Electronic Colloquium on Computational Complexity, Report No. 159 (2018)

Expander-Based Cryptography Meets Natural Proofs

Igor C. Oliveira* Rahul Santhanam† Roei Tell‡

September 11, 2018

Abstract

We introduce new forms of attack on *expander-based cryptography*, and in particular on Goldreich's pseudorandom generator and one-way function. Our attacks exploit *low circuit complexity* of the underlying expander's neighbor function and/or of the local predicate. Our two key conceptual contributions are:

1. The security of Goldreich's PRG and OWF hinges, at least in some settings, on the *circuit complexity* of the underlying expander's neighbor function and of the local predicate. This sharply diverges from previous works, which focused on the *expansion* properties of the underlying expander and on the *algebraic* properties of the predicate.
2. We uncover new connections between long-standing open problems: Specifically, we tie *the security of Goldreich's PRG and OWF* both to the existence of *unbalanced lossless expanders* with low-complexity neighbor function, and to *limitations on circuit lower bounds* (i.e., natural proofs).

We prove two types of technical results that support the above conceptual messages. First, we *unconditionally break Goldreich's PRG* when instantiated with a specific expander (whose existence we prove), for a class of predicates that match the parameters of the currently-best "hard" candidates, in the regime of quasi-polynomial stretch. Secondly, conditioned on the existence of expanders whose neighbor functions have extremely low circuit complexity, we present attacks on Goldreich's generator in the *regime of polynomial stretch*. As one corollary, conditioned on the existence of the foregoing expanders, we show that either the parameters of natural properties for several constant-depth circuit classes *cannot be improved, even mildly*; or Goldreich's generator is insecure in the regime of a large polynomial stretch, *regardless of the predicate used*.

In particular, our results further motivate the investigation of average-case lower bounds against DNF-XOR circuits of exponential size, and of the parameters that can be achieved by affine/local unbalanced expanders.

*Department of Computer Science, University of Oxford. E-mail: igorcarb@gmail.com

†Department of Computer Science, University of Oxford. E-mail: rahul.santhanam@cs.ox.ac.uk

‡Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. E-mail: roei.tell@weizmann.ac.il

ISSN 1433-8092

Contents

1 Introduction

Theoretical results provide strong evidence that if secure cryptography is possible, then many fundamental primitives such as one-way functions (OWF) and pseudorandom generators (PRG) can be implemented with a dramatic level of efficiency and parallelism. Specifically, security against efficient adversaries can be achieved by functions where each output bit only depends on a constant number of input bits (see, e.g., [AIK06], and also [App14] for a survey of recent results).

A concrete type of such construction is a conjectured form of OWF that is based on any *expander graph* and on a *local predicate*. Specifically, about two decades ago, Goldreich [Gol00; Gol11] suggested the following candidate $\texttt{owf} : \{0,1\}^n \to \{0,1\}^n$. Fix any bipartite graph $[n] \times [n]$ of right-degree $\ell \leq O(\log(n))$ in which every set $S \subseteq [n]$ of size up to k on the right-hand side has at least (say) $1.01 \cdot |S|$ neighbors, and also fix a predicate $P : \{0,1\}^\ell \to \{0,1\}$. Then, given input $x \in \{0,1\}^n$, each output bit $\texttt{owf}(x)_i$ is computed by applying P to the bits of x at the ℓ neighbors of $i \in [n]$. The expected running-time of a naive algorithm for inverting $\texttt{owf}$ is at least $\exp(k)$ (see, e.g., [Gol11, Sec. 3.2] and [App14, Sec. 3.1]), and Goldreich conjectured that for an appropriate predicate P, no algorithm can perform significantly better.

In an extensive subsequent line of research (see, e.g., [Ale11; MST06; ABW10; BQ12; ABR16; BR13; CEMT14; OW14; FPV15; AL18; AR16], and also see [App16] for a related survey), Goldreich's construction was conjectured to yield not only a one-way function, but also a pseudorandom generator $\texttt{prg} : \{0,1\}^n \to \{0,1\}^m$. In fact, in some settings the two conjectures are essentially equivalent (see [AR16, Sec. 3]).

The question of whether Goldreich's constructions are secure is a long-standing open problem. Much research has focused on necessary requirements from the *predicate* and from the *parameters* in order for the construction to be secure. Let us, for simplicity of presentation, focus on the PRG. In this case, the locality ℓ cannot be too small: If we want a PRG with super-linear stretch, then we must use $\ell \geq 5$ [MST06];[1] and if we want stretch $m = n^k$ then ℓ must be at least (roughly) $3k$ (see [OW14, Thm. II.11]). Also, the predicate must have *high resilience* (i.e., all of the predicate's Fourier coefficients corresponding to sets of size at most $\Omega(\ell)$ are zero; see Definition 3.4) and high *rational degree* (this is a generalization of the requirement that the degree of the predicate as a polynomial $\mathbb{F}_2^\ell \to \mathbb{F}_2$ is $\Omega(\ell)$; see Definition 3.7).

The foregoing properties capture most existing attacks in the PRG setting. Indeed, as mentioned above, all these attacks exploit vulnerabilities of the *predicate* and of the *parameters*, but not of the underlying expander. In fact, prior to our work, the PRG was conjectured to be secure for *any underlying expander* with sufficiently good expansion properties. For reference, let us state such a strong form of conjectured security of the OWF, from a recent work by Applebaum and Raykov [AR16]. We say that a bipartite graph $G = ([n],[m],E)$ with right-degree ℓ is a $(k, 0.99)$-$\texttt{expander}$ if for every set $S \subseteq [m]$ on the right-hand side of size at most k, the number of neighbors of S in G is at least

[1] This impossibility result holds for *any* construction of a pseudorandom generator in $\mathcal{NC}^0$.

$0.99 \cdot \ell \cdot |S|$. [2] Then, the conjecture is the following:

Assumption 1.1 *(the strong EOWF assumption).*[3] *For a family $\mathcal{P} = \{P_\ell : \{0,1\}^\ell \to \{0,1\}\}_{\ell \in \mathbb{N}}$ of predicates, the* strong EOWF($\mathcal{P}$) assumption *is the following. For any $(n^{.99}, .99)$-expander $G = ([n],[m],E)$ of right-degree $\ell = \ell(n) \leq n^{o(1)}$ such that $n \leq m(n) \leq n^{\alpha \cdot \ell}$ (where $\alpha > 0$ is a sufficiently small universal constant), Goldreich's function instantiated with G and P_ℓ cannot be inverted by circuits of size $\exp(o(n^{.99}))$ with success probability $1/\mathrm{poly}(m)$.*

Applebaum and Raykov [AR16] suggested a suitable candidate predicate, which is the predicate $\mathsf{XOR\text{-}MAJ}(x) = \left(\oplus_{i=1,\ldots,\lfloor \ell/2 \rfloor} x_i\right) \oplus \left(\mathsf{MAJ}(x_{\lfloor \ell/2 \rfloor+1}, \ldots, x_\ell)\right)$; this predicate indeed has both high resiliency and high rational degree.

1.1 A high-level digest of our contributions

Our main contribution is a *new form of attack* on Goldreich's pseudorandom generator, which exploits *computational complexity* properties (and, in particular, circuit complexity properties) of the expander and/or of the predicate on which the generator is based. In particular, our distinguishers are algorithms associated with *natural properties*, in the sense of Razborov and Rudich [RR97]. (Recall that a natural property against a circuit class $\mathcal{C}$ is an efficient algorithm that distinguishes a random string, interpreted as a truth table, from truth tables of $\mathcal{C}$-circuits.)[4]

Using this form of attack, we unconditionally break Goldreich's generator when it is instantiated with a graph that has *optimal expansion properties*, and with predicates with *high resilience and rational degree*, for the setting of quasi-polynomial stretch. Using a known reduction of PRGs to OWFs (by [AR16]), it follows that Assumption 1.1 *does not* hold for some predicates with high resilience and rational degree (in fact, our predicates are variations on the specific XOR-MAJ predicate mentioned above). This attack is based on showing the existence of unbalanced lossless expanders whose neighbor function has "very low" circuit complexity. Our proof of the existence of such expanders, which uses certain *unconditional* PRGs that can be computed in a *strongly explicit* fashion, might be of independent interest. (See Section 1.2.)

In the regime of polynomial stretch, we put forward two assumptions about plausible extensions of known expander constructions in which the neighbor functions have

[2] We stress that lossless expansion (i.e., expansion to $\alpha \cdot \ell \cdot |S|$ vertices for $\alpha > 1/2$) is crucial in the PRG setting. To see this, note that one can duplicate a right-vertex in a $(k, 0.99)$-expander: This will produce a graph that, on the one hand, has good (but not lossless!) expansion properties, and on the other hand yields a corresponding PRG that is clearly insecure, regardless of the predicate.

[3] The name "strong" distinguishes the assumption from an assumption that refers to *specific* values for the stretch m, the locality ℓ, and the adversary size. For simplicity, we dropped two parameters, setting them to values that make the assumption weaker (and therefore harder to refute); specifically, we removed a parameter β that determined the expansion properties, and a parameter ϵ that determined the success probability of the algorithm.

[4] Natural properties are typically used to break *pseudorandom functions*, but the idea of using natural properties to break pseudorandom generators goes back to [RR97, Thm. 4.2]. Nevertheless, implementing this idea in our setting presents specific new challenges; for further discussion see Section 2.4.

even lower circuit complexity (compared to the expander mentioned above). Conditioned on *any* of the two assumptions, we show that exactly one of two options holds: Either the parameters of natural properties for certain restricted constant-depth circuit classes *cannot be improved, even mildly*; or Goldreich's generator is insecure in the regime of a large polynomial stretch, *regardless of the predicate used*. (See Section 1.3.)

The "take-home" messages. The two key conceptual contributions of our work, which follow from the results mentioned above, are the following:

1. The security of Goldreich's PRG and OWF hinges, at least in some settings, on the *circuit complexity of the expander's neighbor function and of the predicate*.

2. There are significant interdependencies between *the security of Goldreich's PRG and OWF*, the existence of *unbalanced lossless expanders* with low-complexity neighbor function, and *limitations on circuit lower bounds* (i.e., natural proofs).

In particular, it follows from Item (1) that the conjecture that the PRG and OWF are secure with *any sufficiently-expanding graph*, given an appropriate predicate, *might be too naive*. Moreover, as further explained below, our work connects the three research areas mentioned in Item (2), and the questions motivated by our results are closely related to existing results and important open problems in each area.

1.2 Unconditional results for quasi-polynomial stretch

Our main result for the setting of quasi-polynomial stretch is an attack that *unconditionally breaks* Goldreich's PRG when it is instantiated with a *specific expander that has optimal expansion properties*, and with a class of predicates that have both high resilience and high rational degree. Specifically:

Theorem 1.2 *(unconditional attack on Goldreich's PRG with quasi-polynomial stretch; informal). There exists a deterministic polynomial-time algorithm A that satisfies the following. Let $n \in \mathbb{N}$ be sufficiently large, let $m = n^{k=\text{poly}\log(n)}$, and let $\ell = O(k)$. Then, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n],[m],E)$ of right-degree ℓ such that for any predicate $P : \{0,1\}^{\ell} \to \{0,1\}$ that can be computed by an $\mathcal{AC}^0[\oplus]$ circuit of sufficiently small sub-exponential size, when Goldreich's generator is instantiated with the expander G and the predicate P, the algorithm A distinguishes the m-bit output of the generator from a uniform m-bit string (with gap $> 1/2$).*

In fact, we actually prove a more general theorem, which exhibits a trade-off between the locality ℓ and the size of the $\mathcal{AC}^0[\oplus]$ circuit for the predicate P (for a precise statement see Theorem 4.6). That is, we are able to break the generator even with much larger locality (e.g., $\ell = n^{.01}$), at the expense of using a more restricted predicate family, namely that of $\mathcal{AC}^0[\oplus]$ circuits of smaller size (e.g., polynomial size). We stress that even the latter predicate family is rich enough to contain predicates that have both high resilience and high rational degree (see below).

Recall that the property of the expander $[n] \times [m]$ that we exploit in our attack is that its *neighbor functions* (i.e., the functions $\Gamma_i : [m] \to [n]$ for $i \in [\ell]$) have *low circuit complexity*. The expander in Theorem 1.2 in particular has neighbor functions that can be computed by $\mathcal{AC}^0[\oplus]$ circuits of small sub-exponential size, and we prove its existence in Section 4.1 (see Section 2.2 for a high-level description).

Combining Theorem 1.2 with Applebaum and Raykov's reduction of expander-based PRGs to expander-based OWFs [AR16, Thm. 3.1] (i.e., they prove that if the OWF is secure, then the PRG is also secure), our attack also breaks Goldreich's OWF. Specifically, we say that a predicate $P : \{0,1\}^\ell \to \{0,1\}$ is *sensitive* if it is "fully sensitive" to one of its coordinates (i.e., if $P(x) = x_i \oplus P'(x)$ for some $i \in [\ell]$ and P' that does not depend on x_i). Then:

Corollary 1.3 *(unconditional attack on Goldreich's OWF with quasi-polynomial stretch; informal). There exists a probabilistic polynomial-time algorithm A' that satisfies the following. Let $n \in \mathbb{N}$ be sufficiently large, let $m' = n^{k'=\text{poly}\log(n)}$, and let $\ell = O(k')$. Then, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m'], E)$ of right-degree ℓ such that for any* sensitive *predicate $P : \{0,1\}^\ell \to \{0,1\}$ that can be computed by an $\mathcal{AC}^0[\oplus]$ circuit of sufficiently small sub-exponential size, when Goldreich's one-way function is instantiated with the expander G and the predicate P, the algorithm A inverts the function with success probability $\Omega(1/m'n)$.*

As immediate corollaries of Theorem 1.2 and of Corollary 1.3, we deduce that Assumption 1.1 does not hold for any sensitive predicate family that can be computed by $\mathcal{AC}^0[\oplus]$ circuits of sufficiently small sub-exponential size; and similarly, that the "PRG analogue" of Assumption 1.1, denoted $EPRG(\mathcal{P})$ in [AR16], does not hold for any predicate family that can be computed by $\mathcal{AC}^0[\oplus]$ circuits of sufficiently small sub-exponential size.

Recall that Applebaum and Raykov suggested the candidate predicate XOR-MAJ; we prove that when replacing majority by *approximate majority* (see Definition 4.9), the resulting predicate XOR-APPROX-MAJ still has both high resilience and high rational degree, and can also be computed by a polynomial-sized $\mathcal{AC}^0[\oplus]$ circuit (see Section 4.3.2). Thus, the predicate families in Theorem 1.2 and Corollary 1.3 contain predicates with high resilience and high rational degree, and even predicates that are variations on the "hard" candidate XOR-MAJ. [5]

Moreover, the predicate XOR-APPROX-MAJ does not even use the "full power" of the predicate family for which Theorem 1.2 allows us to break Goldreich's generator – the predicate XOR-APPROX-MAJ is computable by a circuit of polynomial size, whereas we can break the generator when the predicate can be computed by a circuit of sub-exponential size. We use this to our advantage by relying on the more general version of Theorem 1.2 (i.e., Theorem 4.6), which exhibits a trade-off between locality and the predicate class. Specifically, we obtain the following theorem, which breaks the generator even when the locality ℓ is large (e.g., $\ell = n^{\Omega(1)}$) and the predicate has high resilience and rational degree:

[5] Indeed, the main difference between XOR-MAJ and XOR-APPROX-MAJ seems to be in their *circuit complexity*, which corresponds to our main point that circuit complexity considerations are crucial for the security of Goldreich's PRG and OWF.

Theorem 1.4 *(breaking Goldreich's generator with* XOR-APPROX-MAJ *and high locality). There exists $s > 1$ such that the following holds. Let $n \in \mathbb{N}$, let $m = n^{k=(\log(n))^s}$, and let $c \cdot k \le \ell \le n^{1/c}$, where c is a sufficiently large constant. Then, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m], E)$ of right-degree ℓ and a predicate $P : \{0,1\}^\ell \to \{0,1\}$ with resilience $\Omega(\ell)$ and rational degree $\Omega(\ell)$ (i.e., the predicate* XOR-APPROX-MAJ*) such that the following holds: When Goldreich's generator is instantiated with the expander G and the predicate P, the output of the generator can be distinguished from a uniform string (with gap $> 1/2$) by a deterministic* $\mathrm{poly}(m)$*-time algorithm.*

1.3 Conditional results for large polynomial stretch

Recall that the conjectured "hardness" of Goldreich's PRG (i.e., Assumption 1.1) refers both to the regime of polynomial stretch and to the regime of quasi-polynomial stretch (as long as the locality is sufficiently large to support the corresponding stretch). Could it be that complexity-based attacks separate these two parameter regimes? That is, could the reason that our attacks from Section 1.2 work be that the stretch of the generator is super-polynomial?

As mentioned in Section 1.1 (and will be explained in Section 2), the underlying technical components in our complexity-based attacks are *unbalanced lossless expanders* $[n] \times [m]$ whose neighbor functions have low circuit complexity, and *natural properties* against weak circuit classes. Our main results for the polynomial-stretch regime are of the following form: If lossless expanders $[n] \times [n^{O(1)}]$ with constant degree and (specific) "very simple" neighbor functions exist, then exactly one of two cases holds:

1. Either the parameters of natural properties for certain well-studied weak circuit classes *cannot be improved*, even mildly; or
2. For a sufficiently large polynomial stretch, Goldreich's generator is insecure when instantiated with a specific expander, *regardless of the predicate used*.

We now present two plausible assumptions on existence of suitable expanders, which are essentially improvements or extensions of existing explicit constructions. Conditioned on each assumption, we will contrast the security of Goldreich's PRG with the possibility of extending natural proofs for some well-studied circuit class.

1.3.1 Affine expanders and DNF-XOR circuits

As motivation for our first assumption, let us recall two well-known *explicit* constructions of unbalanced lossless expanders, which were given by Ta-Shma, Umans, and Zuckerman [TUZ07], and later on by Guruswami, Umans, and Vadhan [GUV09]. We note that these two constructions are inherently different (the relevant construction from [TUZ07] is combinatorial, whereas the construction of [GUV09] is algebraic), and yet in both constructions the neighbor function of the expander can be computed by *single layer of parity gates* (see Section 5.1 for further details); we will call expanders with such a neighbor function *affine expanders*.

In the two foregoing affine expanders, the right-degree ℓ is polylogarithmic, and it is an open problem to improve the degree to be constant, which matches the degree of a random construction. However, a random construction is not necessarily affine. Our first assumption is that there indeed exists an *affine expander* with *constant degree*:

Assumption 1.5 *(expanders with an affine neighbor function; informal, see Assumption 5.4). There exists $\beta > 3$ such that for every constant $k \in \mathbb{N}$ and sufficiently large $n \in \mathbb{N}$, there exists an $(n^{.99}, 0.99)$-expander $G = ([n], [m = n^k], E)$ with right-degree $\ell = \beta \cdot k$ whose neighbor function $\Gamma_G : [m] \to ([n])^\ell$ can be computed by a single layer of parity gates.*

An unconditional proof of Assumption 1.5 will contrast the security of Goldreich's PRG with the possibility of extending the known natural properties for DNF-XOR *circuits of exponential size*. [6] Specifically, known lower bounds for DNF-XOR circuits yield natural properties useful against such circuits of size up to $2^{(1-o(1))\cdot n}$ (see Section 5.1.2).[7] Can these natural properties be extended to functions that are *approximated*, in the average-case sense, by DNF-XOR circuits of size $2^{\epsilon \cdot n}$, for some $\epsilon > 0$? This is the natural property that we contrast with the security of Goldreich's PRG:

Theorem 1.6 *(is Goldreich's generator insecure, or are natural properties for* DNF-XOR *circuits "non-extendable"?; informal statement). Suppose that Assumption 1.5 holds. Then, exactly one of the following two options holds:*

1. *For all $\epsilon > 0$, there does not exist a natural property for the class of functions that can be approximated with success $1/2 + o(1)$ by* DNF-XOR *circuits of size $2^{\epsilon \cdot n}$.*
2. *For a sufficiently large $k \in \mathbb{N}$, Goldreich's generator is insecure with stretch $m = n^k$ and locality $\ell = \beta \cdot k$, for some expander and* regardless of the local predicate used.

We stress that for any value of $\beta > 3$ such that Assumption 1.5 holds, Theorem 1.6 follows with that value of β. Also note that Cohen and Shinkar [CS16] specifically conjectured that strong average-case lower bounds for DNF-XOR circuits of size $2^{\Omega(n)}$ hold, and proved a similar statement for the related-yet-weaker model of parity decision trees. (Their proof for parity decision trees indeed yields a natural property; see Proposition 5.13.)

1.3.2 $\mathcal{NC}^0$ expanders and weak $\mathcal{AC}^0_4$ circuits

To motivate our next assumption, recall the recent explicit construction of lossless expanders by Viola and Wigderson [VW17] (which builds on the well-known construction of Capalbo *et al.* [CRVW02]). In this construction the neighbor function can be computed by an $\mathcal{NC}^0$ circuit, but this construction is only for *balanced* expanders, rather than unbalanced ones (see Theorem 5.18). The following assumption is that such a construction is possible also for unbalanced expanders:

[6] Recall that DNF-XOR circuits are depth-3 circuits that consist of a top OR gate, a middle layer of AND gates, and a bottom layer of parities above the inputs.

[7] Some of these natural properties actually run in slightly super-polynomial time, rather than in strictly polynomial time, but this issue is not crucial for our purpose of breaking Goldreich's PRG.

Assumption 1.7 *(expanders with $\mathcal{NC}^0$ neighbor functions; informal, see Assumption 5.4). There exists $\beta > 3$ such that for every constant $k \in \mathbb{N}$ and sufficiently large $n \in \mathbb{N}$, there exists an $(n^{.99}, 0.99)$-expander $G = ([n], [m = n^k], E)$ with right-degree $\ell = \beta \cdot k$ such that the neighbor function $\Gamma_G : [m] \to ([n])^\ell$ can be computed by an $\mathcal{NC}^0$ circuit.*

An unconditional proof of Assumption 1.7 will immediately break Goldreich's PRG in the polynomial-stretch regime by a complexity-based attack, when instantiated with a weak (but non-trivial) predicate class; see Proposition 5.25. But more importantly, such a proof will contrast the security of Goldreich's PRG with the possibility of extending the known natural properties for the class of exponential-sized $\mathcal{AC}^0$ *circuits of depth four with constant bottom fan-in and top fan-in.*

Since the precise trade-off between the parameters is a bit subtle, let us present the theorem in a simplified form (for a discussion of the more general setting, see Sections 5.2 and in particular Section 5.2.3). To do so, consider the (optimistic) possibility that in Assumption 1.7, there exists a single t such for any $k \in \mathbb{N}$ the arity of the $\mathcal{NC}^0$ circuit is t (i.e., the arity does not depend on k; as far as we are aware of, such a hypothesis is possible even with $t = 1$, which is the locality achieved in [VW17]). Relying on Håstad's switching lemma [Hås87], for any $c = O(1)$ there exists a natural property against depth-four circuits with top fan-in c, bottom fan-in t, and size $2^{\epsilon \cdot (n/\log(c))}$ for a tiny universal $\epsilon > 0$ (see Corollary 5.24). In the following theorem, the security of Goldreich's PRG is contrasted with the possibility of extending these natural properties to work against such circuits of size $2^{\beta \cdot (n/\log(c))}$ where $\beta > 3$.

Theorem 1.8 *(is Goldreich's generator insecure, or are natural properties for very restricted $\mathcal{AC}^0$ circuits "non-extendable"?; informal statement). Suppose that Assumption 1.7 holds and that the arity of the $\mathcal{NC}^0$ circuit equals $t = O(1)$ for any $k \in \mathbb{N}$. Then, exactly one of the following two options holds:*

1. *For any $c \in \mathbb{N}$, there does not exist a natural property for depth-four $\mathcal{AC}^0$ circuits with top fan-in c and bottom fan-in t and size $O\left(2^{\beta \cdot (n/\log(c))}\right)$.*
2. *For a sufficiently large $k \in \mathbb{N}$, Goldreich's generator is insecure with stretch $m = n^k$ and predicate locality $\ell = \beta \cdot k$, for some expander and regardless of the predicate used.*

Recall that Assumption 1.7 is parametrized by β and by the arity of the $\mathcal{NC}^0$ circuit; we stress that for *any* values of β and t such that Assumption 1.7 holds, we get a corresponding "win-win" theorem such as Theorem 1.8 (for further details see Section 5.2). We also stress that both the natural properties that we can unconditionally prove and the natural properties referred to in Theorem 1.8 are for circuits of exponential size $2^{\Theta(n/\log(c))}$, and the difference is in the universal constant hidden in the Θ-notation.

1.3.3 An important comment about expander constructions

As mentioned in Section 1.1, the *explicit* construction of highly unbalanced lossless expanders is a long-standing open problem, regardless of the circuit complexity of their neighbor function (see, e.g., [CRVW02], [Vad12, Prob. 5.36 & 6.35], and [Wig18, Chap. 8.7]). Assumptions 1.5 and 1.7, however, *do not concern explicit constructions* of expanders, but only assume their *existence*; in particular, the circuit family for the neighbor function of the graph may be *non-uniform*. (This is indeed the case for our construction of expanders in the quasi-polynomial stretch regime.)

1.4 Organization

In Section 2 we describe our proof approach, in high-level, and give overviews of some of the proofs. Section 3 contains preliminary definitions. In Section 4 we prove our results for the regime of quasi-polynomial stretch, and in Section 5 we prove our results for the regime of polynomial stretch.

2 Overviews of the proofs

2.1 The general form of attack

A natural property for a class $\mathcal{F}$ of functions is a deterministic polynomial-time algorithm that rejects all truth-tables of functions from $\mathcal{F}$, but accepts the truth-tables of almost all functions.[8] Indeed, a natural property for $\mathcal{F}$ exists only if almost all functions are *not* in $\mathcal{F}$. We will show how to use natural properties to break Goldreich's pseudorandom generator.

The key step in our proofs is to show, for every fixed $x \in \{0,1\}^n$, that $\mathtt{prg}(x)$ is the truth-table of a function from some class $\mathcal{F}$ of "simple" functions (e.g., $\mathtt{prg}(x)$ is the truth-table of a small constant-depth circuit). When we are able to show this, it follows that a natural property for $\mathcal{F}$ can distinguish the outputs of the PRG from uniformly-chosen random strings: This is because the natural property rejects any string in the output-set of the PRG (which is the truth-table of a function in $\mathcal{F}$), but accepts a random string, with high probability. (The general idea of using natural properties to break PRGs in this manner goes back to the original work of [RR97].)

Recall that Goldreich's PRG (i.e., the function $\mathtt{prg}$) is *always* a very "simple" function, since each output bit depends on a few (i.e., $\ell \ll n$) input bits. However, in order for our idea to work, we need that a *different* function (i.e., not the function $\mathtt{prg}$) will be simple: Specifically, for every *fixed* input x, we want that the function $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ such that $g_x(i) = \mathtt{prg}(x)_i$ will be "simple". That is, for a *fixed* "seed" x for the PRG, the function g_x gets as input an index i of an output bit, and computes the i^{th} output bit of $\mathtt{prg}(x)$ *as a function of* i. Intuitively, given $i \in [m]$, the function g_x needs to compute three different objects, successively:

- The neighbors $\Gamma_G(i)$ of the vertex $i \in [m]$ in G.

[8] Throughout the paper, we identify a natural property with the "constructive" algorithm that recognizes the property (see Definition 3.8, which also relaxes the requirement from rejecting "almost all" functions to rejecting "most" functions).

- The projections of the (fixed) string x on locations $\Gamma_G(i)$.
- The output of the predicate P on $x\restriction_{\Gamma_G(i)}$.

The proofs of our main theorems consist of showing instantiations of Goldreich's generator (i.e., choices for an expander and a predicate) such that g_x is a function from a class against which we can construct natural properties. We mention that the requirements from the function g_x are technically reminiscent of the requirements from constructions of pseudorandom functions in the work of Applebaum and Raykov [AR16] (although there are several key differences, and the conceptual framework is very different); for further details see Section 2.4.

2.2 The setting of quasi-polynomial stretch

The proof of Theorem 1.2 consists of showing that for a suitable expander G, and for any predicate P computable by an $\mathcal{AC}^0[\oplus]$ circuit of sufficiently small sub-exponential size, the function g_x can be computed by an $\mathcal{AC}^0[\oplus]$ circuit of sufficiently small sub-exponential size. Natural properties for such circuits, based on the lower bounds by Razborov and Smolensky [Raz87; Smo87], are well-known (see, e.g., [RR97; CIKK16]).

To describe the instantiations and the construction of an $\mathcal{AC}^0[\oplus]$ circuit for g_x, let $n \in \mathbb{N}$, and let $m = 2^{(\log(n))^k}$, for a sufficiently large k. The first technical component that we need is an expander graph G such that the function $i \mapsto \Gamma_G(i)$ can be computed by a sub-exponential sized $\mathcal{AC}^0[\oplus]$ circuit. We show that there exists such a graph, with essentially optimal parameters:

Theorem 2.1 *(strongly-explicit lossless expander in $\mathcal{AC}^0[p]$; see Theorem 4.5). There exists a universal constant $d_G \in \mathbb{N}$ such that the following holds. For any $k \in \mathbb{N}$ and sufficiently large n and $m = 2^{(\log(n))^k}$, there exists a $(n^{0.99}, 0.99)$-expander $G = ([n], [m], E)$ of right-degree $\ell = O(\log(m)/\log(n))$, and an $\mathcal{AC}^0[\oplus]$ circuit $C_G : \{0,1\}^{\log(m)} \to \{0,1\}^{\ell \cdot \log(n)}$ of depth d_G and size* $\mathrm{poly}(n)$ *such that for every $i \in [m]$ it holds that $C_G(i)$ outputs the list of ℓ neighbors of i in G.*

We stress that the depth d_G of the circuit in Theorem 2.1 does not depend on the relation between m and n, which is what will allow us to have a natural property for the circuit C_G. Specifically, recall that we have natural properties against $\mathcal{AC}^0[\oplus]$ circuits of depth d_G over $\ell_m = \log(m)$ input bits of sub-exponential size $2^{\Omega(\ell_m^{1/2d_G})}$. The size of C_G is $\mathrm{poly}(n)$, and thus if we take $m = 2^{(\log(n))^k}$, for a sufficiently large k, then the size of C_G is a sufficiently small sub-exponent in its input length $\log(m)$.

In high-level, our construction of the expander in Theorem 2.1 is as follows. Our starting point is the well-known fact that a random graph is, with high probability, a good lossless bipartite expander (see Theorem 3.2). The first step is to construct an *efficient test* that gets as input a string $G \in \{0,1\}^{m'}$, where $m' = m \cdot \ell \cdot \log(n)$, considers G as the incidence-list of a graph, and decides whether or not G is an $(n^{.99}, .99)$-expander. We show that such a test can be implemented by a CNF of size 2^n (see

Claim 4.2). Hence, a pseudorandom generator for CNFs of size 2^n outputs, with high probability, a good expander. Specifically, we will use the pseudorandom generator of Nisan [Nis91], which has seed length $\text{poly}(n)$. Thus, for some fixed "good" seed s, the output $NW(s) \in \{0,1\}^{m'}$ of the generator on s is an $(n^{.99}, .99)$-expander.

Our next step is to show that the expander represented by $NW(s)$ has neighbor functions that can be computed by an $\mathcal{AC}^0[\oplus]$ circuit. In fact, we will show that there exists a circuit that gets as input the index $i \in \{0,1\}^{\log(m')}$ of a bit in $NW(s)$ and outputs $NW(s)_i$. To do so we can rely, for instance, on the recent work of Carmosino *et al.* [CIKK16] , who showed that Nisan's generator can be made "strongly-explicit": That is, there exists an $\mathcal{AC}^0[\oplus]$ circuit of polynomial size that gets as input a seed z and an index i of an output bit, and computes the i^{th} output bit of the generator on seed z. [9] By "hard-wiring" a "good" seed s into the latter circuit, we obtain an $\mathcal{AC}^0[\oplus]$ circuit of size $\text{poly}(n)$ that computes the output bits of the expander $NW(s)$. Indeed, a crucial point is that we did not algorithmically look for a good seed s, but rather non-uniformly fixed a "good" seed and "hard-wired" it into the circuit.

Given this expander construction, g_x can compute $i \mapsto \Gamma_G(i)$ in sub-exponential size, and we now need g_x to compute the projections of x on locations $\Gamma_G(i)$. To do so we simply "hard-wire" the entire string x into g_x. Specifically, after computing the function $i \mapsto \Gamma_G(i)$, the circuit now has the $\ell \cdot \log(n)$ bits of $\Gamma_G(i)$; it then uses ℓ depth-two formulas, each over $\log(n)$ bits and of size n, to compute the mapping $\Gamma_G(i) \mapsto x\restriction_{\Gamma_G(i)}$ by brute-force. This increases the size of the circuit for g_x by $\ell \cdot n < n^2$ gates, which is minor compared to the size $\text{poly}(n)$ of C_G from Theorem 2.1.

Finally, the circuit g_x has now computed the ℓ bits corresponding to $x\restriction_{\Gamma_G(i)}$, and needs to compute the predicate $P : \{0,1\}^\ell \to \{0,1\}$ on these bits. To get the circuit to be of sufficiently small sub-exponential size, we require that the predicate can be computed by a sufficiently small sub-exponential sized $\mathcal{AC}^0[\oplus]$ circuit. Specifically, we want that for some d_P, the predicate P can be computed by an $\mathcal{AC}^0[\oplus]$ circuit of depth d_P and size 2^{ℓ^ϵ}, for a sufficiently small $\epsilon < 1/2(d_G + d_P + 2)$. We thus obtain a circuit for g_x of depth $d = d_G + d_P + 2$ and of size $O\left(2^{\ell^\epsilon}\right) < 2^{\log(m)^{1/2d}}$, [10] which is sufficiently small such that we have natural properties against it (see Theorem 3.9).

2.3 The setting of large polynomial stretch

Why are the results in Section 2.2 applicable only to the setting of quasi-polynomial stretch? The main bottleneck is the expander construction in Theorem 2.1, which is an $\mathcal{AC}^0[\oplus]$ circuit. Specifically, since we only know of natural properties against $\mathcal{AC}^0[\oplus]$ circuits of at most sub-exponential size, and since the circuit that we obtain is of size at least n (because we hard-wire $x \in \{0,1\}^n$ to the circuit), we were forced to take $m = n^{\text{poly}\log(n)}$ such that n will be a small sub-exponential function of $\log(m)$.

[9] A similar observation has appeared in other works, such as in [RR97, Thm. 4.2].

[10] For this calculation we assumed that 2^{ℓ^ϵ} dominates the size of the circuit (since the size of C_G is already sufficiently small); and we used the fact that $\ell = O(\log(m)/\log(n)) < \log(m)$, and that $\epsilon < 1/2d$ is sufficiently small).

In this section we circumvent this obstacle by using the hypothesized existence of expanders whose neighbor functions have "extremely simple" circuits. For simplicity, in the current high-level overview we present the attacks that are based on the existence of an expander as in Assumption 1.5; that is, a lossless expander $G = ([n], [m = n^k], E)$ of right-degree $\ell = O(k)$ whose neighbor function is an *affine function* (i.e., each output bit is a parity of input bits). The ideas that underlie the attacks that are based on expanders whose neighbor function is an $\mathcal{NC}^0$ circuit (as in Assumption 1.7) are similar, yet require a slightly more subtle parametrization (see Section 5.2).

Consider an instantiation of Goldreich's predicate with expander G as above and with a predicate $P : \{0,1\}^\ell \to \{0,1\}$ that can be computed by a CNF of size $2^{\delta \cdot \ell}$, where δ can be an arbitrarily large constant compared to k (or even $\delta = 1$, which allows for *any* predicate). In this case, for any $x \in \{0,1\}^n$, the output $\mathtt{prg}(x)$ of the generator on x is a truth-table of a function g_x over an input $i \in \{0,1\}^{\log(m)}$ that can be computed as follows. One layer of *parity gates* maps $i \in [m]$ to $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$ (this uses our assumption about the expander). Then, ℓ copies of a DNF over $\log(n)$ bits and of size n map the names of the ℓ vertices to $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$, i.e., we project the bits of x that feed the predicate P (this DNF is essentially a "hard-wiring" of x into g_x). Finally, the CNF that computes P of size $2^{\delta \cdot \ell}$ maps $x\restriction_{\Gamma_G(i)}$ to the value $P(x\restriction_{\Gamma_G(i)})$. After collapsing a layer that connects the top CNF and the DNFs, we obtain an AND-OR-AND-XOR circuit g_x over $\ell_m = \log(m)$ input bits of size $O\big(\ell \cdot \log(n) + \ell \cdot n + 2^{\ell \cdot \delta}\big) = O\big(2^{\ell_m / k}\big)$ with top fan-in $2^{\delta \cdot \ell} = 2^{O(\delta \cdot k)}$.

When $\delta > 0$ is sufficiently small, we are able to unconditionally construct a natural property against circuits as above. However, the main point (i.e., Theorem 1.6) comes when considering the case $\delta = 1$; that is, *any* predicate $P : \{0,1\}^\ell \to \{0,1\}$. In this case, we first use the discriminator lemma of [HMP+93] to deduce that g_x can be $(1/2 + 1/2^{O(k)})$-approximated by a DNF-XOR circuit of size $O\left(2^{\ell_m / k}\right)$. Now (still under Assumption 1.5), exactly one of two options holds. The first option is that there exists a natural property for functions on ℓ_m input bits that can be $(1/2 + o(1))$-approximated by DNF-XOR circuits of size $2^{\Omega(\ell_m)}$; in this case, by taking k sufficiently large so that $2^{\ell_m / k}$ is sufficiently small, the natural property breaks the generator. The other option is that no such natural property exists, despite the fact that natural properties exist both for functions computed (in the worst-case) by DNF-XOR circuits of size $2^{(1 - o(1)) \cdot \ell_m}$, and for functions approximated (even weakly) by parity decision trees of such size. This completes the sketch of the proof of Theorem 1.6.

2.4 Natural properties and expander-based one-way functions

As mentioned in Footnote 4, the fact that natural properties can be used to break *pseudorandom functions*, and in particular pseudorandom functions that are obtained when considering the outputs of a pseudorandom generator as truth-tables, goes back to the original work of Razborov and Rudich [RR97]. Nevertheless, to our knowledge, our work is the first to implement this idea in the context of *expander-based pseudorandom generators*; moreover, as detailed in Sections 2.2 and 2.3, in this setting it is a challenge to show that for *specific instantiations* of the generator (i.e., specific expanders and predicate classes) the generator's output is a truth-table of a *simple* function.

It is also known that pseudorandom generators can be transformed into pseudorandom functions with comparable complexity. Such transformations have the potential to yield results conceptually similar to our results, relying on alternative proof approaches;[11] but unfortunately the known transformations do not preserve the complexity of the original function sufficiently well. Specifically, the transformation of Goldreich, Goldwasser, and Micali [GGM86] incurs a blow-up in the circuit depth, which does not maintain the constant depth of the original circuit; whereas the constructions of Applebaum and Raykov [AR16] either yield only a *weak* PRF [AR16, Sec. 4.1] (which is not broken by natural properties, in general) or require complicated computations, which they implement using majority gates (i.e., in $\mathcal{TC}^0$, for which natural properties are neither known nor conjectured to exist; see [AR16, Sec. 4.3]). For further discussion of the work of Applebaum and Raykov, see Appendix B.

3 Preliminaries

We use $[n]$ to denote the set $\{1,\ldots,n\}$. For functions $f,g : \{0,1\}^t \to \{0,1\}$ and $\gamma \geq 0$, we say that f *γ-approximates* g if $\Pr_{x\sim\mathbf{u}_t}[f(x) = g(x)] \geq \gamma$, where $\mathbf{u}_t$ denotes the uniform distribution over $\{0,1\}^t$. We now define the three main object that are studied in the current work: Expanders (possibly with a simple neighbor function), Goldreich's pseudorandom generator and the properties of its predicates, and natural proofs.

3.1 Expander graphs

Throughout the paper, when we refer to expander graphs, we will mean highly unbalanced bipartite lossless expanders. More specifically:

Definition 3.1 *(lossless bipartite expanders). Let $G = ([n],[m],E)$ be a bipartite graph with left vertices $[n]$, right vertices $[m]$, and right-degree ℓ. We say that G is an $(r,0.99)$-*expander *if for every set $S \subseteq [m]$ of size at most r it holds that $|\Gamma(S)| \geq 0.99\cdot\ell\cdot|S|$, where $\Gamma(S)$ denotes the set of neighbors of S in G. The* neighbor function *$\Gamma_G : \{0,1\}^{\log(m)} \to \{0,1\}^{\ell\cdot\log(n)}$ of G maps any $i \in [m]$ to its ℓ neighbors in $[n]$ in G.*

In our constructions, we assume an ordering of the neighbors of each right-vertex of G, and allow parallel edges among vertices. Recall that a random graph is, with high probability, a good lossless expander; that is:

Theorem 3.2 *(a random graph is a good lossless expander). Let $n \leq m \in \mathbb{N}$ and $\ell \in \mathbb{N}$ such that $c\cdot\frac{\log(m)}{\log(n)} \leq \ell \leq n^{1/c}$, where c is a sufficiently large constant. Let $G = ([n],[m],E)$ be a random bipartite graph of right-degree ℓ; that is, the ℓ neighbors of each $i \in [m]$ are chosen*

[11] Since if the (alleged) PRG is a "simple" function, then the resulting (alleged) PRF is also "simple", in which case it can be broken by a corresponding natural property.

uniformly in $[n]$ (allowing repetitions). Then, with probability at least $1 - 1/\mathrm{poly}(n)$ it holds that G is an $(n^{0.99}, 0.99)$-expander.[12]

The proof of Theorem 3.2 is a straightforward probabilistic argument, and in fact a more general theorem has been proved in [AR16, Lem. 2.5]. For completeness, we include a proof of Theorem 3.2 in Appendix A.

3.2 Goldreich's pseudorandom generator and properties of predicates

Let us formally define Goldreich's pseudorandom generator:

Definition 3.3 *(Goldreich's PRG). Let $n \in \mathbb{N}$, let $m = m(n)$ be a* stretch *parameter, and let $d = d(n)$ be a* locality *parameter. Let $P : \{0,1\}^d \to \{0,1\}$ be a* predicate*, and let G be a bipartite graph (typically a "good" expander) with left vertex-set $[n]$, right vertex-set $[m]$, and right-degree d. Then,* Goldreich's PRG using expander G and predicate P *is the function* $\mathtt{prg} : \{0,1\}^n \to \{0,1\}^m$ *defined as follows: For every $x \in \{0,1\}^n$ and $i \in [m]$, the i^{th} output bit of* $\mathtt{prg}$ *equals $P(x\restriction_{\Gamma_G(i)})$, where $\Gamma_G(i)$ is the set of neighbors of $i \in [m]$ in the graph G.*

We now define several standard algebraic properties, which are relevant in the study of suitable predicates for Goldreich's generator (as mentioned in Section 1). For a predicate $P : \{0,1\}^\ell \to \{0,1\}$, the Fourier coefficients of P are given by $\widehat{P}(\gamma) = \mathbb{E}_{z \in \{0,1\}^\ell}[(-1)^{P(z)+\langle z,\gamma\rangle}]$, where $\gamma \in \{0,1\}^\ell$.

Definition 3.4 *(resilience). We say that P has* resilience k *if $\widehat{P}(\gamma) = 0$ for every $\gamma \in \{0,1\}^\ell$ of Hamming weight at most k. In other words, P is uncorrelated with parities of size at most k.*

Definition 3.5 *(degree). We use $\deg_{\mathbb{F}_2}(P)$ to denote the $\mathbb{F}_2$-degree of a predicate $P \colon \{0,1\}^\ell \to \{0,1\}$ viewed as a polynomial.*

Definition 3.6 *(bit-fixing degree). An ℓ-bit predicate P has r-bit fixing degree e if for every restriction $\rho \colon [\ell] \to \{0,1,*\}$ that fixes at most r variables, $P\restriction_\rho$ has $\mathbb{F}_2$-degree at least e.*

Definition 3.7 *(rational degree). For a predicate $P \colon \{0,1\}^\ell \to \{0,1\}$, we let the* rational degree *of Q, denoted by $\mathrm{rdeg}_{\mathbb{F}_2}(Q)$, be the minimum $e \in \mathbb{N}$ such that there exist predicates $Q, R \colon \{0,1\}^\ell \to \{0,1\}$, not both identically zero, such that $\deg_{\mathbb{F}_2}(Q), \deg_{\mathbb{F}_2}(R) \le e$ and*

$$P(z)Q(z) = R(z) \ \text{ for every } \ z \in \{0,1\}^\ell.$$

Clearly, a rational degree lower bound or a bit-fixing degree lower bound imply an equivalent degree lower bound. In addition, as noted for instance in [AL18], rational degree $\geq e$ implies r-bit fixing degree of at least $e - r$ for any $r < e$. The rational degree of any predicate over ℓ input bits is upper bounded by $\lceil \ell/2 \rceil$ (cf. [Car10; AL18]).

[12]The polynomial power in the success probability depends on c and can be made arbitrarily large.

3.3 Natural properties

Let us recall the definition of natural properties by Razborov and Rudich [RR97].

Definition 3.8 *(natural properties). Let $\mathcal{F} \subseteq \{\{0,1\}^* \to \{0,1\}\}$ be a class of Boolean functions. A deterministic polynomial-time algorithm is a* natural property algorithm *for $\mathcal{F}$ if for every sufficiently large $n \in \mathbb{N}$, the algorithm satisfies the following:*

1. *The algorithm rejects every truth-table of $f \in \mathcal{F}$ over n input bits (when f is given to the algorithm as a string in $\{0,1\}^{2^n}$).*
2. *The algorithm accepts more than half of the strings in $\{0,1\}^{2^n}$.*

We will sometimes slightly abuse Definition 3.8 by referring to natural properties that run in super-polynomial time (e.g., in quasi-polynomial time), or that accept less than half of the strings in 2^n (say, a constant fraction of the strings). We will clearly indicate the deviation from Definition 3.8 when we do so.

It is well-known that, relying on the results of Razborov [Raz87] and Smolensky [Smo87], there exist natural properties for $\mathcal{AC}^0[\oplus]$ of subexponential size (see, e.g., [RR97; CIKK16]). Since we will crucially use these natural properties in Section 4, let us formally state the parameters of the natural properties:

Theorem 3.9 *(natural properties for $\mathcal{AC}^0[\oplus]$ of subexponential size; [RR97; CIKK16]). For any constant $d \in \mathbb{N}$, let $\mathcal{F} \subseteq \{\{0,1\}^* \to \{0,1\}\}$ be the class of functions that satisfy the following: For every $\ell \in \mathbb{N}$, all functions in $\mathcal{F} \cap \{\{0,1\}^\ell \to \{0,1\}\}$ are computable by circuits of depth d with parity gates of size at most $2^{\eta \cdot \ell^{1/2d}}$, where $\eta > 0$ is a sufficiently small constant. Then, there exists a natural property algorithm for $\mathcal{F}$.*

4 Unconditional results in the setting of quasi-polynomial stretch

Our goal in this section is to prove Theorem 1.2, which shows that Goldreich's generator can be broken in the quasi-polynomial stretch regime for a large class of predicates. As a first step, in Section 4.1 we will prove Theorem 2.1, which asserts the existence of an expander whose "neighbor functions" can be computed by relatively-small $\mathcal{AC}^0[\oplus]$ circuits. In Section 4.2, we prove Theorem 1.2. And finally, in Section 4.3 we prove that the class of predicates for which we break the generator is quite rich, and in particular contains predicates with high rational degree and high resilience.

4.1 Expanders with $\mathcal{AC}^0[\oplus]$ neighbor functions

In this section we construct a (non-explicit) bipartite lossless expander whose "neighbor function" can be computed by a sub-exponential sized $\mathcal{AC}^0[\oplus]$ circuit. Throughout the section we consider an incidence-list representation of expanders; that is

Definition 4.1 *(incidence-list representation). For $n, m, \ell \in \mathbb{N}$ such that $\ell \le n$, let $m' = m \cdot \ell \cdot \lceil \log(n) \rceil$. We identify any string $G \in \{0,1\}^{m'}$ with a bipartite graph $[n] \times [m]$ of right-degree ℓ such that for $v \in [m]$ and $i \in [\ell]$, the j^{th} neighbor of v is encoded by the bits in $G_{(v-1)\cdot\ell\cdot\lceil\log(n)\rceil+1}, ..., G_{v\cdot\ell\cdot\lceil\log(n)\rceil}$.*

Let us start the proof by showing that for many natural settings of the parameters (e.g., $m \le 2^{n^{o(1)}}$ and $\ell \le n^{o(1)}$) we can test whether a string $G \in \{0,1\}^{m'}$ is an $(n^{.99}, .99)$-expander using a CNF of size at most 2^n:

Claim 4.2 *(testing whether a graph is an expander). Let $n, m, \ell \in \mathbb{N}$ such that $m \ge n$ and $\log(m) + 2\ell \cdot \lceil \log(n) \rceil \le n^{.01}$, and let $m' = m \cdot \ell \cdot \lceil \log(n) \rceil$. Then, there exists a CNF of size at most 2^n that gets as input a string $G \in \{0,1\}^{m'}$ and accepts if and only if G is an incidence-list representation of an $(n^{0.99}, .99)$-expander $[n] \times [m]$ of right-degree ℓ.*

Proof. The CNF is a conjunction of the following set of tests:

- For every $k = 1, ..., n^{0.99}$,
- For every set $S \subseteq [m]$ of size at most k,
- The number of neighbors of S is at least $0.99 \cdot \ell \cdot |S|$.

For a fixed $S \subseteq [m]$ of size $k \le n^{0.99}$, the decision of whether or not $|\Gamma(S)| \ge 0.99 \cdot \ell \cdot |S|$ is a function of $k \cdot \ell \cdot \lceil \log(n) \rceil$ bits, and can therefore be implemented by a CNF of size $2^{k\cdot\ell\cdot\lceil\log(n)\rceil}$. There are at most $k \cdot \binom{m}{k}$ sets to consider, and therefore the final CNF is of size at most $k \cdot \binom{m}{k} \cdot 2^{k\cdot\ell\cdot\lceil\log(n)\rceil} < 2^{\log(k)+k\cdot\log(m)+k\cdot\ell\cdot\lceil\log(n)\rceil} \le 2^n$. ■

Our next step is to use the construction of Carmosino *et al.* [CIKK16] to show that Nisan's generator can be made "strongly-explicit":

Proposition 4.3 *(a strongly-explicit NW-generator). There exists a constant d such that the following holds. Let $n, m' \in \mathbb{N}$ such that $n \le m' \le 2^n$, let $t = \mathrm{poly}(n)$ for a sufficiently large polynomial, and let $\epsilon = 2^{-n}$. Then, there exists a pseudorandom generator $NW : \{0,1\}^t \to \{0,1\}^{m'}$ with error ϵ for CNFs of size 2^n that satisfies the following. There exists a polynomial-sized circuit $E : \{0,1\}^t \times \{0,1\}^{\lceil\log(m')\rceil} \to \{0,1\}$ of depth d with parity gates such that for every $s \in \{0,1\}^t$ and $i \in [m']$ it holds that $E(s,i)$ outputs the i^{th} bit of $NW(s)$.*

Proof. Our generator $NW : \{0,1\}^t \to \{0,1\}^{m'}$ will be the Nisan-Wigderson generator. Specifically, for a sufficiently large $k \ge 6$ and $r = n^k$, let $f : \{0,1\}^r \to \{0,1\}$ be the parity function such that f is (ϵ/m')-average-case hard for circuits of depth five and of size $2^{O(n)}$. [13] We will use a combinatorial design $S_1, ..., S_{m'} \subseteq [t]$ with parameters as in the following lemma, which is due to Carmosino *et al.* [CIKK16].

[13] For $S = 2^{O(n)}$ and a sufficiently large integer $k \ge 6$, the correlation of a size-S circuit of depth five with the parity of $v = n^k$ variables is at most $2^{-\Omega(v/(\log(S))^5)} \le \epsilon/m'$ (see [Hås14]).

Lemma 4.4 (*strongly-explicit NW-designs in $\mathcal{AC}^0[p]$; see* [CIKK16, Thm. 3.3]) *Let p be any prime. There exists a constant $d_{MX} \geq 1$ such that for any $r \in \mathbb{N}$ and $m' < 2^r$ there exists a combinatorial design $S_1, ..., S_{m'} \subseteq [t]$ with $t = O(r^2)$, each $|S_i| = r$, and $|S_i \cap S_j| \leq \lceil \log(m') \rceil$ for all $1 \leq i \neq j \leq m'$, that satisfies the following. The function $MX_{NW} : \{0,1\}^t \times \{0,1\}^{\lceil \log(m') \rceil} \to \{0,1\}^r$, defined by $MX_{NW}(s,i) = s\restriction_{S_i}$, is computable by an $\mathcal{AC}^0[p]$ circuit of size $O(\log(m') \cdot r^3 \cdot \log(r))$ and depth d_{MX}.*

By a standard argument that follows Nisan and Wigderson [NW94], the generator G instantiated with the function f and the design in Lemma 4.4 is ϵ-pseudorandom for CNFs of size $2^{O(n)}$. The circuit $E : \{0,1\}^t \times \{0,1\}^{\lceil \log(m') \rceil} \to \{0,1\}$ acts as follows. Given inputs (s,i), the circuit E first feeds (s,i) to the circuit from Lemma 4.4 to compute $s\restriction_{S_i}$, and then computes the parity of $s\restriction_{S_i}$ (using a single parity gate). The depth of E is $d = d_{MX} + 1$, and its size is $\text{poly}(t, \log(m')) = \text{poly}(n)$. ■

Finally, we prove Theorem 2.1 by combining Claim 4.2 and Proposition 4.3, and "hard-wiring" a good seed s into the circuit E from Proposition 4.3. That is:

Theorem 4.5 *(strongly-explicit lossless expander in $\mathcal{AC}^0[p]$). There exists a constant $d_G \in \mathbb{N}$ such that the following holds. Let $n \leq m \in \mathbb{N}$ such that $m \leq 2^{n^{o(1)}}$, and let $\ell \in \mathbb{N}$ such that $c \cdot \frac{\log(m)}{\log(n)} \leq \ell \leq n^{1/c}$, where c is a sufficiently large constant. Then, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n],[m],E)$ of right-degree ℓ, and an $\mathcal{AC}^0[\oplus]$ circuit $C_G : \{0,1\}^{\log(m)} \to \{0,1\}^{\ell \cdot \log(n)}$ of depth d_G and size $\text{poly}(n)$, such that for every $i \in [m]$ it holds that $C_G(i)$ outputs the list of ℓ neighbors of i in G.*

Proof. Let $m' = m \cdot \ell \cdot \lceil \log(n) \rceil$. By Theorem 3.2, a random string in $\{0,1\}^{m'}$ is an $(n^{0.99}, 0.99)$-expander, with high probability. By Claim 4.2, such expanders can be recognized by CNFs of size 2^n. Thus, considering the pseudorandom generator NW from Proposition 4.3 for such CNFs, with high probability over choice of seed $s \in \{0,1\}^{\text{poly}(n)}$ for NW it holds that $NW(s)$ is an $(n^{0.99}, 0.99)$-expander.

Let us now fix such a good seed $s \in \{0,1\}^{\text{poly}(n)}$, and a corresponding expander $G = NW(s)$. When we "hard-wire" the seed s into the circuit E from Proposition 4.3, we obtain a circuit $E_s : \{0,1\}^{\lceil \log(m') \rceil} \to \{0,1\}$ that on input $j \in [m']$ outputs the j^{th} bit in the representation of G as a bit-string. Thus, our final circuit C_G gets as input $i \in \{0,1\}^{\lceil \log(m) \rceil}$, and uses $\ell \cdot \lceil \log(n) \rceil$ copies of E_s to output the $\ell \cdot \lceil \log(n) \rceil$ bits in the representation of G that correspond to the neighbors of $i \in [m]$ (i.e., $C_G(i) = E_s((i-1) \cdot \lceil \log(n) \rceil), ..., E_s(i \cdot \ell \cdot \lceil \log(n) \rceil - 1))$. The size of C_G is $\ell \cdot \lceil \log(n) \rceil$ times the size of E_s, and is thus upper-bounded by $\text{poly}(n)$; the depth d_G of C_G is just the depth of E_s, which is the universal constant d from Proposition 4.3. ■

4.2 Proof of Theorem 1.2

Relying on Theorem 4.5, we now prove Theorem 1.2. As explained in Section 2.2, the main step is to show that, when Goldreich's generator is instantiated with the expander from Theorem 4.5 and with any predicate that can be computed by an $\mathcal{AC}^0[\oplus]$

circuit of sufficiently small sub-exponential size, the output of the generator (on any seed) is the truth-table of an $\mathcal{AC}^0[\oplus]$ circuit of small sub-exponential size.

Theorem 4.6 *(Theorem 1.2, restated). For any $d_P \in \mathbb{N}$ and sufficiently small $\epsilon = \epsilon(d_P) > 0$ and sufficiently large $k = k(d_P) > 1$, there exists a deterministic polynomial-time algorithm A that satisfies the following. For any sufficiently large $n \in \mathbb{N}$, let $m = 2^{(\log(n))^k}$, and let $c \cdot (\log(n)^{k-1}) \leq \ell \leq n^{1/c}$, where c is a sufficiently large constant. Then, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m], E)$ of right-degree ℓ such that the following holds:*

For any predicate $P : \{0,1\}^\ell \to \{0,1\}$ that can be computed by an $\mathcal{AC}^0[\oplus]$ circuit of depth d_P and size at most $2^{(\log(m))^\epsilon}$, when Goldreich's generator is instantiated with the expander G and the predicate P, the algorithm A distinguishes (with high probability) between a uniform string and the output of the generator.

Proof. Let $G = ([n], [m], E)$ be the expander graph from Theorem 4.5, with right-degree ℓ. Let $\texttt{prg} : \{0,1\}^n \to \{0,1\}^m$ be Goldreich's generator, instantiated with the graph G and with the predicate P. For any fixed $x \in \{0,1\}^n$, we show that $\texttt{prg}(x)$ is the truth-table of an $\mathcal{AC}^0[\oplus]$ circuit $\{0,1\}^{\log(m)} \to \{0,1\}$ of depth d and size $2^{o\left(\log(m)^{1/2d}\right)}$. Assuming such a circuit indeed exists for any fixed x, the natural property algorithm from Theorem 3.9 rejects all m-bit strings in the image of $\texttt{prg}$, but accepts a random m-bit string with probability at least $1/2$.

Thus, let us fix x and construct a circuit $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ whose truth-table is $\texttt{prg}(x)$. Given input $i \in [m]$, the circuit g_x:

- Uses the circuit $C_G : \{0,1\}^{\log(m)} \to \{0,1\}^{\ell \cdot \log(n)}$ from Theorem 4.5 to compute $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$.
- Let $\Phi_x : \{0,1\}^{\log(n)} \to \{0,1\}$ be a depth-two formula of size n such that for any $i \in [n]$ it holds that $\Phi_x(i)$ is the i^{th} bit of x. The circuit g_x uses ℓ copies of Φ_x to map $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$ to $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$.
- Finally, the circuit maps $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$ to an output bit $P(x\restriction_{\Gamma_G(i)})$, using the $\mathcal{AC}^0[\oplus]$ circuit of depth d_P for P.

The total depth of the circuit for g_x is $d = d_G + d_P + 2$, and its size is at most

$$\text{poly}(n) + \ell \cdot n + 2^{(\log(m))^\epsilon} = 2^{o\left((\log(m))^{1/2d}\right)} ,$$

where we relied on the hypothesis that k is sufficiently large (to deduce that $n^{O(1)} = 2^{o\left(\log(n)^{k/2d}\right)}$) and on the hypothesis that ϵ is sufficiently small. ■

4.3 Hard predicates and the class of subexponential sized $\mathcal{AC}^0[\oplus]$ circuits

Our goal in this section is to prove that the class of $\mathcal{AC}^0[\oplus]$ circuits of small sub-exponential size (and even of polynomial size) contains predicates with high rational degree and high resilience. In order to explain the relevance of these complexity measures, we briefly survey related results from the recent work [AL18]; this might also help to put the main result of this section (Theorem 4.6) in perspective.

4.3.1 A brief review of attacks from [AL18]

We provide next a brief informal description of the attacks and parameters from [AL18]. For a fixed predicate $P\colon \{0,1\}^\ell \to \{0,1\}$, we let $f\colon \{0,1\}^n \to \{0,1\}^m$ be a *random ℓ-local function*, i.e., each of its output bits is computed by applying P to a random set of ℓ distinct input bits. We say that f is *s-pseudorandom* if for $m = n^s$ and n sufficiently large the resulting random function (asymptotically almost surely) fools all efficient tests. Note that, after the choice of f, both the hypergraph describing the input-output dependencies and the predicate P are known to the adversary.

We describe two attacks from [AL18]; a *linear* attack, which exploits low resilience of the predicate, and an *algebraic* attack, which exploits low rational degree of the predicate. (See [AL18] for a precise description of these.)

Theorem 4.7 (Linear Attack, Theorem 1.1 from [AL18]) *Let P be a predicate with resilience k and r-bit fixing degree e. Then, for any $s > 1$, if $k < 2s - 1$ or $r + e < s$, P is not s-pseudorandom against linear tests.*

Some comments are in order:

– In the case of resilience $k < 2s - 1$, the distinguishing probability of the linear test is $2^{-2\ell}$, where ℓ is the arity of P [AL18, Lemma 1]. Therefore, in the case where $\ell = \ell(n)$ is polylogarithmic, the advantage is negligible.

– In both cases (low resilience or low bit-fixing degree), the attacks explore properties that hold for a random hypergraph [AL18, Lemmas 1 and 2].[14]

– Theorem 1.1 from [AL18] also provides a converse to Theorem 4.7. Roughly speaking, they show that slightly better resilience and bit-fixing degree parameters imply security against linear tests. This result requires $n \geq \exp(\ell)$ (see [AL18, Theorem 4.1]), and therefore it is not meaningful if $\ell = \ell(n) \gg \log n$.

– The proof of the converse statement relies on the expansion properties of a random hypergraph and on the properties of the predicate. In particular, security against linear tests is guaranteed if the underlying hypergraph is fixed but has good expansion properties.

Theorem 4.8 (Algebraic Attack, Theorem 1.4 from [AL18]) *Let P be a predicate with rational degree e. If $e < s$ then P is not s-pseudorandom against algebraic attacks.*

We also mention some relevant aspects of this result:

– The attack described in [AL18, Section 5.2] runs in time $O(m^{2\ell})$, where ℓ is the arity of P. Therefore, the attack is not efficient in the output length m of f is $\ell = \ell(n)$ is super-constant.

[14] It is possible that some of these properties hold for every fixed expander graph with adequate parameters, since such graphs behave as random graphs in many aspects (e.g. expander mixing lemma and its consequences).

– The attack explores properties of the underlying random hypergraph. Moreover, the attack as described in [AL18] requires a stretch parameter m for f that is lower bounded by a function $\ell^{O(\ell)}$ (see [AL18, Lemma 5]). In particular, this is super-polynomial for $\ell = \ell(n) = \log n$.

– Theorem 4.8 admits a converse that guarantees security against algebraic attacks for predicates of rational degree $e > 8s + 1$ [AL18, Theorem 5.4].

According to [AL18], "*Our results show that any k-resilient predicate with rational degree e is s-pseudorandom against linear attacks and algebraic attacks as long as* $k \geq 5s$ *and* $e > 18s$. *Are there efficient attacks against such predicates? As a concrete example, one may try to break the s-pseudorandomness of the* $\mathsf{XOR\text{-}MAJ}_{a,b}$ *predicate* $(z_1 \oplus \ldots \oplus z_a) \oplus \mathsf{MAJ}(z_{a+1}, \ldots, z_{a+b})$ *with* $a \geq 5s$ *and* $b > 36s$."

We stress that, at least in the case of linear attacks, the security that is alluded to in the quote above is only guaranteed in the regime where the parameters are fixed with respect to n. In both cases (linear and algebraic attacks), the security follows from the parameters of the predicate and the expansion properties of the hypergraph encoding the input-output dependencies. In other words, security does not require a *random local function* as defined earlier in this section; a fixed hypergraph with good expansion properties is sufficient.

4.3.2 The XOR-APPROX-MAJ Predicate

In this section we consider a modification of the $\mathsf{XOR\text{-}MAJ}_{a,b}$ predicate, which we denote by XOR-APPROX-MAJ, where the majority function is replaced by an approximate majority. Towards defining the predicate, let us first recall the standard definition of an *approximate majority* predicate, and the fact that there exist $\mathcal{AC}^0$ circuits computing this predicate. For a string $x \in \{0,1\}^*$, let $|x|_1$ denote its Hamming weight. Then:

Definition 4.9 (Approximate Majority) *We say that a function* $h\colon \{0,1\}^b \to \{0,1\}$ *is an* ε-approximate majority *if* $h(x) = 0$ *for every input* x *such that* $|x|_1 \leq (1/2 - \varepsilon)b$, *and* $h(x) = 1$ *for every input* x *such that* $|x|_1 \geq (1/2 + \varepsilon)b$.

Theorem 4.10 (Approximate Majority in AC^0 ([Ajt90; Vio09])) *For every* $\varepsilon > 0$, *there exists a uniform family* $\{D_b\}_{b \geq 1}$ *of polynomial size depth-3 circuits that compute an* ε*-approximate majority over* b *input bits.*

We define the XOR-APPROX-MAJ predicate using the particular approximate majority function that can be computed by $\mathcal{AC}^0$ circuits as in Theorem 4.10.

Definition 4.11 (The XOR-APPROX-MAJ Predicate) *Let* $a, b \geq 1$ *and* $\varepsilon > 0$. *Let* $h\colon \{0,1\}^b \to \{0,1\}$ *be the* ε*-approximate majority function whose existence is asserted in Theorem 4.10, and let* $\mathsf{XOR\text{-}}\varepsilon\mathsf{\text{-}APPROX\text{-}MAJ}_{a,b}\colon \{0,1\}^{a+b} \to \{0,1\}$ *be the predicate given by* $(z_1 \oplus \ldots \oplus z_a) \oplus h(z_{a+1}, \ldots, z_{a+b})$.

The next proposition asserts that the XOR-APPROX-MAJ predicate behaves similarly to XOR-MAJ with respect to resilience and rational degree. In particular, it also avoids the linear and algebraic attacks from [AL18] described in Section 4.3.1.

Proposition 4.12 *(properties of the* XOR-APPROX-MAJ *predicate). Let $a, b \geq 1$, $\varepsilon > 0$, and let* XOR-ε-APPROX-MAJ$_{a,b}$: $\{0,1\}^{a+b} \to \{0,1\}$ *be the predicate from Definition 4.11. Then P has resilience $\geq a-1$ and rational degree $> (1/2-\varepsilon)b$.*

Proof. For the resilience lower bound, let $\gamma \in \{0,1\}^{a+b}$, $|\gamma|_1 \leq a-1$, and $h\colon \{0,1\}^b \to \{0,1\}$ be the corresponding ε-approximate majority function. Clearly, there exists $i \in \{1,\ldots,a\}$ such that $\gamma_i = 0$. Consequently,

$$\begin{aligned}
\mathbb{E}_{z\in\{0,1\}^{a+b}}[(-1)^{P(z)+\langle z,\gamma\rangle}] &= \mathbb{E}_z[(-1)^{(\oplus_{j=1}^{a} z_j)\oplus h(z_{a+1},\ldots,z_{a+b})\oplus\langle z,\gamma\rangle}] \\
\text{(using independence)} \quad &= E_z[(-1)^{z_i}]\cdot\mathbb{E}_z[(-1)^{(\oplus_{j\in[a]\setminus\{i\}} z_j)\oplus h(z_{a+1},\ldots,z_{a+b})\oplus\langle z,\gamma\rangle}] \\
&= 0,
\end{aligned}$$

where we have used the fact that $\gamma_i = 0$ to decompose the expectation into a product of expectations.

Turning to the rational degree lower bound, first we argue the claim for the ε-approximate majority function $h\colon \{0,1\}^b \to \{0,1\}$. Let $d = \lfloor(1/2-\varepsilon)b\rfloor$, and suppose that $\mathsf{rdeg}_{\mathbb{F}_2}(h) \leq d$. We will rely on the following lemma:

Lemma 4.13 *(see, e.g., [Car10; AL18]). A predicate $h\colon \{0,1\}^b \to \{0,1\}$ has rational degree $e \geq 0$ if and only if there exists a non-zero predicate $\tilde{h}\colon \{0,1\}^b \to \{0,1\}$ with $\mathsf{deg}_{\mathbb{F}_2}(\tilde{h}) \leq e$ such that one of the following conditions hold:*

(a) *For every $z \in \{0,1\}^b$, if $h(z) = 0$ then $\tilde{h}(z) = 0$; or*

(b) *For every $z \in \{0,1\}^b$, if $h(z) = 1$ then $\tilde{h}(z) = 0$.*

Assume that Case (a) from Lemma 4.13 holds (the proof of the other case is very similar), and let $\tilde{h}$ be the non-zero predicate from the lemma. Let $B(0^b, d) \subseteq \{0,1\}^b$ be the set of strings of Hamming distance at most d from 0^b, and note that by the definition of h it holds that $B(0^b,d) \subseteq h^{-1}(0)$. It follows that $\tilde{h}$ vanishes on $B(0^b,d)$ (i.e., for every $z \in B(0^b,d)$ it holds that $\tilde{h}(z) = 0$). However, this contradicts the following lemma:

Lemma 4.14 *(see [KS12]). Let $\tilde{h}(x_1,\ldots,x_b) \in \mathbb{F}_2[x_1,\ldots,x_b]$ be a non-zero polynomial of degree at most d, and let $S_{\tilde{h}} = \{x \in \{0,1\}^b \mid \tilde{h}(x) \neq 0\}$. Then, for every $y \in \{0,1\}^b$ it holds that $S_{\tilde{h}} \cap B(y,d) \neq \emptyset$.*

Finally, note that this rational degree lower bound also holds for the predicate P, since it is easy to see using Lemma 4.13 that the rational degree of a function cannot increase by taking a restriction. ∎

5 Conditional results in the setting of polynomial stretch

In Section 5.1 we present our results that are conditioned on the existence of expanders with a neighbor function that can be computed by a layer of parity gates, and in Section 5.2 we present our results that are conditioned on the existence of expanders with a neighbor function that can be computed in $\mathcal{NC}^0$.

5.1 Expanders with affine neighbor functions

In this section we present results that are conditioned on the existence of expanders in which the neighbor functions are affine functions (i.e., can be computed by a layer of parity gates). In Section 5.1.1 we formally present the assumption that suitable expanders exist. In Section 5.1.2 we prove the existence of natural properties that are useful against DNF-XOR circuits and related functions. In Section 5.1.3 we show that if the aforementioned expanders exist, then an improvement in the parameters of the foregoing natural properties would allow to break the generator with *any* predicate for the setting of a large polynomial stretch.

5.1.1 The affine expander assumption

Towards presenting the assumption regarding the existence of expanders with affine neighbor functions, let us first formally define affine functions:

Definition 5.1 *(multi-output affine functions). We say that a function $f : \{0,1\}^r \to \{0,1\}^{r'}$ is* affine *if for each output bit y_i of $f(x)$ there exists a set $S_i \subseteq [r]$ and $b_i \in \{0,1\}$ such that $y_i = b_i + \sum_{j \in S_i} x_j \pmod 2$.*

To motivate our assumption, we recall two well-known *explicit constructions* of unbalanced lossless expanders, which were given by Ta-Shma, Umans, and Zuckerman [TUZ07], and later on by Guruswami, Umans, and Vadhan [GUV09]. These two constructions are very different (the first is combinatorial whereas the second is algebraic), and yet in both constructions *the neighbor function of the expander is affine*; that is, for each $i \in [\ell]$, the neighbor function $\Gamma_i : [m] \times [n]$ is affine.

Being a bit more specific, let us first recall the construction in [TUZ07]. When their construction is instantiated with Trevisan's extractor [Tre01], which is a "reconstructive extractor" (as explained in [TUZ07, Sec. 4.1]), it yields an expander $[n] \times [n^k]$ (for $k = O(1)$) with right-degree $\ell = \text{poly}\log(n)$, sub-optimal expansion (i.e., only sets of size $n^{o(1)}$ are guaranteed to expand), and an affine neighbor function: [15]

[15] To avoid confusion, let us note in advance that in the literature concerning lossless expanders and condensers (and in particular in [TUZ07; GUV09]) it is typical to think of expanders in which the left-hand side $[n]$ is larger, has left-degree ℓ, and expands into the (smaller) right-hand side $[m]$. In our case, corresponding to the literature concerning expander-based cryptography, the larger side is the right one $[m]$, and it expands into the smaller side $[n]$.

Theorem 5.2 *([TUZ07, Thm. 1.7]). Let $n \in \mathbb{N}$, let $m = \mathrm{poly}(n)$, and let $\ell = \mathrm{poly}\log(n)$. Then, there exists an $(n^{1/(\log(n))^{.01}}, 0.99)$-expander $G = ([n],[m],E)$ of right-degree ℓ such that for each $i \in [\ell]$, the neighbor function $\Gamma_i : [m] \to [n]$ is affine.*

Proof Sketch. For any "reconstructive extractor" (as defined in [TUZ07, Def. 3.3]), the advice function of the extractor yields a condenser (see [TUZ07, Thm. 3.6]), which yields an expander (see [TUZ07, Thm. 8.1]).

We use the specific reconstructive extractor by Trevisan [Tre01], based on [NW94], as described in [TUZ07, Sec. 4.1, "The advice function"]. Let us describe the advice function of this extractor. The main input $v \in [m]$ to the function represents a truth-table of a function $f_v : \{0,1\}^{\log\log(m)} \to \{0,1\}$, and the second input $i \in [\ell]$ represents an index $\alpha \in [\log\log(m)]$ and values $\beta \in \{0,1\}^{t-r}$, where $t = O(\log\log(m))$ and $r = \log\log(m) + O(1)$. (Readers familiar with the Nisan-Wigderson construction will recognize α as the index of an input bit to f that comes from a hybrid argument, and β as values for $t-r$ bits in the t-long seed of the generator in locations outside a specific subset associated with α.) The advice function encodes its main input $v \in [m]$ using a linear error-correcting code (any good linear good suffices), and outputs the projections of the encoding of v to a subset $S = S_{\alpha,\beta} \subseteq [O(\log(m))]$ of coordinates that is determined by α and β. Therefore, for any fixed $i = (\alpha, \beta)$, the function $\Gamma_{i=(\alpha,\beta)}$ outputs (fixed) projections of a linear encoding of the input v. ■

Turning to [GUV09], it is well-known that when their construction is instantiated over a field of characteristic two, with a specific a choice of parameters, it yields an expander that is affine (see, e.g., [Che05, Cor. 2.23]). In particular, it yields an affine expander with the following parameters:

Theorem 5.3 *([GUV09, Thm. 1.3]). Let $n \in \mathbb{N}$, let $m = n^2$, and let $\ell = \mathrm{poly}\log(n)$. Then, there exists an $(n^{.99}, 0.99)$-expander $G = ([n],[m],E)$ of right-degree ℓ such that for each $i \in [\ell]$, the neighbor function $\Gamma_i : [m] \to [n]$ is affine.*

The degree of the expander in Theorems 5.2 and 5.3 is polylogarithmic, whereas the degree of a random construction is constant $\ell = O(k)$ (see Theorem 3.2). Indeed, it is an open problem to improve the degree in these constructions to match the one of a random construction. The following assumption asserts that expanders as in Theorems 5.2 and 5.3 exist, but with small right-degree $\ell = O(k)$:

Assumption 5.4 *(affine expander assumption). There exists a constant $\beta > 3$ such that for every constant $k \in \mathbb{N}$, sufficiently large $n \in \mathbb{N}$, and $m = n^k$, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n],[m],E)$ with right-degree $\ell = \beta \cdot k$ that satisfies the following: The function $\Gamma_G : \{0,1\}^{\log(m)} \to \{0,1\}^{\ell \cdot \log(n)}$ that gets as input $i \in [m]$ and outputs the ℓ neighbors of i in G is affine.*

Under Assumption 5.4, there exists an expander G such that when Goldreich's generator is instantiated with G and with any predicate P, the output of the generator on any seed is the truth-table of a depth-four circuit with constant top fan-in and a bottom layer of parity gates. In more detail:

Claim 5.5. *If Assumption 5.4 holds with parameter $\beta > 3$, then for every $k \in \mathbb{N}$, every sufficiently large $n \in \mathbb{N}$, and $m = n^k$, there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m], E)$ with right-degree $\ell = \beta \cdot k$ such that for every predicate $P : \{0,1\}^\ell \to \{0,1\}$ that can be computed by a CNF of size $2^{\delta \cdot \ell}$ the following holds: When Goldreich's generator is instantiated with G and P, for every fixed $x \in \{0,1\}^n$, the output of the generator on x is the truth-table of a function $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ such that :*

1. *The function g_x can be computed by an* AND-DNF-XOR *circuit of top fan-in $2^{\delta \cdot \beta \cdot k}$ and size at most $O\left(2^{(1/k) \cdot \log(m)}\right)$ containing at most $2 \cdot \beta \cdot \log m$ parity gates.*

2. *The function g_x can be $(1/2 + 1/2^{\delta \cdot \beta \cdot k})$-approximated by a* DNF-XOR *circuit of size at most $O\left(2^{(1/k) \cdot \log(m)}\right)$ containing at most $2 \cdot \beta \cdot \log m$ parity gates.*

Proof. For $k \in \mathbb{N}$ and sufficiently large $n \in \mathbb{N}$, let $G = ([n], [m = n^k], E)$ be the expander from Assumption 5.4, let $P : \{0,1\}^\ell \to \{0,1\}$ be any predicate that can be computed by a CNF of size $2^{\delta \cdot \ell}$, and let $\texttt{prg} : \{0,1\}^n \to \{0,1\}^m$ be Goldreich's generator, instantiated with the graph G and with the predicate P.

For any fixed $x \in \{0,1\}^n$, we construct a circuit $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ whose truth-table is $\texttt{prg}(x)$. Given input $i \in [m]$, the circuit g_x:

- Uses a single layer of at most $2 \cdot \ell \cdot \log n$ parity gates to compute the string $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$ and its negation.

- Let $\Phi_x : \{0,1\}^{\log(n)} \to \{0,1\}$ be a DNF of size at most n such that for any $i \in [n]$ it holds that $\Phi_x(i)$ is the i^{th} bit of x; we view this DNF as a De Morgan formula over input literals $z_1, \ldots, z_{\log(n)}, \neg z_1, \ldots, \neg z_{\log(n)}$. The circuit g_x uses ℓ copies of Φ_x to map $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$ to $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$, and similarly uses ℓ copies of a DNF of size at most n for $\neg\Phi_x$ to compute the negations of $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$.

- Finally, the circuit g_x computes the value of the predicate P on the input string $x\restriction_{\Gamma_G(i)}$ by using a CNF of size $2^{\delta \cdot \ell}$ for P over the corresponding $\ell = \beta \cdot k$ input variables.

After collapsing two layers (of the top CNF with the DNFs below it), we obtain an exact AND-OR-AND-XOR circuit of size $O(2^{(1/k) \cdot \log(m)})$ that computes $g_x \colon \{0,1\}^{\log(m)} \to \{0,1\}$, where negation gates appear next to input variables only. Moreover, the number of parity gates is at most $2 \cdot \ell \cdot \log n = 2 \cdot \ell \cdot \log(m)/k = 2 \cdot \beta \cdot \log(m)$, and the top fan-in of the circuit is $2^{\delta \cdot \ell} = 2^{\delta \cdot \beta \cdot k}$.

The claimed DNF-XOR circuit that $(1/2 + 1/2^{O(k)})$-approximates g_x follows from the Discriminator Lemma of [HMP+93] using that the top gate of g_x computes a symmetric gate over at most $2^\ell = 2^{\beta \cdot k}$ input wires. ■

5.1.2 Natural properties against DNF-XOR circuits

Our goal in this section is to show natural properties that are useful against DNF-XOR circuits of exponential size, against conjunctions of a constant number of DNF-XOR circuits of exponential size, and against functions that can be approximated by exponential-sized parity decision trees (which are related to DNF-XOR circuits, but weaker).

DNF-XOR circuits of exponential size. The first natural property that we show is useful against DNF-XOR circuits over ℓ_m bits of exponential size $2^{(1/2-o(1))\cdot\ell_m}$:

Proposition 5.6 *(natural property against exponential size DNF-XOR circuits). There exists a natural property against the class of functions $f : \{0,1\}^{\ell_m} \to \{0,1\}$ that can be computed by DNF-XOR circuits of size $2^{(1/2-o(1))\cdot\ell_m}$.*

Proof. We rely on the following result from [ABGKR14], which shows that a DNF-XOR circuit of bounded size has a Fourier coefficient of relatively large magnitude.

Proposition 5.7 ([ABGKR14]) *Let $g : \{0,1\}^{\ell_m} \to \{0,1\}$ be computed by a DNF-XOR circuit of top fan-in s. Then there exists $S \subseteq [\ell_m]$ such that $|\widehat{g}(S)| \geq 1/(2s+1)$.*

In particular, if $s = o\left(2^{\ell/2-\frac{1}{2}\cdot\log(\ell_m)}\right)$, then some Fourier coefficient of g is of magnitude $\omega\left(\sqrt{\ell_m}\cdot 2^{-\ell_m/2}\right)$. On the other hand, if $h : \{0,1\}^{\ell_m} \to \{0,1\}$ is a random Boolean function, it follows from a standard concentration bound that for any fixed $S \subseteq [\ell_m]$, with probability at least $1 - 2\cdot\exp(-10\cdot\ell_m)$ it holds that $|\widehat{h}(S)| \leq 10\cdot\sqrt{\ell_m}\cdot 2^{-\ell_m/2}$. By taking a union bound over all such $S \subseteq [\ell_m]$, we get:

Proposition 5.8 (see, e.g. [OS08]) *With probability $1 - 2^{-\ell_m}$ over choice of a random Boolean function $h : \{0,1\}^{\ell_m} \to \{0,1\}$, for every $S \subseteq [\ell_m]$ it holds that $|\widehat{h}(S)| \leq O(\sqrt{\ell_m}\cdot 2^{-\ell_m/2})$.*

The natural property claimed in the statement of the lemma can therefore be defined as follows: Given the truth table of a Boolean function f, exactly compute each Fourier coefficient of f, and accept the input function if and only if every coefficient has magnitude bounded by $O(\sqrt{\ell_m}\cdot 2^{-\ell_m/2})$. This computation can be done in time $2^{O(\ell_m)}$, which is polynomial in the size of the truth table. ■

Extending Proposition 5.6, we can get a property that is useful against larger circuits of size $2^{(1-o(1))\cdot\ell_m}$, at the cost of having a slightly super-polynomial running time. This natural property is based on a result of Cohen and Shinkar [CS16]:

Proposition 5.9 *(natural property against strongly exponential size DNF-XOR circuits). There exists a natural property computed in time $T(2^{\ell_m}) = 2^{O(\ell_m\cdot\log(\ell_m))}$ against the class of functions $f : \{0,1\}^{\ell_m} \to \{0,1\}$ that can be computed by DNF-XOR circuits of strongly exponential size $2^{(1-o(1))\cdot\ell_m}$.*

Proof. An *affine disperser* for dimension k is a function $h : \{0,1\}^{\ell_m} \to \{0,1\}$ with the following property: For every affine subspace $U \subseteq \{0,1\}^{\ell_m}$ of dimension k, the restriction of h to U is non-constant. A standard argument shows that a random function is, with probability more than $1/2$, an affine disperser for dimension $k = \log(\ell_m) + \log\log(\ell_m) + C$, where C is a large enough constant.

On the other hand, Cohen and Shinkar [CS16] established the following result. For a function g, let $\mathsf{DNF\text{-}XOR}(g)$ be the number of gates in the smallest DNF-XOR circuit for g. Then, if g is an affine disperser for dimension k, $\max(\mathsf{DNF\text{-}XOR}(g), \mathsf{DNF\text{-}XOR}(1-g)) = \Omega(2^{\ell_m - k})$. An immediate consequence is that for $k^\star = \log(\ell_m) + \log\log(\ell_m) + C$, by checking if an input truth-table or its negation is a $k^\star$ affine disperser it is possible to distinguish a random function from a function of DNF-XOR circuit complexity at most $2^{\ell}_{m}/(\ell_m)^2$. ∎

Conjunctions of DNF-XOR circuits of exponential size. We now turn to show a natural property that is useful against conjunctions of $c = O(1)$ DNF-XOR circuits over ℓ_m input bits of size $2^{(\epsilon/2)\cdot\ell_m}$, for any $\epsilon < 1/c$. Specifically:

Proposition 5.10 *(a natural property against conjunctions of c DNF-XOR's of size $2^{(1/2c - \Omega(1))\cdot\ell_m}$). For any $c \in \mathbb{N}$ and $\epsilon < 1/c$, there exists a natural property against the class of functions $f : \{0,1\}^{\ell_m} \to \{0,1\}$ that can be computed by a conjunction of c DNF-XOR circuits of size $2^{(\epsilon/2)\cdot\ell_m}$.*

Proof. We prove Proposition 5.10 using Proposition 5.6. To do so, note that any conjunction of c DNF-XOR circuits of size s can be computed by a single DNF-XOR circuit of size s^c (using the distributivity of the top AND gate with the OR gates below it). In particular, if $s = 2^{(\epsilon/2)\cdot\ell_m}$, where $\epsilon < 1/c$, then the conjunction can by computed by a single DNF-XOR circuit of size $s^c = 2^{(\epsilon\cdot c/2)\cdot\ell_m}$. Since $\epsilon \cdot c/2 < 1/2$, the natural property from Proposition 5.6 works against such a function. ∎

The natural property from Proposition 5.10 can be extended to work against conjunctions of c DNF-XOR circuits of size $2^{\epsilon\cdot\ell_m}$ (rather than $2^{(\epsilon/2)\cdot\ell_m}$), for any $\epsilon < 1/c$, at the cost of a slightly super-polynomial running time. Specifically:

Proposition 5.11 *(an almost-natural property against conjunctions of c DNF-XOR's of size $2^{(1/c-\Omega(1))\cdot\ell_m}$). For any $c \geq 2$ and $\epsilon < 1/c$, there exists a natural property that runs in time $T(2^{\ell_m}) = 2^{O(\ell_m\cdot\log(\ell_m))}$ for the class of functions $\{0,1\}^{\ell_m} \to \{0,1\}$ that can be computed by a conjunction of c DNF-XOR circuits of size at most $2^{\epsilon\cdot\ell_m}$.*

Proof. To present this property we first need the following structural lemma for conjunctions of DNF-XOR circuits. (The proof of the lemma is reminiscent of the proof of [CS16, Second Item of Thm. 1.1].)

Lemma 5.12 *(a structural property of conjunctions of DNF-XOR circuits). Let $g_x : \{0,1\}^{\ell_m} \to \{0,1\}$ be computed by a conjunction of $c \geq 1$ DNF-XOR circuits, where each DNF-XOR circuit is of size at most $s(\ell_m) = 2^{\epsilon\cdot\ell_m}$.* [16] *Then, for every $k \in \mathbb{N}$, at least one of the following holds:*

[16]Here we consider the size of a DNF-XOR circuit to be its top fan-in (the number of affine subspaces such that the DNF-XOR circuit is a union of the subspaces).

1. *The function g_x is constant on a subspace of dimension $c \cdot k$.*

2. *The acceptance probability of g_x is at most $2^{\ell_m \cdot (\epsilon - 1/c) + k}$.*

Proof. Let $g_x(z) = \wedge_{i \in [c]} h_i(z)$, where each h_i is a DNF-XOR circuit of size at most $s(\ell_m)$. For each $i \in [c]$, we view h_i as the union of at most $s(\ell_m)$ affine subspaces, and denote by $\dim(h_i)$ the maximal dimension of an affine subspace in h_i; that is, if $h_i(n) = \cup_{j \in [s(n)]} H_j$, where each $H_j \subseteq \{0,1\}^n$ is an affine subspace, then $\dim(h_i) = \max_{j \in [s(n)]} \dim(H_j)$. We consider two separate cases:

Case 1: $\min_{i \in [c]} \{\dim(h_i)\} > \ell_m - \ell_m/c + k$. In this case it holds that g_x is constant on a subspace of dimension $c \cdot k$. Specifically, for each $i \in [c]$ let $H^{(i)}$ be an affine subspace in h_i such that $\dim(H^{(i)}) > \ell_m - \ell_m/c + k$, and let $H = \bigcap_{i \in [c]} H^{(i)}$. Then, the co-dimension of H is less than $c \cdot (\ell_m/c - k) = \ell_m - c \cdot k$, and for every $z \in H$ it holds that $g_x(z) = 1$.

Case 2: $\min_{i \in [c]} \{\dim(h_i)\} \leq \ell_m - \ell_m/c + k$. In this case it holds that g_x has acceptance probability at most. To see this, let $i \in [c]$ be such that $\dim(h_i) \leq \ell_m - \ell_m/c + k$, and note that the acceptance probability of g_x is upper-bounded by the acceptance probability of h_i. However, since h_i is the union of at most $s(\ell_m)$ subspaces of dimension $\ell_m - \ell_m/c + k$, the number of inputs accepted by h_i is at most

$$s(\ell_m) \cdot 2^{\ell_m - \ell_m/c + k} = 2^{\ell_m \cdot (1 - 1/c + \epsilon) + k} . \qquad \square$$

Now, let $k = \log(\ell_m)$. When given the truth-table of a function $f : \{0,1\}^n \to \{0,1\}$, the algorithm iterates over all affine subspaces in $\{0,1\}^{\ell_m}$ of dimension $c \cdot k$, and rejects if f is the constant one function on any such subspace. Also, the algorithm rejects if the acceptance probability of f is less than $1/3$.

To see that the algorithm rejects any conjunction of c DNF-XOR circuits of size at most $2^{\epsilon \cdot \ell_m}$, note that by Lemma 5.12, any such function is either constant on a subspace of dimension $c \cdot k$, or has acceptance probability at most $2^{\ell_m \cdot (\epsilon - 1/c) + k} = 2^{-\Omega(\ell_m)}$; in both cases, the algorithm rejects. To see that the algorithm accepts a random function, note a random function is very likely (with probability $1 - \exp(-\ell_m)$) to have acceptance probability more than $1/3$. Also, for any fixed subspace of dimension $c \cdot k$, the probability that a random function is the constant one function on the subspace is $2^{-2^{c \cdot k}} \leq 2^{-\ell_m^2}$. Since there are at most $2^{(k+1) \cdot \ell_m} = 2^{O(\ell_m \cdot \log(\ell_m))}$ affine subspaces of dimension k, by a union-bound, with probability more than $1/2$, a random function is not constant on any subspace of such dimension k.

Finally, the algorithm is deterministic, and its runtime is polynomial in the number of affine subspaces of dimension $k = \log(\ell_m)$, and is thus at most $2^{O(\ell_m \cdot \log(\ell_m))}$. ■

Functions that are approximated by parity decision trees Finally, consider natural properties that are useful against functions that can be *approximated*, in the average-case sense, by simple functions such as DNF-XOR circuits. Specifically, we consider approximation by *parity decision trees* (PDTs), a model that is known to be related to but strictly weaker than DNF-XOR circuits. Recall that a parity decision tree is simply a decision tree where the nodes can query the value of any parity function of the input variables. We measure the size of such a decision tree by its number of nodes. (See [CS16] for a comparison of PDTs and DNF-XOR circuits.) Then:

Proposition 5.13 *(average-case natural property against PDTs). There exists a function $\gamma(\ell_m) \to 0$ and a natural property computed in time $T(2^{\ell_m}) = 2^{O(\ell_m \cdot \log(\ell_m))}$ against the class of functions $f : \{0,1\}^{\ell}_m \to \{0,1\}$ that can be $(1/2 + \gamma(\ell_m))$-approximated by a parity decision tree of size $2^{(1-o(1))\cdot \ell_m}$.*

Proof. A (k, ϵ)-*affine extractor* is a function $h : \{0,1\}^{\ell_m} \to \{0,1\}$ with the following property: For every affine subspace $U \subseteq \{0,1\}^{\ell_m}$ of dimension k, the restriction of h to U has bias at most ϵ. A standard argument shows that a random function is a (k, ϵ)-affine extractor with probability at least $1/2$, where $k = \log(\ell_m/\epsilon^2) + \log\log(\ell_m/\epsilon^2) + C$, and C is a large enough constant.

Let $\gamma(\ell_m) = (\ell_m)^{-1/4}$, and let $\mathcal{A}_{\ell_m}$ be the class of ℓ_m-bit Boolean functions that compute a (k, γ)-affine extractor, where $k = 2 \cdot \log(\ell_m)$. First, $\mathcal{A}_{\ell_m}$ is a dense property for every large enough ℓ_m, by the result mentioned above. Secondly, arguing as in the proof of Proposition5.9, $\mathcal{A}_{\ell_m}$ can be decided in time $2^{O(\ell_m \cdot \log(\ell_m))}$. Finally, we claim that $\mathcal{A}_{\ell_m}$ does not accept functions that can be approximated by PDTs. We use the following result from [CS16].

Theorem 5.14 ([CS16]) *Let $f : \{0,1\}^{\ell_m} \to \{0,1\}$ be a (k, ϵ)-affine extractor. For any function $g : \{0,1\}^{\ell_m} \to \{0,1\}$ of parity decision tree complexity $\leq \epsilon \cdot 2^{\ell_m - k}$, we have $\Pr_x[f(x) = g(x)] < 1/2 + 4\epsilon$.*

It follows from Theorem 5.14 and from our choice of the parameters k and γ that any Boolean function that can be $(1/2 + \gamma(\ell_m))$-approximated by a parity decision tree g of size $\leq 2^{\ell_m}/\ell_m^{10}$ cannot be in $\mathcal{A}_{\ell_m}$. This completes the proof. ∎

5.1.3 Natural properties that would allow to break the generator with any predicate, under Assumption 5.4

In this section we consider two "mild strengthenings" of natural properties that were presented in Section 5.1.2. We then show that if Assumption 5.4 holds and any of the two "mildly stronger" natural properties exist, then Goldreich's generator is insecure for the setting of large polynomial stretch with *any* predicate P.

First setting: Worst-case natural property. Recall that in Proposition 5.11 we showed an almost-natural property against conjunctions of $c = O(1)$ DNF-XOR circuits, each of which is of size at most $2^{\epsilon \cdot \ell_m}$, where $\epsilon < 1/c$. We first consider the setting in which there exists a natural property against conjunctions of $c = O(1)$ DNF-XOR circuits, each of which is of size at most $2^{\epsilon \cdot \ell_m}$, where $\epsilon > \Omega(1/\log(c)))$.

Theorem 5.15 *(breaking Goldreich's generator for any predicate, under Assumption 5.4 and a natural property mildly stronger than Proposition 5.10). Suppose that Assumption 5.4 holds with parameter $\beta > 3$. Also assume that for some $k \in \mathbb{N}$, there exists a natural property against the class of Boolean functions on ℓ_m input bits that are computed by* AND-DNF-XOR *circuits of top fan-in $2^{\beta \cdot k}$ and size at most $O\left(2^{(1/k)\cdot \log(m)}\right)$ containing at most $2 \cdot \beta \cdot \log m$ parity gates.*

Then, there exists a polynomial-time algorithm A that satisfies the following. For every $n \in \mathbb{N}$ there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m = n^k], E)$ with right-degree $\ell = \beta \cdot k$ such that for every predicate $P : \{0,1\}^\ell \to \{0,1\}$, the algorithm A distinguishes the outputs of Goldreich's generator from random strings (with constant gap $\Omega(1)$).

Proof. Let k be as in the hypothesis, let n be sufficiently large, let G be the expander from Assumption 5.4 with right-degree $\ell = \beta \cdot k$, and let $P : \{0,1\}^\ell \to \{0,1\}$ be an arbitrary predicate.

We instantiate Goldreich's generator with the expander G and predicate P. By Claim 5.5, every output string g_x of the generator can be computed by an AND-DNF-XOR circuit of top fan-in $2^{\beta \cdot k}$ and size at most $O\left(2^{(1/k)\cdot \log(m)}\right)$ containing at most $2 \cdot \beta \cdot \log m$ parity gates. Thus, the natural property from our hypothesis distinguishes the output of the generator from a random string with probability at least $1/2$. ■

Second setting: Average-case natural property. Recall that in Proposition 5.13 we showed a natural property (based on [CS16]) that is useful against functions that can be *approximated* by parity decision trees. We now consider the existence of natural properties with similar parameters, but that are useful against the class of functions approximated by DNF-XOR circuits, rather than by PDTs. [17]

Theorem 5.16 *(breaking Goldreich's generator for any predicate, under Assumptions 5.4 and an average-case natural property for* DNF-XOR *circuits). Suppose that Assumption 5.4 holds with parameter $\beta > 3$. Also assume that for some function $\gamma(\ell_m) \to 0$ and a fixed $\epsilon > 0$, there exists a natural property computable in quasi-polynomial time against the class of Boolean functions on ℓ_m input bits that can be $(1/2 + \gamma(\ell_m))$-approximated by* DNF-XOR *circuits of size $2^{\epsilon \cdot \ell_m}$ containing at most $2 \cdot \beta \cdot \ell_m$ parity gates at the bottom layer.*

Then, for every $k > 1/\epsilon$ there exists a quasi-polynomial-time algorithm A that satisfies the following. For a sufficiently large $n \in \mathbb{N}$ there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m = n^k], E)$ with right-degree $\ell = \beta \cdot k$ such that for every predicate $P : \{0,1\}^\ell \to \{0,1\}$, the algorithm A distinguishes the outputs of Goldreich's generator from random strings (with constant gap $\Omega(1)$).

Proof. Let $k > 1/\epsilon$, let n be sufficiently large, let G be the expander from Assumption 5.4 with right-degree $\ell = \beta \cdot k$, and let $P : \{0,1\}^\ell \to \{0,1\}$ be an arbitrary predicate.

[17] To our knowledge, the same natural property behind Proposition 5.13 might also be useful against functions that are approximated by DNF-XOR circuits; we refer to [CS16] for related open problems.

We instantiate Goldreich's generator with the expander G and predicate P. By Claim 5.5, every output string g_x of the generator can be $(1/2+\Omega_k(1))$-approximated by a DNF-XOR circuit over $\ell_m = \log(m)$ input variables and of size $O(2^{(1/k)\cdot\ell_m}) \le 2^{\epsilon\cdot\ell_m}$. In addition, the number of parity gates in g_x is at most $2\cdot\beta\cdot\ell_m$, and we can assume that $\gamma(\ell_m)$ is sufficiently small (since n is sufficiently large). Thus, the natural property from our hypothesis distinguishes the output of the generator from a random string with probability at least $1/2$. ∎

5.2 Expanders with local neighbor functions

In this section we present results that are conditioned on the existence of expanders in which the neighbor functions are computable in $\mathcal{NC}^0$. In Section 5.2.1 we formally present the assumption that suitable expanders exist. Conditioned on this assumption, in Section 5.2.2 we prove the existence of natural properties that allow to break the generator with a limited class of predicates (i.e., predicates computable by DNFs of sufficiently small exponential size). In Section 5.2.3 we show that a mild improvement in the parameters of known natural properties allows to break the generator with *any* predicate for the setting of a large polynomial stretch.

5.2.1 The local expander assumption

Towards presenting the assumption regarding the existence of expanders with $\mathcal{NC}^0$ neighbor functions, let us first formally define local functions:

Definition 5.17 *(multi-output local functions). We say that a function $f : \{0,1\}^r \to \{0,1\}^{r'}$ is t-*`local` *if each output bit of f depends on at most t input bits.*

To motivate our assumption, we recall the recent construction of expanders with local neighbor functions, by Viola and Wigderson [VW17] (building on the work of Capalbo *et al.* [CRVW02], which used the Zig-Zag product).

Theorem 5.18 ([VW17, Thm. 4]) *There is $\ell \ge 1$ such that for every $n \ge 1$, the following holds. There exists a* balanced *$(n^{0.99}, 0.99)$-expander $G = ([n],[n],E)$ of right-degree ℓ such that for each $i \in [\ell]$, the neighbor function $\Gamma_i : [n] \to [n]$ can be realized by an $\mathcal{NC}^0$ circuit (i.e., each output bit of Γ_i depends on a constant number of input bits).*

The construction in Theorem 5.18 works for *balanced* expanders, whereas we are interested in highly imbalanced expanders (i.e., on expanders with $m = \text{poly}(n)$ vertices on the right side). Our main assumption in this section is that the construction of [VW17] can be extended to the imbalanced case:

Assumption 5.19 *(local expander assumption). There is a function $t : \mathbb{N} \to \mathbb{N}$ and a constant $\beta > 3$ such that for every $k \in \mathbb{N}$ the following holds. For every $n \in \mathbb{N}$ and $m = n^k$ there exists an $(n^{0.99}, 0.99)$-expander $G = ([n],[m],E)$ with right-degree $\ell = \beta\cdot k$ that satisfies the following: The function $\Gamma_G : \{0,1\}^{\log(m)} \to \{0,1\}^{\ell\cdot\log(n)}$ that gets as input $i \in [m]$ and outputs the ℓ neighbors of i in G is $t(k)$-local.*

Assumption 5.19 makes no requirements on locality t of the function Γ_G, except that it is only a function of k (and is thus constant, independent of $\log(m(n))$). The requirements of our attack satisfy a trade-off depending on the actual function t. It is consistent with our knowledge that even $t(k) = 1$ might suffice in Assumption 5.19.

Under Assumption 5.19, there exists an expander G such that when Goldreich's generator is instantiated with G and with any predicate P, the output of the generator on any seed is the truth-table of a depth-four circuit with constant top fan-in and constant bottom fan-in. In more detail:

Claim 5.20. *If Assumption 5.19 holds with parameters $t : \mathbb{N} \to \mathbb{N}$ and $\beta > 3$, then for every $k \in \mathbb{N}$, and every $n \in \mathbb{N}$ there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m = n^k], E)$ with right-degree $\ell = \beta \cdot k$ such that for every predicate $P : \{0,1\}^\ell \to \{0,1\}$ that can be computed by a depth-two circuit of size $2^{\delta \cdot \ell}$ the following holds: When Goldreich's generator is instantiated with expander G and predicate P, for every fixed $x \in \{0,1\}^n$, the output of the generator on x is the truth-table of an $\mathcal{AC}^0$ circuit $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ of depth four, with top fan-in $2^{\delta \cdot \ell} = 2^{\beta \cdot \delta \cdot k}$, bottom fan-in $t = t(k)$, and size $O\left(2^{(1/k) \cdot \log(m)}\right)$.*

Proof. For $k \in \mathbb{N}$ and $n \in \mathbb{N}$, let $G = ([n], [m = n^k], E)$ be the expander from Assumption 5.19, let $P : \{0,1\}^\ell \to \{0,1\}$ be any predicate that can be computed by a DNF of size $2^{\delta \cdot \ell}$, and let $\mathtt{prg} : \{0,1\}^n \to \{0,1\}^m$ be Goldreich's generator, instantiated with the graph G and with the predicate P. (We assumed that P can be computed by a DNF of size $2^{\delta \cdot \ell}$ only for simplicity; the proof of the case where P can be computed by a CNF of such size is very similar.)

For any fixed $x \in \{0,1\}^n$, we construct a circuit $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ whose truth-table is $\mathtt{prg}(x)$. Given input $i \in [m]$, the circuit g_x:

- Uses $\ell \cdot \log(n)$ DNFs over t bits (each of size $\le 2^t$) to compute $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$. Additionally, g_x uses the same number of DNFs to compute the negation of each output bit.
- Let $\Phi_x : \{0,1\}^{\log(n)} \to \{0,1\}$ be a CNF of size at most n such that for any $i \in [n]$ it holds that $\Phi_x(i)$ is the i^{th} bit of x; we view this CNF as a De Morgan formula over input literals $z_1, \ldots, z_{\log(n)}, \neg z_1, \ldots, \neg z_{\log(n)}$. The circuit g_x uses ℓ copies of Φ_x to map $\Gamma_G(i) \in \{0,1\}^{\ell \cdot \log(n)}$ to $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$, and similarly uses ℓ copies of a CNF of size at most n for $\neg\Phi_x$ to compute the negations of $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$.
- Finally, the circuit g_x maps $x\restriction_{\Gamma_G(i)} \in \{0,1\}^\ell$ to an output bit $P(x\restriction_{\Gamma_G(i)})$, using a DNF of size $2^{\delta \cdot \ell}$. This DNF is viewed as a de Morgan formula over input literals $y_1, \ldots, y_\ell, \neg y_1, \ldots, \neg y_\ell$.

The total depth of the circuit for g_x is four (after collapsing two layers), its bottom fan-in is t, and its top fan-in (which is the size of the DNF that computes P) is $c \le 2^{\delta \cdot \ell} = 2^{\delta \cdot \beta \cdot k}$. Finally, the size of the circuit for g_x is less than

$$2 \cdot \ell \cdot \log(n) \cdot 2^t + 2 \cdot \ell \cdot n + 2^\ell = O(2^{(1/k) \cdot \log(m)}),$$

for a constant k and as a function of n. ■

5.2.2 Natural properties that allow to break the generator with a limited class of predicates, under Assumption 5.19

Our goal is to show natural properties for a function $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ as in Claim 5.20, when $\delta > 0$ is sufficiently small (recall that we assume that the predicate $P : \{0,1\}^{\ell} \to \{0,1\}$ is computable by a depth-two circuit of size $2^{\delta \cdot \ell}$). Let us recall the definition of the *average sensitivity* of a function:

Definition 5.21 *(average sensitivity). We say that two strings $x, y \in \{0,1\}^r$ are* neighbors *if x and y disagree on a single bit. The* average sensitivity *of a function $f : \{0,1\}^r \to \{0,1\}$ is defined as the average, over $x \in \{0,1\}^r$, of the number of neighbors y of x such that $f(x) \neq f(y)$.*

To show the natural property, let g_x be a circuit of depth four of exponential size (i.e., size $S \leq 2^{\epsilon \cdot \ell_m}$) with constant top fan-in and constant bottom fan-in. We show that each g_x of such form simplifies to a depth-two circuit of small width (say, width 10) under a random restriction that keeps significantly more variables alive in expectation (say, 1000); the proof of this result is a standard application of Håstad's switching lemma. Thus, under the restriction, the circuit g_x has average sensitivity that is much smaller than the expected average sensitivity of a random function.

For the first part of the proof, let us recall Håstad's switching lemma. For $p \in [0,1]$ and $\ell_m \in \mathbb{N}$, we denote by $\mathcal{R}_p$ the distribution over restrictions in $\{0,1,\star\}^{\ell_m}$ that is obtained by independently setting each coordinate to $\star$ with probability p, and to a uniformly-chosen bit otherwise. Then,

Theorem 5.22 *(Håstad's switching lemma [Hås87]). Let $g : \{0,1\}^{\ell_m} \to \{0,1\}$ be a function computed by a depth-2 circuit of bottom fan-in w, and let $\rho \sim \mathcal{R}_p$. Then, for any $s \geq 1$, the probability that $g\restriction_\rho$ cannot be computed by a decision tree of depth s is at most $(5pw)^s$.*

Lemma 5.23. *Let $f : \{0,1\}^{\ell_m} \to \{0,1\}$ be computed by a circuit of depth four and size at most $S = 2^{\epsilon \cdot \ell_m}$ that has top fan-in $2 \leq c \leq O(1)$ and bottom fan-in $t = O(1)$ such that $\epsilon \leq \frac{1}{3 \cdot 5^2 \cdot 2^{24} \cdot t \cdot \log(c)}$. Then, for $p = \frac{1/\epsilon}{5^2 \cdot 2^{22} \cdot t} \cdot \frac{1}{\ell_m}$, with probability at least $1 - 4/c$ over choice of restriction $\rho \sim \mathcal{R}_p$ it holds that:*

1. *The restricted function $f\restriction_\rho$ is a depth-two circuit with bottom fan-in $\log(c)/10$.*
2. *The restriction ρ keeps at least $\log(c)$ variables alive.*

Proof. We prove the first item using Theorem 5.22. Specifically, we first apply a restriction with $p = 1/20t$. For each gate of distance two from the inputs we use Theorem 5.22 with values $s = \log(S)$ and $w = t$, and deduce that with probability $1 - 1/S^2$ the gate can be computed by a decision tree of depth $\log(S)$. After union-bounding over at most S such gates we, get that with probability $1 - 1/S = 1 - o(1)$ the circuit is of depth three with at most S^2 gates in the bottom layer, each of fan-in $\log(S)$, and at most c gates in the middle layer. Now condition on this event happening.

Next, we apply another restriction with $p = 1/(5 \cdot 2^{20} \cdot \log(S))$. Again, for each gate of distance two from the inputs we use Theorem 5.22 with values $w = \log(S)$ and

$s = \log(c)/10$, and deduce that with probability $1 - 1/c^2$ the gate can be computed by a decision tree of depth $\log(c)/10$. After union-bounding over the c (or less) gates in the middle layer, with probability at least $1 - 1/c$, after the restriction the circuit is of depth two with at most $c^{1+1/10}$ gates in the bottom layer, each of fan-in $\log(c)/10$.

To prove the second item, note that by the hypothesis that ϵ is sufficiently small, the expected number of living variables under $\rho \sim \mathcal{R}_p$ is at least $2\log(c)$. By a Chernoff bound (see, e.g., [Gol08, Apdx. D.1.2.3]), the probability that less than $\log(c)$ variables remain alive is at most $2\exp\left(-(1/3p) \cdot (p^2/4) \cdot \ell_m\right) = 2/c$. ■

We get the following corollary of Lemma 5.23:

Corollary 5.24. *Let $f : \{0,1\}^{\ell_m} \to \{0,1\}$ be computed by a circuit of depth four and size at most $S = 2^{\epsilon \cdot \ell_m}$ that has top fan-in $c = O(1)$ and bottom fan-in $t = O(1)$, where $\epsilon(t,c) = \frac{1}{3 \cdot 5^2 \cdot 2^{24} \cdot t \cdot \log(c)}$ and c is sufficiently large. Then, for $p = \frac{1/\epsilon}{5^2 \cdot 2^{22} \cdot t} \cdot \frac{1}{\ell_m}$ it holds that:*

1. *The expected value over $\rho \sim \mathcal{R}_p$ of the average sensitivity of $f\restriction_\rho$ is less than $\frac{1}{10} \cdot \log(c)$.*
2. *With probability $\Omega(1)$ over choice of random function g, the expected value over $\rho \sim \mathcal{R}_p$ of the average sensitivity of $g\restriction_\rho$ is at least $\frac{1}{5} \cdot \log(c)$.*

Consequently, there exists a polynomial-time algorithm that is given as input the truth-table of a Boolean function h over ℓ_m input bits, and satisfies the following: If h is a depth four circuits of top fan-in c, bottom fan-in t, and size $\leq 2^{\epsilon(t,c)\ell_m}$, then the algorithm rejects; and the algorithm accepts a constant fraction of all functions.

Proof. To prove the first item we use Lemma 5.23 as well as the fact (which was proved by Amano [Ama11]) that depth-two circuits with bottom fan-in $\log(c)/10$ have average sensitivity at most $\log(c)/10$. Thus, the expected average sensitivity of $f\restriction_\rho$ is at most $(4/c + 1/10) \cdot \log(c) < \frac{3}{20} \cdot \log(c)$, assuming c is sufficiently large.

For the second item, let us denote by $\mathtt{sens}(g\restriction_\rho)$ the average sensitivity of $g\restriction_\rho$; then, we have that:

Fact 5.24.1. *It holds that $\mathbb{E}_{g,\rho\sim\mathcal{R}_p}[\mathtt{sens}(g\restriction_\rho)] \geq (1 - 4/c)^2 \cdot \log(c)/4$.*

Proof. Condition on any restriction ρ that keeps at least $\log(c)$ variables alive. Then, the expected average sensitivity of g is at least $\log(c)/2$, and by a standard analysis, with probability at least $1 - 4/c$, the actual average sensitivity of g is at least $\log(c)/4$. [18] The statement follows since the choices of ρ and of g are independent. □

[18] In more detail, suppose g is a random function over $n' = \log c$ input variables, and let $E = n'2^{n'}/2$ be the number of edges in the n'-dimensional hypercube. Let X_e for $e \in [E]$ be the 0/1-random variable that is 1 if and only g flips its value along edge e. Then, for $X = \sum X_e$, we get $\mathbb{E}[X] = E/2$. Note that the average sensitivity of the random function g is precisely $2X/2^{n'}$. Therefore, it is enough for us to upper bound $\Pr[|X - E/2| \geq E/4]$, since if this event does not happen then the average sensitivity of g is at least $(1/2^{n'-1})(E/2 - E/4) \geq n'/4 = \log(c)/4$. For that, it suffices to use Chebyshev's inequality together with the pairwise independence of $\{X_e\}_{e\in[E]}$: $\mu_X = E/2$, $\sigma_X^2 = E/4$, and consequently the aforementioned probability is at most $4/E = 8/(\log(c) \cdot c) \leq 4/c$, where we have used $c \geq 4$.

Relying on Fact 5.24.1, the probability over g that $\mathbb{E}_\rho[\texttt{sens}(g\restriction_\rho)] < \log(c)/5$ is at most $\frac{1-(1-4/c)^2/4}{4/5}$, which is a constant smaller than one for a sufficiently large c.

For the "consequently" part, consider an algorithm that, when given a truth-table $h \in \{0,1\}^{2^{\ell_m}}$, enumerates over all restrictions $\rho \in \{0,1,\star\}^{\ell_m}$, computes the average sensitivity of $h\restriction_\rho$, and accepts if and only if $\mathbb{E}_\rho[\texttt{sens}(h\restriction_\rho)] > \log(c)/10$ (note that the algorithm weighs the restrictions according to their probabilities in $\mathcal{R}_p$, which can be done in time $\text{poly}(2^{\ell_m})$). Indeed, this algorithm rejects every h in the circuit class, but accepts $\Omega(1)$ of all functions. ■

By combining Claim 5.20 and Corollary 5.24, we get the following:

Proposition 5.25 *(breaking Goldreich's generator with a limited class of predicates, under Assumption 5.19). Suppose that Assumption 5.19 holds with parameters $t(k) \equiv \gamma$, for some constant $\gamma \in \mathbb{N}$, and $\beta > 3$. Then, there exists $\delta > 0$ such that for every $k \in \mathbb{N}$ there exists a polynomial-time algorithm A that satisfies the following. For every sufficiently large $n \in \mathbb{N}$ there exists an $(n^{0.99}, 0.99)$-expander $G = ([n], [m = n^k], E)$ with right-degree $\ell = \beta \cdot k$ such that for every predicate $P : \{0,1\}^\ell \to \{0,1\}$ that can be computed by a depth-two circuit of size $2^{\delta \cdot \ell}$, the algorithm A distinguishes the outputs of Goldreich's generator from random strings (with constant gap $\Omega(1)$).*

Proof. Let $k \in \mathbb{N}$, let δ be sufficiently small such that $1/k < \frac{1}{3 \cdot 5^2 \cdot 2^{24} \cdot \gamma \cdot \delta \cdot \beta \cdot k}$ (i.e., $\delta < 1/O(\gamma \cdot \beta)$, where the O-notation hides a universal constant), and let $n \in \mathbb{N}$ be sufficiently large. Let G be the expander from Assumption 5.19 with right-degree $\ell = \beta \cdot k$, and let $P : \{0,1\}^\ell \to \{0,1\}$ be a predicate computed by a depth-two circuit of size $2^{\delta \cdot \ell}$.

We instantiate Goldreich's generator with the expander G and predicate P. By Claim 5.20, for every fixed $x \in \{0,1\}^n$, the output of the generator on x is the truth-table of an $\mathcal{AC}^0$ circuit $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ of depth four, with top fan-in $c = 2^{\delta \cdot \ell}$, bottom fan-in γ, and size $O\left(2^{(1/k) \cdot \log(m)}\right)$. Using the fact that δ is sufficiently small, the size of g_x is $o\left(2^{\epsilon(\gamma,c) \cdot \ell_m}\right)$, where $\ell_m = \log(m)$ and $\epsilon(\gamma, c) = \frac{1}{3 \cdot 5^2 \cdot 2^{24} \cdot \gamma \cdot \log(c)}$. Thus, the natural property from Corollary 5.24 can distinguish the outputs of Goldreich's generator from random strings. ■

5.2.3 Natural properties that would suffice to break the generator with any predicate, under Assumption 5.19

We note that Assumption 5.19, Claim 5.20, and Corollary 5.24 do not suffice to break Goldreich's pseudorandom generator (in the polynomial stretch regime) for an *arbitrary predicate*. To see this, suppose that Assumption 5.19 holds with some $\beta > 3$ and function $t(k)$. Then, Claim 5.20 implies that to break the generator (with any predicate) it suffices to have a natural property against depth-four circuits over $\ell_m = \log(m)$ input bits with top fan-in $c = 2^{\beta \cdot k}$, bottom fan-in $t = t(k) = t(\log(c)/\beta)$ and size $O\left(2^{(1/k) \cdot \ell_m}\right) = O\left(2^{(\beta/\log(c)) \cdot \ell_m}\right)$. However, Corollary 5.24 only provides a natural property against circuits of size $2^{(1/O(t\cdot\log(c)))\cdot\ell_m}$; that is, the constant that precedes $(1/\log(c))\cdot\ell_m$ in the exponent is $1/O(t)<1$, instead of $\beta>3$.

Let us therefore consider the possibility that a natural property that is *mildly* stronger than the one in Corollary 5.24 exists. Intuitively, we restrict our attention to circuits with sufficiently large top fan-in c, and appropriately bounded bottom fan-in (i.e., the bottom fan-in is at most $t(\log(c)/\beta)$; this is in contrast to Corollay 5.24, which does not impose any dependency among the top and bottom fan-ins, as long as both are constant). In this setting, we ask for a natural property against depth-four circuits of size $O\left(2^{(\beta/\log(c))\cdot\ell_m}\right)$. To the best of our knowledge, such a natural property might exist even when the bound on the bottom fan-in is much more relaxed (e.g., for bottom fan-in $\log^*(n)$).

Theorem 5.26 *(breaking Goldreich's generator for any predicate, under Assumptions 5.19 and a natural property that is mildly stronger than Corollay 5.24). Suppose that Assumption 5.19 holds with parameters $t:\mathbb{N}\to\mathbb{N}$ and $\beta>3$. Also assume that for a sufficiently large constant $c>2^\beta$, there exists a natural property against the class of Boolean functions on ℓ_m input bits that are computed by depth four circuits of top fan-in c, bottom fan-in $t(\log(c)/\beta)$, and size $O\left(2^{(\beta/\log(c))\cdot\ell_m}\right)$, where the O-notation hides an arbitrarily large constant.*

Then, for a sufficiently large constant $k\in\mathbb{N}$, there exists a polynomial-time algorithm A that satisfies the following. For every $n\in\mathbb{N}$ there exists an $(n^{0.99},0.99)$-expander $G=([n],[m=n^k],E)$ with right-degree $\ell=\beta\cdot k$ such that for every predicate $P:\{0,1\}^\ell\to\{0,1\}$, the algorithm A distinguishes the outputs of Goldreich's generator from random strings (with constant gap $\Omega(1)$).

Proof. Let $k\in\mathbb{N}$ be sufficiently large such that $c=2^{\beta\cdot k}$ is sufficiently large to satisfy the hypothesis regarding the natural property, let $n\in\mathbb{N}$ be sufficiently large, let G be the expander from Assumption 5.19 with right-degree $\ell=\beta\cdot k$, and let $P:\{0,1\}^\ell\to\{0,1\}$ be an arbitrary predicate.

We instantiate Goldreich's generator with the expander G and predicate P. By Claim 5.20, every output string g_x of the generator can be computed by a depth four circuit over $\ell_m=\log(m)$ input variables with top fan-in $c=2^{\beta\cdot k}$, bottom fan-in $t(k)=t(\log(c)/\beta)$, and size $O\left(2^{(1/k)\cdot\ell_m)}\right)=O\left(2^{(\beta/\log(c))\cdot\ell_m}\right)$. Thus, the natural property from our hypothesis distinguishes the output of the generator from a random string with probability at least $1/2$. ■

Acknowledgements

The authors thank Emanuele Viola and Avi Wigderson for useful e-mail exchanges about the complexity of constructing unbalanced bipartite expander graphs. The authors are also grateful to Mahdi Cheraghchi, Oded Goldreich, and Benny Applebaum for very useful discussions. The first and second authors are supported by the second author's ERC-CoG grant no. 615075. The third author is partially supported by Irit Dinur's ERC-CoG grant no. 772839.

References

[ABGKR14] Adi Akavia, Andrej Bogdanov, Siyao Guo, Akshay Kamath, and Alon Rosen. "Candidate weak pseudorandom functions in $\mathsf{AC}^0 \circ \mathsf{MOD}\,2$". In: *Proc. Annual Innovations in Theoretical Computer Science Conference (ITCS)*. 2014, pp. 251–260.

[ABR16] Benny Applebaum, Andrej Bogdanov, and Alon Rosen. "A dichotomy for local small-bias generators". In: *Journal of Cryptology* 29.3 (2016), pp. 577–596.

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson. "Public-key cryptography from different assumptions". In: *Proc. 42nd Annual ACM Symposium on Theory of Computing (STOC)*. 2010, pp. 171–180.

[AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. "Cryptography in NC^0". In: *SIAM Journal of Computing* 36.4 (2006), pp. 845–888.

[Ajt90] Miklós Ajtai. "Approximate Counting with Uniform Constant-Depth Circuits." In: *Advances in computational complexity theory*. 1990, pp. 1–20.

[AL18] Benny Applebaum and Shachar Lovett. "Algebraic attacks against random local functions and their countermeasures". In: *SIAM Journal of Computing* 47 (1 2018), pp. 52–79.

[Ale11] Michael Alekhnovich. "More on average case vs approximation complexity". In: *Computational Complexity* 20.4 (2011), pp. 755–786.

[Ama11] Kazuyuki Amano. "Tight bounds on the average sensitivity of k-CNF". In: *Theory of Computing* 7 (2011), pp. 45–48.

[App14] Benny Applebaum. *Cryptography in Constant Parallel Time*. Information Security and Cryptography. Springer, 2014.

[App16] Benny Applebaum. "Cryptographic Hardness of Random Local Functions". In: *Computational Complexity* 25.3 (2016), pp. 667–722.

[AR16] Benny Applebaum and Pavel Raykov. "Fast pseudorandom functions based on expander graphs". In: *Theory of cryptography. Part I*. Vol. 9985. Lecture Notes in Comput. Sci. Springer, Berlin, 2016, pp. 27–56.

[BQ12] Andrej Bogdanov and Youming Qiao. "On the security of Goldreich's one-way function". In: *Computational Complexity* 21.1 (2012), pp. 83–127.

[BR13] Andrej Bogdanov and Alon Rosen. "Input locality and hardness amplification". In: *Journal of Cryptology* 26.1 (2013), pp. 144–171.

[Car10] Claude Carlet. "Boolean functions for cryptography and error-correcting codes". In: *Boolean models and methods in mathematics, computer science, and engineering*. Cambridge Univerity Press, 2010, pp. 257–397.

[CEMT14] James Cook, Omid Etesami, Rachel Miller, and Luca Trevisan. "On the one-way function candidate proposed by Goldreich". In: *ACM Transactions of Computation Theory* 6.3 (2014), Art. 14, 35.

[Che05] Mahdi Cheraghchi. “Applications of Derandomization Theory in Coding”. PhD thesis. École Polytechnique Fédérale de Lausanne, 2005.

[CIKK16] Marco L. Carmosino, Russell Impagliazzo, Valentine Kabanets, and Antonina Kolokolova. “Learning algorithms from natural proofs”. In: *Proc. 31st Annual IEEE Conference on Computational Complexity (CCC)*. 2016, 10 (24).

[CRVW02] Michael Capalbo, Omer Reingold, Salil Vadhan, and Avi Wigderson. “Randomness Conductors and Constant-degree Lossless Expanders”. In: *Proc. 34th Annual ACM Symposium on Theory of Computing (STOC)*. 2002, pp. 659–668.

[CS16] Gil Cohen and Igor Shinkar. “The complexity of DNF of parities”. In: *Proc. 7th Annual Innovations in Theoretical Computer Science Conference (ITCS)*. 2016, pp. 47–58.

[FPV15] Vitaly Feldman, Will Perkins, and Santosh Vempala. “On the Complexity of Random Satisfiability Problems with Planted Solutions”. In: *Proc. 47th Annual ACM Symposium on Theory of Computing (STOC)*. 2015, pp. 77–86.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. “How to construct random functions”. In: *Journal of the ACM* 33.4 (1986), pp. 792–807.

[Gol00] Oded Goldreich. “Candidate One-Way Functions Based on Expander Graphs”. In: *Electronic Colloquium on Computational Complexity: ECCC* 7 (2000), p. 90.

[Gol08] Oded Goldreich. *Computational Complexity: A Conceptual Perspective*. New York, NY, USA: Cambridge University Press, 2008.

[Gol11] Oded Goldreich. “Candidate one-way functions based on expander graphs”. In: *Studies in complexity and cryptography*. Vol. 6650. Lecture Notes in Computer Science. Springer, Heidelberg, 2011, pp. 76–87.

[GUV09] Venkatesan Guruswami, Christopher Umans, and Salil Vadhan. “Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes”. In: *Journal of the ACM* 56.4 (2009), Art. 20, 34.

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Márió Szegedy, and György Turán. “Threshold circuits of bounded depth”. In: *Journal of Computer and System Sciences* 46.2 (1993), pp. 129–154.

[Hås14] Johan Håstad. “On the correlation of parity and small-depth circuits”. In: *SIAM Journal of Computing* 43.5 (2014), pp. 1699–1708.

[Hås87] Johan Håstad. *Computational Limitations of Small-depth Circuits*. MIT Press, 1987.

[KS12] Swastik Kopparty and Srikanth Srinivasan. “Certifying polynomials for $\mathrm{AC}^0[\oplus]$ circuits, with applications”. In: *Proc. 32nd Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. 2012, pp. 36–47.

[MST06] Elchanan Mossel, Amir Shpilka, and Luca Trevisan. “On ϵ-biased generators in NC^0”. In: *Random Structures & Algorithms* 29.1 (2006), pp. 56–81.

[Nis91] Noam Nisan. “Pseudorandom bits for constant depth circuits”. In: *Combinatorica* 11.1 (1991), pp. 63–70.

[NW94] Noam Nisan and Avi Wigderson. “Hardness vs. randomness”. In: *Journal of Computer and System Sciences* 49.2 (1994), pp. 149–167.

[OS08] Ryan O’Donnell and Rocco A. Servedio. “Extremal properties of polynomial threshold functions”. In: *Journal of Computer and System Sciences* 74.3 (2008), pp. 298–312.

[OW14] Ryan O’Donnell and David Witmer. “Goldreich’s PRG: evidence for near-optimal polynomial stretch”. In: *Proc. 29th Annual IEEE Conference on Computational Complexity (CCC)*. 2014, pp. 1–12.

[Raz87] Alexander A. Razborov. “Lower bounds on the size of constant-depth networks over a complete basis with logical addition”. In: *Mathematical Notes of the Academy of Science of the USSR* 41.4 (1987), pp. 333–338.

[RR97] Alexander A. Razborov and Steven Rudich. “Natural proofs”. In: *Journal of Computer and System Sciences* 55.1, part 1 (1997), pp. 24–35.

[Smo87] Roman Smolensky. “Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity”. In: *Proc. 19th Annual ACM Symposium on Theory of Computing (STOC)*. 1987, pp. 77–82.

[Tre01] Luca Trevisan. “Extractors and Pseudorandom Generators”. In: *Journal of the ACM* 48.4 (2001), pp. 860–879.

[TUZ07] Amnon Ta-Shma, Christopher Umans, and David Zuckerman. “Lossless condensers, unbalanced expanders, and extractors”. In: *Combinatorica* 27.2 (2007), pp. 213–240.

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2012.

[Vio09] Emanuele Viola. “On Approximate Majority and Probabilistic Time”. In: *Computational Complexity* 18.3 (2009), pp. 337–375.

[VW17] Emanuele Viola and Avi Wigderson. “Local Expanders”. In: *Computational Complexity* (2017).

[Wig18] Avi Wigderson. Mathematics and Computation (book draft). Accessed at `https://www.math.ias.edu/avi/book`, August 26, 2018. August 26, 2018.

Appendix A Proof of Theorem 3.2: Unbalanced Expanders

Let us recall the statement of Theorem 3.2 and detail the proof.

Theorem A.1 *(a random graph is a good lossless expander). Let $n \leq m \in \mathbb{N}$ and $d \in \mathbb{N}$ such that $c \cdot \frac{\log(m)}{\log(n)} \leq \ell \leq n^{1/c}$, where c is a sufficiently large constant. Let $G = ([n], [m], E)$ be a random bipartite graph of right-degree ℓ; that is, the ℓ neighbors of each $i \in [m]$ are chosen uniformly in $[n]$ (allowing repetitions). Then, with probability at least $1 - 1/\mathrm{poly}(n)$ it holds that G is an $(n^{0.99}, 0.99)$-expander.*[19]

Proof. We say that a set $S \subseteq [m]$ expands if $|\Gamma(S)| \geq 0.99 \cdot \ell \cdot |S|$. For any fixed $k \leq n^{0.99}$, we show that with probability $1 - 1/\mathrm{poly}(n)$ it holds that all sets of size k expand (and the theorem follows by union-bounding over all values of k).

To do so, fix $k \leq n^{0.99}$, and fix a set $S \subseteq [m]$ of size $|S| = k$. For any fixed set $T \subseteq [n]$ of size less than $0.99 \cdot \ell \cdot k$, the probability that $\Gamma(S) \subseteq T$ is $\left(\frac{|T|}{n}\right)^{\ell \cdot k}$. Union-bounding over all $T \subseteq [n]$ of such size, the probability that S does not expand is at most

$$\binom{n}{0.99 \cdot \ell \cdot k} \cdot \left(\frac{|T|}{n}\right)^{\ell \cdot k} < \left(\frac{e \cdot n}{0.99 \cdot \ell \cdot k}\right)^{0.99 \cdot \ell \cdot k} \cdot \left(\frac{|T|}{n}\right)^{\ell \cdot k} < \left(\frac{O(|T|)}{n}\right)^{0.01 \cdot \ell \cdot k} .$$

Union-bounding over all sets $S \subseteq [m]$ of size $|S| = k$, the probability that there exists a non-expanding set of size k is at most

$$\begin{aligned} \binom{m}{k} \cdot \left(\frac{O(|T|)}{n}\right)^{0.01 \cdot \ell \cdot k} &< \left(\frac{e \cdot m}{k}\right)^{k} \cdot \left(O(\ell) \cdot \frac{k}{n}\right)^{0.01 \cdot \ell \cdot k} \\ &= 2^{O(\ell \cdot \log(\ell) \cdot k)} \cdot \left(\frac{m}{k}\right)^{k} \cdot \left(\frac{k}{n}\right)^{0.01 \cdot \ell \cdot k} \\ &< 2^{O(\ell \cdot \log(\ell) \cdot k)} \cdot n^{k\ell/c - 0.01 \cdot \ell \cdot k} \cdot k^{0.01 \cdot \ell \cdot k} , \qquad \text{(A.1)} \end{aligned}$$

where the last inequality relied on the fact that $m < n^{\ell/c}$ (since $\ell \geq c \cdot (\log(m)/\log(n))$). The expression in Eq. (A.1) is exponential in

$$\begin{aligned} &O(k \cdot \ell) \cdot \Big(\log(\ell) + (1/c - 0.01) \cdot \log(n) + 0.01 \cdot \log(k)\Big) \\ &\leq O(k \cdot \ell) \cdot \left(2/c - 10^{-4}\right) \cdot \log(n) , \end{aligned}$$

where the last inequality relied on the fact that $k \leq n^{.99}$ and on the fact that $c \leq \ell \leq n^{1/c}$. Assuming that c is sufficiently large such that $2/c - 0.01^2 < -10^{-5}$, the probability that there exists a set of size k that does not expand is at most $n^{-10^{-5} \cdot c}$. As mentioned above, the theorem follows by union-bounding over $k = 1, ..., n^{0.99}$. ■

We note that the right-degree bound provided by Theorem A.1 does not contradict the impossibility results discussed for instance in **[DBLP:journals/jacm/GuruswamiUV09]**. This is because the expanding sets in Theorem A.1 are of size at most $n^{1-\Omega(1)}$ instead of $\Omega(n)$.

[19] The polynomial power depends on c and can be arbitrarily large.

Appendix B A comparison with the work of Applebaum and Raykov [AR16]

A specific construction of pseudorandom functions based on Goldreich's pseudorandom generator was studied by Applebaum and Raykov [AR16]. We note several technical similarities between our work and their constructions. In their work they constructed a pseudorandom function $f_x : \{0,1\}^r \to \{0,1\}^m$, where $x \in \{0,1\}^n$ is a fixed key, as follows. The function f_x uses x as a fixed seed for Goldreich's generator, gets as input a description $G \in \{0,1\}^r$ of a bipartite graph, and outputs the evaluation of Goldreich's generator instantiated with G and with some "good" universal predicate P on the fixed seed x. Their main idea is that if the queries $G_1, ..., G_q \in \{0,1\}^r$ for the function are chosen at random, then the concatenation of the outputs $f_x(G_1) \circ ... \circ f_x(G_q)$ is the evaluation of Goldreich's generator on the seed x and the graph $G_1 \cup ... \cup G_q$, which is (with high probability) an expander. Thus, assuming that Goldreich's generator is secure, the outputs of f_x cannot be distinguished from random strings.[20]

The function $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$ defined in the beginning of Section 2 is similar to their pseudorandom function f_x in the sense that both functions are defined with respect to a fixed seed x for Goldreich's generator, get *some* description of an ℓ-tuple of coordinates in $[n]$ (or of a list of such ℓ-tuples), and output evaluation of Goldreich's generator with x on the corresponding ℓ-tuple/s.

However, in their setting the output of Goldreich's generator is used as the *non-Boolean output* of a function $f_x : \{0,1\}^r \to \{0,1\}^m$, whereas our main conceptual step is to consider the output of Goldreich's generator as the *truth-table* of a Boolean function $g_x : \{0,1\}^{\log(m)} \to \{0,1\}$. Moreover, our function g_x is defined with respect to a *fixed* expander, whereas their function f_x gets as input the (potential) hyperedges of an expander, and a specific expander is only defined after queries to f_x are issued. These are crucial differences, since our function needs to efficiently map $\log(m)$ input bits, which represent a vertex $i \in [m]$, to $\ell \cdot \log(n) \gg \log(m)$ bits representing the neighbors of i. In contrast, in their setting the input size can be much larger (e.g., $\ell \cdot \log(n)$, or even close to n), and thus no concise representation of the expander is needed. (Indeed, they use either the naive mapping of $\ell \cdot \log(n)$ bits to ℓ vertices in $[n]$, or a mapping which relies on a (non-expanding) hash function.)

[20]This description only yields a *weak* pseudorandom function; they also get (standard, non-weak) pseudorandom functions against non-adaptive adversaries by refining the construction.

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