BOUNDED-DEPTH CIRCUITS CANNOT SAMPLE GOOD CODES

Shachar Lovett and Emanuele Viola

February 8, 2012

Abstract. We study a variant of the classical circuit-lower-bound problems: proving lower bounds for sampling distributions given random bits. We prove a lower bound of $1 - 1/n^{\Omega(1)}$ on the statistical distance between (i) the output distribution of any small constant-depth (a.k.a. AC^0) circuit $f : \{0,1\}^{\text{poly}(n)} \rightarrow \{0,1\}^n$, and (ii) the uniform distribution over any code $\mathcal{C} \subseteq \{0,1\}^n$ that is "good", i.e. has relative distance and rate both $\Omega(1)$. This seems to be the first lower bound of this kind.

We give two simple applications of this result: (1) any data structure for storing codewords of a good code $\mathcal{C} \subseteq \{0,1\}^n$ requires redundancy $\Omega(\log n)$, if each bit of the codeword can be retrieved by a small AC^0 circuit; (2) for some choice of the underlying combinatorial designs, the output distribution of Nisan's pseudorandom generator against AC^0 circuits of depth d cannot be sampled by small AC^0 circuits of depth less than d.

Keywords. Sampling, Small depth circuits, noise sensitivity, isoperimetric inequalities

Subject classification. 68Q17

1. Introduction

The classical problem in computational complexity is to prove "lower bounds," that is to show that certain functions cannot be computed or approximated in various computational models. A few works, such as the ones by Jerrum *et al.* (1986), Ambainis *et al.* (2003), Goldreich *et al.* (2010), and by Viola (2010) study instead the complexity of generating – or sampling – certain distributions.

Viola (2010) raises the problem of exhibiting any explicit boolean function $b : \{0,1\}^n \rightarrow \{0,1\}$ such that no small, unbounded fan-in constant-depth circuit (i.e., AC^0) can generate the distribution $(x, b(x))$ given random bits.

To illustrate the differences between computing a function and sampling a distribution, consider for example the Parity function $\text{Parity}(x_1, \ldots, x_n) := x_1 \oplus \ldots \oplus x_n$. A classical result of Håstad (1987) shows that Parity cannot be approximated by unbounded fan-in constant-depth (i.e., AC^0) small circuits with better than exponentially small bias. It is possible however to sample an (input,output) pair $(x_1, \ldots, x_n, \text{Parity}(x_1, \ldots, x_n))$ in AC^0: let $y_1, \ldots, y_{n+1}$ be uniform bits, and take $x_i = y_i \oplus y_{i+1}$ and $\text{Parity}(x_1, \ldots, x_n) = y_1 \oplus y_{n+1}$.

In this work we solve the variant of the problem raised in Viola (2010) where b is a function with long output length (not boolean). Specifically, we prove that small AC^0 circuits cannot approximate uniform distributions over good codes, where approximation is measured by the statistical distance between the two corresponding distributions D' and D'':

$$\text{sd}(D', D'') = \max_S |\Pr[D' \in S] - \Pr[D'' \in S]|.$$

A subset $\mathcal{C} \subset \{0,1\}^n$ is an (n, k, d) code if $|\mathcal{C}| = 2^k$ and the hamming distance between any two distinct codewords $x, y \in \mathcal{C}$ is at least d. A code $\mathcal{C}$ is *good* if $k = \Omega(n)$ and $d = \Omega(n)$. As is well known, there exist explicit constructions of good codes. We denote by U_m the uniform distribution over $\{0,1\}^m$ and by $U_{\mathcal{C}}$ the uniform distribution over codewords of $\mathcal{C}$.

THEOREM 1.1 (Small AC^0 circuits cannot sample codes). *Let* $F : \{0,1\}^m \rightarrow \{0,1\}^n$ *be a function computable by an AC^0 circuit of depth t and size M. Let $\mathcal{C} \subset \{0,1\}^n$ be an (n, k, d)-code. Then*

$$\text{sd}(F(U_m), U_{\mathcal{C}}) \geq 1 - O\left(\frac{n}{dk} \cdot \log^{t-1} M\right)^{1/3}.$$

In particular, if $\mathcal{C}$ is a good code, $t = O(1)$, and $M = \text{poly}(n)$ then $\text{sd}(F(U_m), U_{\mathcal{C}}) \geq 1 - 1/n^{\Omega(1)}$.

It is well-known (and we review it in Section 1.1) that small AC^0 circuits cannot compute the encoding function of any good error-correcting code; our lower bound is stronger in that it applies even if the circuit is given as input a number of random bits that is longer than the message length of the code. Furthermore, we achieve statistical distance approaching one, which is crucial for a couple of applications mentioned below. It may even be true that the statistical distance approaches 1 exponentially fast, as opposed to polynomially fast in our result. But our techniques seem unable to establish this, and more generally we raise the question of proving such a statistical bound for any explicit distribution.

We mention that after this work, and using the ideas in it, there has been progress on proving sampling lower bounds on AC^0 circuits, see Viola (2011).

We next discuss two applications of Theorem 1.1. From a technical point of view, the applications are straightforward corollaries to the theorem.

Data structures.   As pointed out in Viola (2010), proving lower bounds approaching 1 on the statistical distance between the output of a circuit and some flat distribution T on $\{0, 1\}^n$ implies data structures lower bounds for storing elements t in the support of T succinctly while retrieving each bit of t efficiently. In particular, one obtains the following lower bound for storing codewords.

COROLLARY 1.2. *Let $\mathcal{C}$ be an (n, k, d) code with $kd \geq n^{1+\Omega(1)}$. Suppose we can store codewords of $\mathcal{C}$ using only $k + r$ bits so that each bit of the codeword can be computed by an AC^0 circuit of depth $O(1)$ and size $\mathrm{poly}(n)$. Then $r \geq \Omega(\log n)$.*

PROOF.   Assume for the sake of contradiction that it is possible. Consider the AC^0 circuit $F : \{0, 1\}^{k+r} \to \{0, 1\}^n$ computing the codeword bits. For a random input to F, the output distribution of F has statistical distance $\leq 1 - 2^{-r}$ from the uniform distribution over codewords. By Theorem 1.1, $2^{-r} \leq 1/n^{\Omega(1)}$ hence $r \geq \Omega(\log n)$. $\square$

Note that without the restriction that the bits are retrievable by small AC^0 circuits, $r = 0$ is possible.

The model in Corollary 1.2 generalizes standard models such as bit-probe and cell-probe (for background, see Miltersen (1999)): it is easy to see that one can simulate cell-probes by small AC^0 circuits, while the lower bound in Corollary 1.2 holds even if one is allowed to look at the entire data structure, as long as the computation is done efficiently in AC^0. One can think of this as placing a lower bound on data structures where queries are answered quickly *in parallel*. This seems to be the first result of this kind.

We note that Gál & Miltersen (2007) prove a bit-probe lower bound for the same data structure problem as in Corollary 1.2. If the data structure is allowed q queries, then they show that it must have redundancy $r \geq \Omega(kd/nq)$. It is conceivable that one can obtain their result (or even improve it) by improving the bound in Theorem 1.1 to be exponentially close to one.

The complexity of Nisan's generator against AC^0. In this section we discuss the consequences of our results for the complexity of Nisan's generator given in Nisan (1991) against small bounded-depth circuits (AC^0 circuits). As typical of the Nisan-Wigderson style pseudorandom generators, computing Nisan's generator requires more resources than the circuits it is supposed to fool: Nisan's generator against circuits of depth d and size n (taking $\leq n$ input bits) computes the parity function on inputs of length ℓ that, loosely, is $\geq \log^{d+1} n$, and thus to be computed in size $\text{poly}(n)$ the generator requires depth $\geq d+1$.[1] However, it was not clear if such a lower bound on the complexity of computing the generator still holds if we only want to produce its output distribution, which is all that matters for pseudorandomness purposes. In this section we give a first answer to this question by showing that qualitatively the same lower bound applies for this task too, even up to

---

[1]The distinction between $d+1$ and d is irrelevant for the main message of this section. But we mention it arises because (1) up to lower order terms, the minimum size of depth-d circuits for parity on ℓ bits is $\exp(\ell^{1/(d-1)})$ Håstad (1987), and (2) the depth increases by 1 in the proof of correctness of Nisan's generator.

a constant statistical distance, for a particular implementation of Nisan's generator as we explain next.

Nisan's generator $G : \{0,1\}^k \to \{0,1\}^n$ can be written as $G(x) = Mx$ where M is an $n \times k$ matrix and multiplication is modulo 2. The rows of M are characteristic vectors of a design with set-size ℓ and intersection size $\leq \log n$, which means that each row has hamming weight exactly ℓ and any two rows share at most $\log n$ ones. To fool circuits of depth d, one sets ℓ sufficiently larger than $\log^{d+1} n$ and $k = \text{poly}(\ell)$. Nisan's proof works for any choice of M satisfying the above constraints. We now exhibit a particular matrix M satisfying the constraints such that generating the distribution Mx requires circuits of depth $\geq d$. This is accomplished by showing a matrix satisfying the constraints that is also the generator matrix of a good code, and then applying Theorem 1.1.

THEOREM 1.3. *Let $\ell = \ell(n)$ and $k = k(n)$ be functions such that $k \geq 4\ell^2, n = \omega(k^3)$, and $\ell(n) \geq \log(n)$ is odd.*
For arbitrarily large n, there is an $n \times k$ matrix such that:

(1) *M forms a design: each row of M has hamming weight ℓ, and any two rows share at most $\log n$ ones, and*

(2) *Any AC^0 circuit of size s and depth c whose output distribution (over uniform input) has statistical distance less than $1/2$ from Mx (over uniform $x \in \{0,1\}^k$) satisfies $\log^{c-1} s = \Omega(\ell)$.*

In particular, if one wants to compute the generator for $\ell \geq \log^{d+1} n$ by an AC^0 circuit of size $s = \text{poly}(n)$ then depth $c \geq d$ is required. Except for the arbitrariness in the choice of the underlying designs, this theorem shows an inherent inefficiency in Nisan's generator. By contrast, there is an alternative generator in Viola (2010) (based on the results in Bazzi (2009); Braverman (2010); Razborov (2009) and in Guruswami *et al.* (2009)) which fools circuits of depth d and can be computed by small depth-2 circuits.

1.1. Techniques.   In this section we explain the techniques behind the proof of Theorem 1.1. In short, the result is obtained

by combining bounds on the noise-sensitivity (a.k.a. average-sensitivity) of small AC^0 circuits with isoperimetric inequalities for the boolean cube. The techniques apply to any model with "low" noise-sensitivity; we focus on AC^0 circuits for concreteness.

We start by recalling the low noise-sensitivity of AC^0 circuits Boppana (1997); Linial *et al.* (1993). We use the following version, given explicitly in (Viola 2004, Lemma 6.6). Let $f : \{0,1\}^m \to \{0,1\}$ be an AC^0 circuit of depth t and size M. Then

$$\Pr_{x \in U_m, e \in \mu_p} [f(x) \neq f(x+e)] \leq O\left(p \cdot \log^{t-1} M\right),$$

where U_m is the uniform distribution over $\{0,1\}^m$, '$+$' denotes bit-wise xor, and $e \in \mu_p$ is obtained by setting each bit independently to 1 with probability p. We explain our ideas in stages, thinking of $M = \mathrm{poly}(n), t = O(1)$, so that $\log^{t-1} M = \mathrm{poly} \log n$.

Why small AC^0 circuits cannot compute good codes. Using the low noise-sensitivity of AC^0 it is easy to see that a small AC^0 circuit f cannot compute the encoding function $E : \{0,1\}^{k=\Omega(n)} \to \{0,1\}^n$ of a code with minimum distance $d = \Omega(n)$: If we choose $x \in \{0,1\}^k$ at random and let $e \in \mu_{1/k}$ then $f(x)$ and $f(x+e)$ have expected hamming distance only $n(1/k) \mathrm{poly} \log n = \mathrm{poly} \log n$, while on the other hand (if $e \neq 0$) the two codewords should have hamming distance $\geq d = \Omega(n)$. This gives a contradiction and proves that small AC^0 circuits cannot compute good codes.

Warm-up: lower bound for generating a codeword. Imagine now that the circuit is given as input not a number of bits equal to the message length, but $m \gg k$ bits, think $m = n^{100}$, and we would like to show that its output distribution cannot be uniform over codewords (statistical distance 0). The argument from the previous paragraph does not apply any more because it could happen that $f(x) = f(x+e)$ with high probability. We reason as follows. For any codeword $y \in \{0,1\}^n$ let $f^{-1}(y) \subseteq \{0,1\}^m$ be the set of input bits causing the circuit to output y. If we show that, no matter how the sets $f^{-1}(y)$ are placed, with high probability over the choice of x and e the inputs $x, x+e$ fall into different sets $f^{-1}(y)$, then we can carry through the same argument as before.

To argue this, we use the edge-isoperimetric inequality over the hamming cube Harper (1964); Hart (1976). This states that for any set $S \subseteq \{0,1\}^m$, the number of edges (unordered pairs of nodes at distance 1) with one endpoint in S and the other outside of S is $\geq |S|(m - \log_2 |S|)$, which is tight if S is a subcube. Therefore, no matter where x lands, assuming for simplicity that e has hamming weight 1, we have over the choice of such an e that the probability that $x + e$ lands in a different set is

$$\geq \frac{m - \log_2 2^m/2^k}{m} = \frac{k}{m}.$$

Hence the expected hamming distance between $f(x)$ and $f(x + e)$ is $\geq (k/m)d$. On the other hand, by low noise-sensitivity of AC^0 ($p = 1/m$) it is only $(n \operatorname{poly} \log n)/m$, which yields a contradiction as long as $kd \gg n$.

Obtaining statistical distance $1-\epsilon$.   To explain the techniques we use to improve the bound in the previous paragraph to a $1 - \epsilon$ statistical distance bound, consider the model case in which the circuit f outputs a codeword with probability ϵ over the input, and we have no control on its output for the other $1 - \epsilon$ fraction of inputs. To use noise-sensitivity, we need to argue that both $f(x)$ and $f(x + e)$ are valid codewords.

We note that using the edge-isoperimetric inequality in a straight forward manner one cannot get error below $\epsilon < 1/2$, since there are sets $S \subset \{0,1\}^m$ of size $|S| \geq 2^{m-1}$ which contain no edges (e.g., the set of all $\{0,1\}^m$ strings with parity 0). Thus, if F maps S to codewords and $\{0,1\}^m \setminus S$ to non-codewords, then at least one of $f(x), f(x+e)$ is always a non-codeword and we cannot argue by using the minimal distance of the code.

To improve the statistical distance bound to make it approach 1, we increase the noise parameter p in the definition of e. Using a symmetrization argument this resolves the problem of showing that both $f(x)$ and $f(x + e)$ are codewords with noticeable probability, but leaves the problem of analyzing the boundary of sets with respect to noise. We make use of a more sophisticated isoperimetric inequality that applies to vectors perturbed to noise: for any set

$A \subset \{0,1\}^m$, and any $0 \le p \le 1/2$

$$\left(\frac{|A|}{2^m}\right)^2 \le \Pr_{x \in U_m, e \in \mu_p}[x \in A, x + e \in A] \le \left(\frac{|A|}{2^m}\right)^{1/(1-p)}. \qquad (\star)$$

These inequalities and their proofs were pointed out to us by Alex Samorodnitsky. The first inequality is the "symmetrization argument" we alluded to before, and it is proved via the Cauchy-Schwarz inequality. The second inequality is based on the hypercontractivity theorem (often credited to Bonami, Beckner, and Gross). The inequalities appear to be folklore but we could not find them in the literature. Note that we do not claim that the inequalities are a contribution of this paper.

The proof then proceeds as follows. For simplicity, consider again a model case in which the input universe $\{0,1\}^m$ is made of a $1 - \epsilon$ fraction of inputs over which we have no control, and the other $\epsilon 2^m$ inputs are uniformly partitioned into 2^k sets $A_1, \ldots, A_{2^k}$ each corresponding to a codeword. Following the previous outline, we would like to argue that with noticeable probability $x \in A_i$ and $x + e \in A_j$ for $i \ne j$. We set the noise parameter to

$$p := \log(4/\epsilon)/k.$$

Now, by the left inequality in $(\star)$ we get that the probability that both x and $x + e$ fall into $\bigcup_i A_i$ is $\ge \epsilon^2$. On the other hand, by the right inequality in $(\star)$ the probability of falling into the same set A_i is at most

$$\sum_i (|A_i|/2^m)^{1/(1-p)} \le \sum_i (|A_i|/2^m)^{1+p} \le \epsilon \cdot (\epsilon/2^k)^p \le \epsilon^2/2.$$

Thus with probability $\ge \epsilon^2/2$ we have that $x \in A_i$ and $x + e \in A_j$ for $i \ne j$, in which case the hamming distance between the output of f should be d. Thus the expected hamming distance between $f(x)$ and $f(x + e)$ is $\Omega(d\epsilon^2)$.

On the other hand, the same expected hamming distance is at most $n \cdot p \cdot \operatorname{poly} \log n = n(\operatorname{poly} \log n)/k$ by the low noise-sensitivity of AC^0 circuits. Combining these two bounds gives the result:

$$\Omega(d\epsilon^2) \le n(\operatorname{poly} \log n)/k.$$

Organization: We prove our main lower bound in Section 2. Theorem 1.3 is proved in Section 3. In Section 4 we discuss a possible way to attack the problem, mentioned at the beginning of the introduction, of exhibiting an explicit *boolean* function b such that AC^0 cannot generate $(x, b(x))$. For completeness, a proof of the isoperimetric inequalities is in Section 5.

2. Lower bound for sampling good codes in AC^0

We prove Theorem 1.1 in this section, restated next.

Theorem 1.1, restated. (Small AC^0 circuits cannot sample codes). *Let $F : \{0,1\}^m \to \{0,1\}^n$ be a function computable by an AC^0 circuit of depth t and size M. Let $\mathcal{C} \subset \{0,1\}^n$ be an (n, k, d)-code. Then*

$$\mathrm{sd}(F(U_m), U_{\mathcal{C}}) \geq 1 - O\left(\frac{n}{dk} \cdot \log^{t-1} M\right)^{1/3}.$$

In particular, if $\mathcal{C}$ is a good code, $t = O(1)$, and $M = \mathrm{poly}(n)$ then $\mathrm{sd}(F(U_m), U_{\mathcal{C}}) \geq 1 - 1/n^{\Omega(1)}$.

Although one could work with expected hamming distance as in the introduction, we prove Theorem 1.1 using a certain extension of the notion of noise sensitivity of a function, which we now define.

DEFINITION 2.1 (Noise sensitivity). *Let $x \in U_m$ be uniform over $\{0,1\}^m$. A sample $e \in \mu_p$ from the p-biased distribution μ_p on $\{0,1\}^m$ is obtained by setting each bit e_i of e independently to 1 with probability p. For any $x \in \{0,1\}^m$, we denote by $x + e$ the bit-wise xor of x and e. We define the noise sensitivity of $f : \{0,1\}^m \to \{0,1\}^n$ with regards to a set $S \subseteq \{0,1\}^n$ as the probability that f maps $x, x + e$ to distinct elements of S:*

$$\mathrm{NS}_p(f; S) := \Pr_{x \in U_m, e \in \mu_p}[f(x) \in S, f(x + e) \in S, f(x) \neq f(x + e)].$$

The standard noise sensitivity of a function corresponds to $n = 1$ and $S = \{0,1\}$.

The proof of the theorem is deduced from the following lemmas. The first shows that if $\mathcal{C}$ is large enough, then for any function F

whose output distribution is not too far from $U_{\mathcal{C}}$, we must have that $NS_p(F;\mathcal{C})$ is relatively large. In fact, we prove this for any large enough set S.

LEMMA 2.2. *Let $S \subset \{0,1\}^n$ be a set. Let $F : \{0,1\}^m \to \{0,1\}^n$ be a function such that $\mathrm{sd}(F(U_m), U_S) \leq 1 - \epsilon$. Then for any $0 < p \leq 1/2$, if $|S| \geq (4/\epsilon)^{1/p}$ then*

$$\mathrm{NS}_p(F;S) \geq \epsilon^2/8.$$

We prove Lemma 2.2 in Section 2.1. We then show that any function in AC^0 must have small noise sensitivity with regards to a code $\mathcal{C}$ of good distance.

LEMMA 2.3 (AC^0 circuits have low noise sensitivity w.r.t. codes). *Let $F : \{0,1\}^m \to \{0,1\}^n$ be an AC^0 circuit of depth t and size M. Let $\mathcal{C}$ be an (n, k, d) code. Then for any $0 < p \leq 1/2$ we have*

$$\mathrm{NS}_p(F;\mathcal{C}) \leq O\left(p \cdot \frac{n}{d} \log^{t-1} M\right).$$

We prove Lemma 2.3 in Section 2.2. We now deduce Theorem 1.1 from Lemma 2.2 and Lemma 2.3.

PROOF (Proof of Theorem 1.1). Let $\mathrm{sd}(F(U_m), U_{\mathcal{C}}) := 1 - \epsilon$. First, note that we can assume $k \geq 2\log(4/\epsilon)$, for else the conclusion of the theorem holds (using $d \leq n, M \geq 2$ and a sufficiently large constant in the $O(\cdot)$). Let

$$p := \frac{\log(4/\epsilon)}{k},$$

so that $|\mathcal{C}| = 2^k \geq (4/\epsilon)^{1/p}$. Since $k \geq 2\log(4/\epsilon)$, we have $p \leq 1/2$. Applying Lemma 2.2 for $S = \mathcal{C}$ and Lemma 2.3 we get that

$$\epsilon^2/8 \leq NS_p(F;\mathcal{C}) \leq O\left(p \cdot \frac{n}{d} \log^{t-1} M\right).$$

Hence we deduce that

$$\epsilon \leq O\left(\frac{n}{dk} \cdot \log^{t-1} M\right)^{1/3}.$$

$\square$

2.1. Noise sensitivity of distributions close to uniform over a large set. We prove Lemma 2.2 in this subsection. We restate it below for the convenience of the reader.

Lemma 2.2, restated. *Let $S \subset \{0,1\}^n$ be a set. Let $F : \{0,1\}^m \to \{0,1\}^n$ be a function such that $\mathrm{sd}(F(U_m), U_S) \leq 1 - \epsilon$. Then for any $0 < p \leq 1/2$, if $|S| \geq (4/\epsilon)^{1/p}$ then*

$$\mathrm{NS}_p(F; S) \geq \epsilon^2/8.$$

We will need the following lemma, already stated in the introduction $(\star)$.

LEMMA 2.4. *Let $A \subseteq \{0,1\}^m$ and $\alpha := |A|/2^m$. Then for any $0 \leq p \leq 1/2$ we have*

$$\alpha^2 \leq \Pr_{x \in U_m, e \in \mu_p} [x \in A, x + e \in A] \leq \alpha^{1/(1-p)} \leq \alpha^{1+p}.$$

The lower bound follows from a simple convexity argument. The upper bound uses hypercontractivity. Both are given in Section 5. For the familiar reader, we note that the upper bound is equivalent to small set expansion for the noisy hypercube.

We conclude the following corollary. If $A_1, \ldots, A_t \subset \{0,1\}^m$ are disjoint subsets, each of which is small, but whose union $\cup_{i=1}^t A_i$ is large, then with good probability x and $x + e$ belong to distinct sets.

COROLLARY 2.5. *Let $A_1, \ldots, A_t \subset \{0,1\}^m$ be disjoint subsets, such that $|A_1|, \ldots, |A_t| \leq \alpha \cdot 2^m$ and $|\cup_{i=1}^t A_i| = \epsilon \cdot 2^m$. Then*

$$\Pr_{x \in U_m, e \in \mu_p} [\exists i \neq j \text{ such that } x \in A_i, x + e \in A_j] \geq \epsilon(\epsilon - \alpha^p).$$

PROOF. Let $A = \cup_{i=1}^t A_i$. We have

$$\Pr[\exists i \neq j \text{ such that } x \in A_i, x + e \in A_j]$$

$$= \Pr[x, x + e \in A] - \sum_{i=1}^t \Pr[x, x + e \in A_i].$$

Thus, we need to lower bound the probability that both $x, x+e \in A$ and to upper bound the probability that $x, x + e \in A_i$ for any specific set A_i. By Lemma 2.4 we have

$$\Pr[x \in A, x + e \in A] \geq \epsilon^2$$

and, for any set A_i,

$$\Pr[x \in A_i, x + e \in A_i] \leq \left(\frac{|A_i|}{2^m}\right)^{1+p} \leq \frac{|A_i|}{2^m} \cdot \alpha^p.$$

Since $\sum_{i=1}^{t} |A_i| = |A| = \epsilon \cdot 2^m$ we conclude that

$$\Pr[\exists i \neq j \text{ s.t. } x \in A_i, x + e \in A_j] \geq \epsilon(\epsilon - \alpha^p).$$

$\square$

We also use the following claim.

CLAIM 2.6. *Let $S \subset \{0,1\}^n$ be a set. Let D be a distribution over $\{0,1\}^n$ such that $\mathrm{sd}(D, U_S) \leq 1 - \epsilon$. Let $E := \{x \in S : D(x) \leq \frac{2/\epsilon}{|S|}\}$. Then*

$$\Pr_{x \in D}[x \in E] \geq \epsilon/2.$$

PROOF.   We will show that $\Pr_{x \in U_S}[x \in E] \geq 1 - \epsilon/2$. Since by assumption $\mathrm{sd}(D, U_S) \leq 1 - \epsilon$ this will imply that $\Pr_{x \in D}[x \in E] \geq \epsilon/2$ as claimed. Let $E' = S \setminus E = \{x \in S : D(x) \geq \frac{2/\epsilon}{|S|}\}$. Note that since $\sum_{x \in E'} D(x) \leq 1$ we get that $|E'| \leq (\epsilon/2)|S|$. Thus $\Pr_{x \in U_S}[x \in E'] \leq \epsilon/2$. Since U_S is supported on S we conclude since $\Pr_{x \in U_S}[x \in E] = 1 - \Pr_{x \in U_S}[x \in E'] \geq 1 - \epsilon/2$. $\square$

We now have all the ingredients to prove Lemma 2.2.

PROOF (Proof of Lemma 2.2).   Let $D = F(U_m)$ be the output distribution of F. Let $E = \{x \in S : D(x) \leq \frac{2/\epsilon}{|S|}\}$. Since $\mathrm{sd}(D, U_S) \leq 1 - \epsilon$ we have by Claim 2.6 that

$$\Pr_{x \in \{0,1\}^m}[F(x) \in E] \geq \epsilon/2.$$

For any $y \in E$ let $A_y \subset \{0,1\}^m$ be the preimage of y under F,

$$A_y = F^{-1}(y) = \{x \in \{0,1\}^m : F(x) = y\}.$$

Note that by definition we have that the sets $\{A_y\}$ are disjoint, that $|A_y| \leq \frac{2/\epsilon}{|S|} \cdot 2^m$ for every $y \in E$, and that $|\cup_{y \in E} A_y| \geq (\epsilon/2) \cdot 2^m$. Let $\epsilon' = \epsilon/2$. Hence by Corollary 2.5 we have that

$$\Pr_{x \in U_m, e \in \mu_p} [F(x) \in S, F(x+e) \in S, F(x) \neq F(x+e)]$$

$$\geq \Pr_{x \in U_m, e \in \mu_p} [\exists y' \neq y'' \in E \text{ such that } x \in A_{y'}, x+e \in A_{y''}]$$

$$\geq \epsilon'(\epsilon' - (\epsilon'/|S|)^p).$$

Thus to conclude we just need to verify that the condition $|S| \geq (4/\epsilon)^{1/p}$ implies that

$$(\epsilon'/|S|)^p \leq 1/|S|^p \leq \epsilon'/2,$$

and we get that

$$\mathrm{NS}_p(F; S) = \Pr_{x \in U_m, e \in \mu_p} [F(x) \in S, F(x+e) \in S, F(x) \neq F(x+e)]$$

$$\geq (\epsilon')^2/2 = \epsilon^2/8$$

as claimed.   □

2.2. Noise sensitivity of AC^0 functions with respect to codes. We prove Lemma 2.3 in this subsection. We restate it below for the convenience of the reader.

Lemma 2.3, restated. (AC^0 circuits have low noise sensitivity w.r.t. codes). Let $F : \{0,1\}^m \to \{0,1\}^n$ be an AC^0 circuit of depth t and size M. Let $\mathcal{C}$ be an (n,k,d) code. Then for any $0 < p \leq 1/2$ we have

$$\mathrm{NS}_p(F; \mathcal{C}) \leq O\left(p \cdot \frac{n}{d} \log^{t-1} M\right).$$

The proof of Lemma 2.3 uses the low noise-sensitivity of AC^0 circuits Boppana (1997); Linial *et al.* (1993). We use the following version, given explicitly in (Viola 2004, Lemma 6.6).

LEMMA 2.7. *Let* $f : \{0,1\}^m \to \{0,1\}$ *be an* AC^0 *circuit of depth* t *and size* M. *Then*

$$\Pr_{x \in U_m, e \in \mu_p} [f(x) \neq f(x+e)] \leq O\left(p \cdot \log^{t-1} M\right).$$

We now prove Lemma 2.3.

PROOF (Proof of Lemma 2.3).   The proof will follow by analysis of the average distance between $F(x)$ and $F(x+e)$.   Let $F = (f_1, \ldots, f_n)$ where each $f_i : \{0,1\}^m \to \{0,1\}$ is an AC^0 function of depth t and size at most M.   By Lemma 2.7 we know that $\Pr_{x,e}[f_i(x) \neq f_i(x+e)] \leq O\left(p \cdot \log^{t-1} M\right)$.   Since $\mathbb{E}_{x,e}[\mathrm{dist}(F(x), F(x+e)) = \sum_{i=1}^{n} \Pr_{x,e}[f_i(x) \neq f_i(x+e)]$ we deduce that

$$\mathbb{E}_{x,e}[\mathrm{dist}(F(x), F(x+e))] \leq O(n \cdot p \cdot \log^{t-1} M).$$

On the other hand, as $\mathcal{C}$ is a code with minimal distance d, whenever $F(x), F(x+e) \in \mathcal{C}$ such that $F(x) \neq F(x+e)$ we must have $\mathrm{dist}(F(x), F(x+e)) \geq d$. Hence we get that

$$\mathbb{E}_{x,e}[\mathrm{dist}(F(x), F(x+e))]$$
$$\geq d \cdot \Pr_{x,e}[F(x) \in \mathcal{C}, F(x+e) \in \mathcal{C}, F(x) \neq F(x+e)].$$

Thus we deduce that

$$\Pr_{x,e}[F(x) \in \mathcal{C}, F(x+e) \in \mathcal{C}, F(x) \neq F(x+e)] \leq O\left(p \cdot \frac{n}{d} \log^{t-1} M\right).$$

$\square$

3. Complexity of Nisan's generator against AC^0

We prove Theorem 1.3 in this section, which we restate for the convenience of the reader.

Theorem 1.3, restated. *Let* $\ell = \ell(n)$ *and* $k = k(n)$ *be functions such that* $k \geq 4\ell^2, n = \omega(k^3)$, *and* $\ell(n) \geq \log(n)$ *is odd.*
For arbitrarily large n, *there is an* $n \times k$ *matrix such that:*

(1) *M forms a design: each row of M has hamming weight ℓ, and any two rows share at most $\log n$ ones, and*

(2) *Any AC^0 circuit of size s and depth c whose output distribution (over uniform input) has statistical distance less than $1/2$ from Mx (over uniform $x \in \{0,1\}^k$) satisfies $\log^{c-1} s = \Omega(\ell)$.*

A natural approach is to choose each row of M to be a random string with ℓ ones. However, we find it easier to analyze a different, block-wise construction.

PROOF.   To construct the matrix M, divide $[k]$ into ℓ blocks of size k/ℓ each. In order to construct a row of M, with probability $1/2$ independently choose one bit from every block, and with probability $1/2$ shift all the blocks by 1 to the right rolling over (so the last bit of the last block is the first bit of the row) and again independently choose one bit from every block. This "trick" of shifting is useful when arguing that the matrix generates a good enough code.

Do this independently across rows. We show that each of (1) and (2) holds with probability $> 1/2$, hence there exists a matrix as claimed.

(1) The hamming weight of the rows is ℓ by construction. To analyze the intersection size, consider any two rows r and r'. Fix arbitrarily r', and also fix arbitrarily the choice of whether or not to shift the blocks of r by 1. Note that each block of r intersects at most 2 blocks of r'. Hence for every block i of r, the probability that the choice of the bit in the i-th block of r overlaps a bit of r' is $\leq 2\ell/k$. Consequently, the probability that r and r' share more than $\log n$ ones is at most

$$\binom{\ell}{\log n}(2\ell/k)^{\log n} \leq (\ell^2/k)^{\log n} \leq 1/n^2.$$

Hence the probability that there exist two rows sharing more than $\log n$ ones is at most $\binom{n}{2}1/n^2 < 1/2$.

(2) We show that with probability $> 1/2$ the matrix M is the generator matrix of a code with "good" parameters, and then apply Theorem 1.1. M corresponds to a code with block-length n and

message-length k. We now analyze the distance. Since the code is linear, it is sufficient to bound from below the hamming weight of any non-zero codeword, which we accomplish by bounding each fixed codeword and then applying a union bound.

First we claim that for any fixed nonzero $x \in \{0,1\}^k$ and row index, the probability (over the bits in that row) that the inner product between x and that row is 1 is at least

$$p := 0.5\ell/k.$$

If $x = 1^k$, then $Mx = 1^n$ since ℓ is odd, with probability 1. Fix any $x \notin \{0^k, 1^k\}$. With probability $\geq 1/2$ over the choice of whether or not to shift the blocks of the row by 1, there is a block of k/ℓ bits of x with both a 0 and a 1. Consider the inner product between the row and x. Whatever the choice for the row in the other blocks, the choice in this block guarantees that this inner product is 1 with probability at least ℓ/k. This establishes the claim.

Thus, Mx has expected hamming weight pn. By a standard Chernoff bound, the probability that Mx has hamming weight less than $(p/2)n$ is at most

$$e^{-2(p/2)^2 n} \leq 2^{-\Omega(\ell^2/k^2)n} < 2^{-\Omega(n/k^2)} < (1/2)2^k$$

using that $n = \omega(k^3)$. By a union bound, with probability bigger than $1/2$ it holds that Mx has hamming weight at least $(p/2)n$ for every non-zero x. This means that M generates a code with hamming distance $\geq (p/2)n = 0.25\ell n/k$. By Theorem 1.1, any circuit of depth c and size s has an output distribution (over uniform input) whose statistical distance from the distribution Mx (for uniform $x \in \{0,1\}^k$) is $\geq 1 - \epsilon$ for

$$\epsilon = O\left(\frac{n}{(0.25\ell n/k)k}\log^{c-1} s\right)^{1/3} = O\left(\frac{\log^{c-1} s}{\ell}\right)^{1/3}.$$

If one wants $\epsilon \geq 1/2$ then $\log^{c-1} s = \Omega(\ell)$, concluding the proof. $\square$

Not every matrix M corresponding to a design is the generator matrix of a "good" code, e.g. let one column of M be 0. However it may be possible that every matrix M corresponding to a design

contains as a submatrix a "good" code. This would generalize our results showing that the lower bound applies regardless of the choice of the design.

4. Open problems

In this section we discuss a possible way to attack the problem of exhibiting an explicit boolean function b such that AC^0 cannot generate the distribution $(x, b(x))$. Let b be the n-bit Majority function, for n odd. As shown in Viola (2010), there are small AC^0 circuits that generate $(x, b(x))$ with exponentially small error and using $\geq n \log n$ input random bits. We discuss a possible way to show that small AC^0 circuits cannot generate $(x, b(x))$ with error 0 (i.e., exactly) and using n random bits, which is open.

It is easy to see (see Viola (2010)) that any, say, AC^0 circuit $C : \{0,1\}^n \to \{0,1\}^{n+1}$ whose output distribution equals $(X, b(X))$ can be transformed into an AC^0 circuit $C' : \{0,1\}^n \to \{0,1\}^n$ that generates the distribution A over n-bit strings whose hamming weight is $\geq n/2$. We would like to show that the latter is impossible. For simplicity, let us start with the simpler setting in which the input length to C' is $n-1$ (as opposed to n) which is also open (hopefully a solution to this case can be lifted to a solution for input length n).

In other words, we are trying to rule out that there exists an easily computable (say AC^0) bijection from the hamming cube $\{0,1\}^{n-1}$ into the "upper half" of the hamming cube $\{0,1\}^n$. Using, like in this paper, the low noise-sensitivity of AC^0, what stands in the way of a lower bound is a (positive) solution to the following seemingly new and interesting open problem:

OPEN PROBLEM 4.1. *Prove that any bijection $f : \{0,1\}^{n-1} \to \{x \in \{0,1\}^n : \sum x_i \geq n/2\}$, n odd, has high average distortion: the expected hamming distance D between $f(x)$ and $f(y)$ for uniform (x, y) at hamming distance 1 is $D \geq \log^{\omega(1)} n$.*

Even a weaker, $\omega(1)$ lower bound would be interesting and would have consequences for NC^0. In general, proving lower bounds on the distortion necessary to embed hamming cubes into

various subsets of larger hamming cubes seems an interesting approach to prove lower bounds for generating distributions, and an approach that could leverage from the existing body of knowledge on embeddings.

5. Proof of noise sensitivity isoperimetric inequality

We prove in this section Lemma 2.4, restated next.

Lemma 2.4, restated. *Let $A \subseteq \{0,1\}^m$ and $\alpha := |A|/2^m$. Then for any $0 \le p \le 1/2$ we have*

$$\alpha^2 \le \Pr_{x \in U_m, e \in \mu_p}[x \in A, x + e \in A] \le \alpha^{1/(1-p)} \le \alpha^{1+p}.$$

The third inequality is obvious. We now prove the first.

We can view $e \in \mu_p$ as the bit-wise xor of $e', e'' \in \mu_{p'}$ if $p = 2p'(1 - p')$. Fix such a p'. Thus, the joint distribution $(x, x + e)_{x \in U_n, e \in \mu_p}$ is equivalent to the joint distribution $(x + e', x + e'')_{x \in U_n, e', e'' \in \mu_{p'}}$. Let $\mathbf{1}_A : \{0,1\}^m \to \{0,1\}$ denote the indicator function for A. We thus have

$$\Pr_{x \in U_m, e \in \mu_p}[x \in A, x + e \in A]$$

$$= \Pr_{x \in U_m, e', e'' \in \mu_{p'}}[x + e' \in A, x + e'' \in A]$$

$$= \mathbb{E}_{x \in U_m, e', e'' \in \mu_{p'}} \mathbf{1}_A(x + e') \mathbf{1}_A(x + e'')$$

$$= \mathbb{E}_{x \in U_m} \left(\mathbb{E}_{e' \in \mu_{p'}} \mathbf{1}_A(x + e') \right)^2$$

$$\ge \left(\mathbb{E}_{x \in U_m, e' \in \mu_{p'}} \mathbf{1}_A(x + e') \right)^2$$

$$= \left(\mathbb{E}_{x \in U_m} \mathbf{1}_A(x) \right)^2$$

$$= \left(\frac{|A|}{2^m} \right)^2,$$

where the inequality follows from the Cauchy-Schwarz inequality.

We now prove the second inequality.

Let $f : \{0,1\}^m \to \mathbb{R}$ be a function. The p-noise sensitivity of f is given by

$$\mathrm{NS}_p(f) := \mathbb{E}_{x \in U_m, e \in \mu_p}[f(x)f(x + e)].$$

We are interested in the p-noise sensitivity of the indicator function of A, $\mathbf{1}_A(x) := 1_{x \in A}$.

One may define a *noise operator* as follows.

DEFINITION 5.1. *Let* $f : \{0,1\}^m \to \mathbb{R}$ *and* $0 \le p \le 1/2$. *Then* $T_{1-2p}f(x) = \mathbb{E}_{e \in \mu_p} f(x + e)$.

We define an inner product between two functions as

$$\langle f, g \rangle = \mathbb{E}_{x \in U_m} f(x)g(x).$$

Also, for $1 \le q \le \infty$, the L_q norm of f is

$$\|f\|_q = (\mathbb{E}_{x \in U_m}[|f(x)|^q])^{1/q}.$$

Noise sensitivity can be described as the L_2 norm of the noise operator (this is proved below in Fact 5.3). Thus, to study the noise sensitivity of f is equivalent to studying the L_2 norm of $T_\rho f$. The following hypercontractivity theorem relates the L_2 norm of $T_\rho f$ to norms of f, cf. (O'Donnell 2007, Lecture 16).

THEOREM 5.2 (Hypercontractivity). *Let* $f : \{0,1\}^m \to \mathbb{R}$. *Then for any* $0 \le \rho \le 1$ *we have* $\|T_\rho f\|_2 \le \|f\|_{1+\rho^2}$.

We can now immediately derive a somewhat weaker bound than that guaranteed by Lemma 2.4 by applying the Cauchy-Schwarz inequality $\langle f, g \rangle \le \|f\|_2 \cdot \|g\|_2$.

$$\begin{aligned}
\mathrm{NS}_p(\mathbf{1}_A) &= \langle \mathbf{1}_A, T_{1-2p}\mathbf{1}_A \rangle && \text{(follows from the definition)} \\
&\le \|\mathbf{1}_A\|_2 \cdot \|T_{1-2p}\mathbf{1}_A\|_2 && \text{(Cauchy-Schwarz)} \\
&\le \|\mathbf{1}_A\|_2 \|\mathbf{1}_A\|_{1+(1-2p)^2} && \text{(hypercontractivity)} \\
&= \left(\frac{|A|}{2^n}\right)^{\frac{1}{2} + \frac{1}{1+(1-2p)^2}} && (\mathbf{1}_A \text{ is a } \{0,1\} \text{ function}) \\
&= \left(\frac{|A|}{2^n}\right)^{\frac{1-p+p^2}{1-2p+2p^2}}.
\end{aligned}$$

This bound is sufficient for our applications.

To derive the better bound in Lemma 2.4 we use the following.

FACT 5.3. *Let $f : \{0,1\}^m \to \mathbb{R}$. Then $\mathrm{NS}_p(f) = \|T_{\sqrt{1-2p}}f\|_2^2$.*

PROOF.   We have:

$$
\begin{aligned}
\|T_{\sqrt{1-2p}}f\|_2^2 &= \mathbb{E}_x \left(T_{\sqrt{1-2p}}f(x)\right)^2 = \mathbb{E}_x \left(\mathbb{E}_{e\in\mu_{\frac{1-\sqrt{1-2p}}{2}}}f(x+e)\right)^2 \\
&= \mathbb{E}_x \mathbb{E}_{e,e'\in\mu_{\frac{1-\sqrt{1-2p}}{2}}}f(x+e)f(x+e') \\
&= \mathbb{E}_x \mathbb{E}_{e,e'\in\mu_{\frac{1-\sqrt{1-2p}}{2}}}f(x)f(x+e+e') \\
&= \mathbb{E}_x \mathbb{E}_{e\in\mu_p}f(x)f(x+e) = \mathrm{NS}_p(f).
\end{aligned}
$$

$\square$

We can now prove the second inequality in Lemma 2.4. We have

$$
\begin{aligned}
\mathrm{NS}_p(\mathbf{1}_A) = \|T_{\sqrt{1-2p}}\mathbf{1}_A\|_2^2 &\leq \|\mathbf{1}_A\|_{2(1-p)}^2 \qquad\qquad \text{(hypercontractivity)} \\
&= \left(\mathbb{E}_{x\in U_m}[|\mathbf{1}_A(x)|^{2(1-p)}]\right)^{1/(1-p)} \\
&= \left(\mathbb{E}_{x\in U_m}[\mathbf{1}_A(x)]\right)^{1/(1-p)} \qquad (\mathbf{1}_A \text{ is a } \{0,1\} \text{ function}) \\
&= \left(\frac{|A|}{2^m}\right)^{1/(1-p)}.
\end{aligned}
$$

This concludes the proof of Lemma 2.4.

Acknowledgements

References

ANDRIS AMBAINIS, LEONARD J. SCHULMAN, AMNON TA-SHMA, UMESH V. VAZIRANI & AVI WIGDERSON (2003). The Quantum Com-

munication Complexity of Sampling. *SIAM J. Comput.* **32**(6), 1570–1585.

LOUAY M. J. BAZZI (2009). Polylogarithmic Independence Can Fool DNF Formulas. *SIAM J. Comput.* **38**(6), 2220–2272.

RAVI BOPPANA (1997). The average sensitivity of bounded-depth circuits. *Information Processing Letters* **63**(5), 257–261. ISSN 0020-0190.

MARK BRAVERMAN (2010). Polylogarithmic independence fools AC^0 circuits. *J. of the ACM* **57**(5).

ANNA GÁL & PETER BRO MILTERSEN (2007). The cell probe complexity of succinct data structures. *Theoret. Comput. Sci.* **379**(3), 405–417. ISSN 0304-3975.

ODED GOLDREICH, SHAFI GOLDWASSER & ASAF NUSSBOIM (2010). On the Implementation of Huge Random Objects. *SIAM J. Comput.* **39**(7), 2761–2822.

VENKATESAN GURUSWAMI, CHRISTOPHER UMANS & SALIL P. VADHAN (2009). Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. of the ACM* **56**(4).

L. H. HARPER (1964). Optimal Assignments of Numbers to Vertices. *SIAM Journal on Applied Mathematics* **12**(1), 131–135.

SERGIU HART (1976). A note on the edges of the n-cube. *Discrete Mathematics* **14**(2), 157–163.

JOHAN HÅSTAD (1987). *Computational limitations of small-depth circuits.* MIT Press. ISBN 0262081679.

MARC R. JERRUM, LESLIE G. VALIANT & VIJAY V. VAZIRANI (1986). Random Generation of Combinatorial Structures from a Uniform Distribution. *Theoretical Computer Science* **43**(2–3), 169–188.

NATHAN LINIAL, YISHAY MANSOUR & NOAM NISAN (1993). Constant depth circuits, Fourier transform, and learnability. *J. of the ACM* **40**(3), 607–620. ISSN 0004-5411.

SHACHAR LOVETT & EMANUELE VIOLA (2011). Bounded-depth circuits cannot sample good codes. In *IEEE Conf. on Computational Complexity (CCC)*.

PETER BRO MILTERSEN (1999). Cell probe complexity - a survey. Invited talk/paper at Advances in Data Structures (Pre-conference workshop of FSTTCS'99).

NOAM NISAN (1991). Pseudorandom bits for constant depth circuits. *Combinatorica* **11**(1), 63–70. ISSN 0209-9683.

RYAN O'DONNELL (2007). Analysis of Boolean Functions. Lecture notes. Available at http://www.cs.cmu.edu/ odonnell/boolean-analysis/.

ALEXANDER A. RAZBOROV (2009). A Simple Proof of Bazzi's Theorem. *ACM Transactions on Computation Theory (TOCT)* **1**(1).

EMANUELE VIOLA (2004). The Complexity of Constructing Pseudorandom Generators from Hard Functions. *Computational Complexity* **13**(3-4), 147–188.

EMANUELE VIOLA (2010). The complexity of distributions. *SIAM J. on Computing* To appear.

EMANUELE VIOLA (2011). Extractors for circuit sources. In *IEEE Symp. on Foundations of Computer Science (FOCS)*.

SHACHAR LOVETT   EMANUELE VIOLA
The Institute of Advanced Study   Northeastern University
Princeton, NJ   Boston, MA
slovett@math.ias.edu   viola@ccs.neu.edu