola [15].

**Lemma 17** (Meka, Nguyen, and Vu [10])**.** *There exists an absolute constant $B > 0$ so that the following holds. Let $p(x) \in \mathbb{R}[X]$ be a degree $d$ multilinear polynomial with at least $r$ disjoint degree $d$ terms. Then $\Pr_{x \sim \mathcal{U}_X}[p(x) = 0] \leq Bd^{4/3} r^{-\frac{1}{4d+1}} \sqrt{\log r}$.*

Given a polynomial $q \in \mathbb{R}[X]$, we denote by $\mathrm{Err}_i^X(q)$ the error of polynomial $q$ w.r.t. distribution $\mu_i^X$. Formally,

$$\mathrm{Err}_i^X(q) = \Pr_{x \sim \mu_i^X}[q(x) \neq \mathrm{OR}_X(x)]$$

For a set of variables $X$, $\ell \in \mathbb{N}$ and $\delta \in \mathbb{R}^{\geq 0}$, call a polynomial $q \in \mathbb{R}[X]$ $(X, \ell, \delta)$-good if

$$\mathop{\mathbf{E}}_{i \in [\ell]}[\mathrm{Err}_i^X(q)] \leq \delta.$$

A random restriction on the variable set $X$ with $*$-probability $p \in [0,1]$ will be a function $\rho : X \to \{*, 0\}$ with each variable set independently to $*$ with probability $p$ and to 0 otherwise. We use $X_\rho$ to denote $\rho^{-1}(*)$. The restriction of a polynomial $q$ under $\rho$ is denoted $q|_\rho$.

**Observation 18.** Let $q \in \mathbb{R}[X]$ and $\rho$ be a random restriction on the variable set $X$ with $*$-probability $p = \frac{1}{2^b}$ where $b \in \mathbb{N}$. For any $i \geq 1$,

$$\mathop{\mathbf{E}}_{\rho}[\mathrm{Err}_i^{X_\rho}(q|_\rho)] = \mathrm{Err}_{i+b}^X(q)$$

(I.e., setting bits independently to 1 with probability $\frac{1}{2^{i+b}}$ is the same as first applying a random restriction with $*$-probability $\frac{1}{2^b}$ and then setting each surviving variable to 1 with probability $\frac{1}{2^i}$.)

### 3.1 Proof of Theorem 15

We argue by contradiction. Let $\mathbf{P}$ be a $1/8$-error probabilistic polynomial for $\mathrm{OR}_{X_0}$ of degree $D < \sqrt{\log n}/(\log\log n)^2$. In particular, we have

$$\Pr_{\mathbf{P}}[\mathbf{P}(0, 0, \ldots, 0) \neq 0] \leq \frac{1}{8}$$

We discard all such polynomials from the distribution underlying $\mathbf{P}$ (e.g. if such a bad polynomial is sampled, then we could just output 0). The resulting probabilistic polynomial $\mathbf{P}'$ is supported only on polynomials $p \in \mathbb{R}[X_0]$ such that $p(0, 0, \ldots, 0) = 0$ and further, $\mathbf{P}'$ is a $(1/4)$-error probabilistic polynomial for $\mathrm{OR}_{X_0}$ of degree $D$.

[2]The result of Meka et al. is actually stated for polynomials over the *Fourier basis* of Parity functions (see, e.g., the book of O'Donnell [12]). However, it is an easy observation that a polynomial of degree $d$ has $r$ disjoint terms of degree $d$ in the standard monomial basis if and only if it has $r$ disjoint terms of degree $d$ in the Fourier basis. Hence, the result holds in the standard basis as well.

Let $\mathbf{P}'_1, \ldots, \mathbf{P}'_s$ be $s = \log\log n$ independent instances of $\mathbf{P}'$ and let $\mathbf{Q} = 1 - \prod_{i\in[s]}(1 - \mathbf{P}'_i)$. Then, $\mathbf{Q}$ is an error $\frac{1}{4^s} = \frac{1}{\log^2 n}$ probabilistic polynomial for $\mathrm{OR}_n$ of degree at most $sD < \sqrt{\log n}/\log\log n$. In particular, there is a polynomial $q_0 \in \mathbb{R}[x_1, \ldots, x_n]$ of degree $d_0 < \sqrt{\log n}/\log\log n$ such that $q_0(0, 0, \ldots, 0) = 0$ and for $\varepsilon_0 = \frac{1}{\log^2 n}$ we have

$$\mathop{\mathbf{E}}_{i\in[(\log n)/2]}[\mathrm{Err}_i^{X_0}(q_0)] \leq \varepsilon_0$$

Define $n_0 = |X_0| = n$ and $\ell_0 = (\log n)/2$. By the above inequality, the polynomial $q_0$ is $(X_0, \ell_0, \varepsilon_0)$-good. Also define parameters $r = (d_0 \cdot \log^2 n)^{10d_0}$ and $p = \frac{1}{2^b}$ where $b \in \mathbb{N}$ is chosen so that $p \in [\frac{1}{2r^2}, \frac{1}{r^2}]$. Note that $r = n^{o(1)}$ and hence $p = \frac{1}{n^{o(1)}}$.

We now define a sequence of polynomials $q_1, q_2, \ldots, q_t$ such that:

- Each $q_i \in \mathbb{R}[X_i]$ where $X_i \subseteq X_0$ and has degree $d_i \geq 0$. Also, $|X_i| = n_i$ where $n_i \in [pn_{i-1}/2, 3pn_{i-1}/2]$. Further $\deg(q_i) = d_i < d_{i-1}$. The polynomial $q_i = q_{i-1}|_{\rho_i}$ for some restriction $\rho_i : X_{i-1} \to \{*, 0\}$.
- Each polynomial $q_i$ is $(X_i, \ell_i, \varepsilon_i)$-good where $\ell_i = \ell_{i-1} - b$ and $\varepsilon_i = \varepsilon_{i-1} \cdot \exp(\frac{16b}{\log n})$.
- $d_t = \deg(q_t) = 0$. That is, $q_t$ is a constant polynomial.

Before we describe how to construct this sequence, let us see how it implies the desired contradiction. Note that since $d_i < d_{i-1}$ for each $i \geq 1$, the length $t$ of the sequence is bounded by $d_0 < \sqrt{\log n}/\log\log n$.

We first make the following simple claim.

**Claim 19.** *For each $i \in [t]$, $n_i \geq \sqrt{n}$, $\ell_i \geq \frac{\log n}{4}$, and $\varepsilon_i < \frac{1}{\log n}$.*

*Proof.*

$$n_i \geq n_t \geq n_0 \cdot (p/2)^t = n \cdot (d_0 \log n)^{-O(d_0^2)} \geq \sqrt{n}.$$

Also, note that $\ell_i = \ell_0 - bi \geq \ell_0 - bt = (\log n)/2 - O(d_0^2 \log\log n) \geq \frac{\log n}{4}$ and

$$\varepsilon_i = \varepsilon_0 \cdot \exp(\frac{16bi}{\log n}) \leq \varepsilon_0 \cdot \exp(\frac{16bt}{\log n}) = \frac{1}{\log^2 n} \cdot \exp(\frac{O(d_0^2 \log\log n)}{\log n}) < \frac{1}{\log n}.$$

□

In particular, since $q_t$ is $(X_t, \ell_t, \varepsilon_t)$-good, we must have

$$\mathrm{Err}_1^{X_t}(q_t) \leq \ell_t \mathop{\mathbf{E}}_{i\in[\ell_t]}[\mathrm{Err}_i^{X_t}(q_t)] < \varepsilon_t \ell_t < \frac{1}{2} \tag{1}$$

using the fact that $\ell_t \leq \ell_0 = (\log n)/2$ and $\varepsilon_t < \frac{1}{\log n}$.

Since $n_t \geq \sqrt{n}$, the function $\mathrm{OR}_{X_t}(x)$ evaluates to 1 under the distribution $\mu_1^{X_t} = \mathcal{U}_{X_t}$ with probability $1 - o(1)$. Thus, $q_t$ must also evaluate to 1 on some input. However, since $q_t$ is a constant polynomial, this implies that $q_t = 1$. But this implies that $q_t(0, 0, \ldots, 0) = 1$ as well, which leads to a contradiction, since $q_t$ is obtained by setting some input bits of $q_0$ to 0 and $q_0(0, 0, \ldots, 0) = 0$ by our choice of $q_0$. This completes the proof of the theorem.

Now we describe how to obtain the sequence $q_1, \ldots, q_t$. More precisely, we describe how to obtain $q_i$ from $q_{i-1}$ assuming $d_{i-1} \geq 1$. Fix any $i \geq 1$ such that $d_{i-1} \geq 1$. We assume that the sequence $q_1, \ldots, q_{i-1}$ of polynomials constructed so far satisfy the above properties.

For brevity, let $q, X, m, d, \ell, \varepsilon$ denote $q_{i-1}, X_{i-1}, n_{i-1}, d_{i-1}, \ell_{i-1}, \varepsilon_{i-1}$ respectively.

We know that $q$ is $(X, \ell, \varepsilon)$-good. As we did in (1) for $q_t$, we can use this to show that $\mathrm{Err}_1^X(q) < \frac{1}{2}$ and since $\mathrm{OR}_X(x)$ takes the value 1 on an input $x \sim \mathcal{U}_X$ with probability $1 - o(1)$, we see that

$$\Pr_{x \sim \mathcal{U}_X}[q(x) = 1] \geq \frac{1}{2} - o(1) \geq \frac{1}{3}. \tag{2}$$

Lemma 17 then implies that there cannot $r$ disjoint monomials of degree $d$ in $q$. To see this, assume that there are indeed $r$ many disjoint monomials of degree $d$ in $q$. Then by Lemma 17, the probability that $q(x) - 1 = 0$ for a random $x \sim \mathcal{U}_X$ is at most

$$\begin{aligned} Bd^{4/3} r^{-\frac{1}{4d+1}} \sqrt{\log r} &\leq B d_0^{4/3} r^{-\frac{1}{5d_0}} \sqrt{\log r} \\ &\leq B d_0^{4/3} \cdot \frac{\sqrt{10 d_0 \log(d_0 \log^2 n)}}{d_0^2 \log^4 n} = o(1). \end{aligned}$$

This contradicts (2).

Hence, we know that $q$ cannot be contain more than $r$ many disjoint monomials of degree $d$. Let $S$ be any maximal set of disjoint monomials appearing $q$. Note that by definition, every monomial of degree $d$ contains at least one variable from $S$ and hence setting all the variables in $S$ reduces the degree of the polynomial. The number of variables appearing in $S$ is at most $d|S| \leq dr$.

We now choose a random restriction $\rho$ with $*$-probability $p$ as defined above and consider the polynomial $q|_\rho$. Define the following "bad" events:

- $\mathcal{E}_1(\rho)$ is the event that $|X_\rho| \notin [pm/2, 3pm/2]$.
- $\mathcal{E}_2(\rho)$ is the event that some variable in $S$ is not set to 0.
- $\mathcal{E}_3(\rho)$ is the event that $q|_\rho$ is not $(X_\rho, \ell', \varepsilon')$-good where $\ell' = \ell - b$ and $\varepsilon' = \varepsilon \cdot \exp(\frac{Cb}{\log n})$.

We claim that there is a $\rho$ so that none of the bad events $\mathcal{E}_1(\rho), \mathcal{E}_2(\rho)$ or $\mathcal{E}_3(\rho)$ occur. This will imply that we can take $q_i = q|_\rho, X_i = X_\rho, \ell_i = \ell', \varepsilon_i = \varepsilon'$ and we will be done. So we only need to show that $\Pr_\rho[\mathcal{E}_1(\rho) \vee \mathcal{E}_2(\rho) \vee \mathcal{E}_3(\rho)] < 1$. This is done as follows.

- $\Pr_\rho[\mathcal{E}_1(\rho)]$: By Claim 19, we know that $m \geq \sqrt{n}$ and hence $\mathbf{E}_\rho[|X_\rho|] = pm = m \cdot \frac{1}{n^{o(1)}} \geq n^{1/4}$. Hence, by a Chernoff bound, the probability that $|X_\rho| \notin [pm/2, 3pm/2]$ is bounded by $\exp(-\Omega(n^{1/4}))$.
- $\Pr_\rho[\mathcal{E}_2(\rho)]$: By a union bound over $S$, this probability is bounded by $p|S| \leq rd_0/r^2 < \frac{1}{\log n}$.
- $\Pr_\rho[\mathcal{E}_3(\rho)]$: By Observation 18, we know that for any $i$,

$$\mathop{\mathbf{E}}_\rho[\mathrm{Err}_i^{X_\rho}(q|_\rho)] = \mathrm{Err}_{i+b}^X(q).$$

Hence,

$$\mathop{\mathbf{E}}_{\rho}[\mathop{\mathbf{E}}_{i\in[\ell']}[\mathrm{Err}_i^{X_\rho}(q|_\rho)]] = \mathop{\mathbf{E}}_{i\in[\ell']}[\mathrm{Err}_{i+b}^{X}(q)] = \mathop{\mathbf{E}}_{i\in\{b+1,\ldots,b+\ell'\}}[\mathrm{Err}_i^{X}(q)] = \mathop{\mathbf{E}}_{i\in\{b+1,\ldots,\ell\}}[\mathrm{Err}_i^{X}(q)]. \quad (3)$$

We can bound the right hand side of the above equation by

$$\mathop{\mathbf{E}}_{i\in\{b+1,\ldots,\ell\}}[\mathrm{Err}_i^{X}(q)] \leq \frac{1}{(1-\frac{b}{\ell})}\mathop{\mathbf{E}}_{i\in[\ell]}[\mathrm{Err}_i^{X}(q)] \leq \frac{\varepsilon}{(1-\frac{b}{\ell})}$$

where the final inequality follows from the fact that $q$ is $(X,\ell,\varepsilon)$-good. Further, by Claim 19, we know that $\ell \geq \frac{\log n}{4} \gg b$, and hence we can bound the above as follows.

$$\mathop{\mathbf{E}}_{i\in\{b+1,\ldots,\ell\}}[\mathrm{Err}_i^{X}(q)] \leq \frac{\varepsilon}{(1-\frac{b}{\ell})} \leq \varepsilon\cdot(1+\frac{2b}{\ell}) \leq \varepsilon\cdot(1+\frac{8b}{\log n}).$$

Plugging the above bound into (3), we obtain

$$\mathop{\mathbf{E}}_{\rho}[\mathop{\mathbf{E}}_{i\in[\ell']}[\mathrm{Err}_i^{X_\rho}(q|_\rho)]] \leq \varepsilon\cdot(1+\frac{8b}{\log n}) \leq \varepsilon\cdot\exp(\frac{8b}{\log n}).$$

By Markov's inequality,

$$\Pr_{\rho}[\mathop{\mathbf{E}}_{i\in[\ell']}[\mathrm{Err}_i^{X_\rho}(q|_\rho)] > \varepsilon\cdot\exp(\frac{16b}{\log n})] \leq \exp(-\frac{8b}{\log n}) = 1-\Omega(\frac{b}{\log n}) \leq 1-\frac{2}{\log n}.$$

Thus, $\Pr_\rho[\mathcal{E}_3(\rho)] \leq 1-\frac{2}{\log n}$.

By a union bound, we have

$$\Pr_{\rho}[\mathcal{E}_1(\rho)\vee\mathcal{E}_2(\rho)\vee\mathcal{E}_3(\rho)] \leq \exp(-\Omega(n^{1/4})) + \frac{1}{\log n} + 1 - \frac{2}{\log n} < 1.$$

**Acknowledgements.** We thank Swagato Sanyal and Madhu Sudan for encouragement and useful discussions which greatly simplified our proofs. We also thank Avishay Tal for his generous feedback and comments and also for showing us the improvement in seedlength mentioned in Remark 14.

## References

[1] Miklós Ajtai and Michael Ben-Or. A theorem on probabilistic constant depth computations. In *Proc. 16th ACM Symp. on Theory of Computing (STOC)*, pages 471–474, 1984. doi:10.1145/800057.808715. 5

[2] Richard Beigel, Nick Reingold, and Daniel A. Spielman. The perceptron strikes back. In *Proc. 6th IEEE Conf. on Structure in Complexity Theory*, pages 286–291, 1991. doi:10.1109/SCT.1991.160270. 1, 2, 3, 4, 5

[3] Mark Braverman. Polylogarithmic independence fools $AC^0$ circuits. *J. ACM*, 57(5), 2010. (Preliminary version in *24th IEEE Conference on Computational Complexity*, 2009). eccc:TR09-011, doi:10.1145/1754399.1754401. 2, 5, 6

[4] Johan Håstad. Almost optimal lower bounds for small depth circuits. In Silvio Micali, editor, *Randomness and Computation*, volume 5 of *Advances in Computing Research*, pages 143–170. JAI Press, Greenwich, Connecticut, 1989. (Preliminary version in *18th STOC* 1986). URL: http://www.csc.kth.se/~johanh/largesmalldepth.pdf. 2

[5] Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. Comput.*, 43(5):1699–1708, 2014. eccc:TR12-137, doi:10.1137/120897432. 2, 3, 12

[6] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for $AC^0$. In *Proc. 23rd Annual ACM-SIAM Symp. on Discrete Algorithms (SODA)*, pages 961–972, 2012. arXiv:1107.3127. 2, 3, 12

[7] Swastik Kopparty and Srikanth Srinivasan. Certifying polynomials for $AC^0(\text{parity})$ circuits, with applications. In Deepak D'Souza, Telikepalli Kavitha, and Jaikumar Radhakrishnan, editors, *Proc. 32nd IARCS Annual Conf. on Foundations of Software Tech. and Theoretical Comp. Science (FSTTCS)*, volume 18 of *LIPIcs*, pages 36–47. Schloss Dagstuhl, 2012. eccc:TR12-102, doi:10.4230/LIPIcs.FSTTCS.2012.36. 2, 4

[8] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993. (Preliminary version in *30th FOCS*, 1989). doi:10.1145/174130.174138. 2

[9] Michael Luby and Boban Velickovic. On deterministic approximation of DNF. *Algorithmica*, 16(4/5):415–433, 1996. (Preliminary version in *23rd STOC*, 1991). doi:10.1007/BF01940873. 2, 6

[10] Raghu Meka, Oanh Nguyen, and Van Vu. Anti-concentration for polynomials of independent random variables. 2015. arXiv:1507.00829. 3, 7

[11] Noam Nisan and Avi Wigderson. Hardness vs. randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, October 1994. (Preliminary version in *29th FOCS*, 1988). doi:10.1016/S0022-0000(05)80043-1. 3, 12

[12] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. URL: http://analysisofbooleanfunctions.org/, doi:10.1017/CBO9781139814782. 7

[13] Igor Carboni Oliveira and Rahul Santhanam. Majority is incompressible by $AC^0[p]$ circuits. In *Proc. 30th Computational Complexity Conf.*, pages 124–157, 2015. eccc:TR14-173, doi:10.4230/LIPIcs.CCC.2015.124. 2

[14] Alexander A. Razborov. Нжние оценки размера схем ограниченной глубины в полном базисе, содержащем функцию логического сложения (Russian) [Lower bounds on the size of bounded depth circuits over a complete basis with logical addition]. *Mathematicheskie Zametki*, 41(4):598–607, 1987. (English translation in *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987). URL: http://mi.mathnet.ru/eng/mz4883, doi:10.1007/BF01137685. 1, 3

[15] Alexander A. Razborov and Emanuele Viola. Real advantage. *ACM T. Comput. Theory*, 5(4):17, 2013. eccc:TR12-134, doi:10.1145/2540089. 7

[16] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proc. 19th ACM Symp. on Theory of Computing (STOC)*, pages 77–82, 1987. doi:10.1145/28395.28404. 2

[17] Avishay Tal. Tight bounds on the fourier spectrum of $AC^0$. Technical Report TR14-174, Elect. Colloq. on Comput. Complexity (ECCC), 2014. eccc:TR14-174. 2, 6

[18] Jun Tarui. Probablistic polynomials, $AC^0$ functions, and the polynomial-time hierarchy. *Theoret. Comput. Sci.*, 113(1):167–183, 1993. (Preliminary Version in *8th STACS*, 1991). doi:10.1016/0304-3975(93)90214-E. 1, 2, 3, 4, 5

[19] Seinosuke Toda and Mitsunori Ogiwara. Counting classes are at least as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 21(2):316–328, 1992. (Preliminary version in *6th Structure in Complexity Theory Conference*, 1991). doi:10.1137/0221023. 1, 4, 5

[20] Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of $AC^0$. In *Proc. 28th IEEE Conf. on Computational Complexity*, pages 242–247, 2013. doi:10.1109/CCC.2013.32. 3

[21] Ryan Williams. New algorithms and lower bounds for circuits with linear threshold gates. In *Proc. 46th ACM Symp. on Theory of Computing (STOC)*, pages 194–202, 2014. arXiv:1401.2444, doi:10.1145/2591796.2591858. 2

# A The limitations of the Nisan Wigderson paradigm

In this section, we show that the general hardness-to-randomness tradeoff of Nisan and Wigderson [11] does not yield a PRG with optimal seedlength as a function of $\varepsilon$ *given our current knowledge* of circuit lower bounds.

We start by describing the meta-result of Nisan and Wigderson [11] that allows us to convert any sufficiently hard-to-compute function for a class of circuits to a PRG for a slightly weaker class of circuits. The result is true in greater generality than we describe here but to keep things concrete, we stick to the setting of $\mathrm{AC}^0(s,d)$.

We say that a function $f : \{0,1\}^r \to \{0,1\}$ is $(s,d,\varepsilon)$-hard if given any circuit $C$ from $\mathrm{AC}^0(s,d)$ of size $s$, we have

$$\Pr_{x\in\{0,1\}^r}[C(x) = f(x)] \leq \frac{1}{2} + \varepsilon.$$

For non-negative integers $m, r, \ell, s$, we say that a family $\mathcal{F} \subseteq \binom{[m]}{r}$, we say that $\mathcal{F}$ is an $(m,r,\ell,s)$ design if $|\mathcal{F}| = s$ and for any distinct $S, T \in \mathcal{F}$, we have $|S \cap T| \leq \ell$.

Nisan and Wigderson [11] show the following.

**Theorem 20** ([11])**.** *Let $m, r, \ell, s \in \mathbb{N}$ be positive parameters such that $m \geq r \geq \ell$. Given an explicit $f : \{0,1\}^r \to \{0,1\}$ that is $(s \cdot 2^\ell, d+1, \varepsilon/s)$-hard and an explicit $(m,r,\ell,s)$-design, we can construct an explicit PRG $G : \{0,1\}^m \to \{0,1\}^s$ that fools circuits from $\mathrm{AC}^0(s,d)$ with error at most $\varepsilon$.*

To use this theorem, we need a hard function for circuits in $\mathrm{AC}^0$. The best such result known currently is the following due to Impagliazzo, Matthews, and Paturi [6] (see also Håstad [5]).

**Theorem 21.** *Let $d \geq 1$ be a constant. The Parity function on $r$ is bits is $(s_1, d_1, \delta)$-hard if $r \geq A(\log s_1)^{d_1-1} \cdot \log(1/\delta)$ for some constant $A > 0$ depending on $d$.*

Thus, if we want to apply Theorem 20 alongside the lower bound given by Theorem 21 to construct PRGs that $\varepsilon$-fool $\mathrm{AC}^0(s,d)$, then we need

$$r \geq A(\log s + \ell)^d \cdot \log(s/\varepsilon) \geq A(\log s + \ell)^d \cdot \log(1/\varepsilon) \tag{4}$$

for some constant $A > 0$ depending on $d$.

Further, to construct an $(m,r,\ell,s)$-design, we claim that we further need

$$m \geq \min\{r^2/2\ell, s\}. \tag{5}$$

We justify (5) below, but first we use it to prove that the Nisan-Wigderson paradigm cannot be used to obtain seedlength optimal in terms of $\varepsilon$ for a large range of $\varepsilon$.

We assume that $\varepsilon \geq \exp(-s^{1/4})$ (the same proof works as long as $\varepsilon \geq \exp(-s^{\frac{1}{2}-\Omega(1)})$). In this setting, we show that $m \geq B(\log s)^{2d-1} \cdot (\log(1/\varepsilon))^2$ for some constant $B$ depending on $d$.

To see this, note that if $m \geq s$, then trivially we have $(\log s)^{2d-1} \cdot (\log 1/\varepsilon)^2 \leq s^{\frac{1}{2}+o(1)} < s \leq m$. So we assume that $m < s$.

In this case, (5) tells us that $m \geq r^2/2\ell$, which yields

$$\begin{aligned} m &\geq \frac{r^2}{2\ell} \geq \frac{A^2(\log s + \ell)^{2d} \cdot (\log 1/\varepsilon)^2}{2\ell} \\ &\geq \frac{A^2(\log s)^{2d-1}\ell(\log 1/\varepsilon)^2}{2\ell} = \Omega(A^2(\log s)^{2d-1} \cdot (\log(1/\varepsilon))^2) \end{aligned}$$

as required.

The inequality (5) is a standard combinatorial fact and can be found in many standard textbooks. For completeness, here is a simple proof using inclusion-exclusion.

Note that if $s \leq r$, then we immediately have $m \geq r \geq s$ and (5) is proved. So assume that $s > r$ and in particular given any $(m, r, \ell, s)$-design $\mathcal{F}$, we can choose $t = r/\ell$ sets $T_1, \ldots, T_t$ from $\mathcal{F}$. By inclusion-exclusion, we have

$$\begin{aligned} m &\geq |\bigcup_{i\in[t]} T_i| \geq \sum_i |T_i| - \sum_{i<j} |T_i \cap T_j| \\ &\geq rt - \frac{t^2}{2} \cdot \ell \geq \frac{r^2}{2\ell} \end{aligned}$$

which concludes the proof of (5).



ECCC ISSN 1433-8092

http://eccc.hpi-web.de