

Top-Down Lower Bounds for All Depths

Oliver Korten *

Abstract

We prove that Parity requires $2^{n^{\Omega(1)}}$ size De Morgan circuits of constant depth using a new method which is completely “top-down” in the sense of [HJP95]. The proof relies crucially on the core ideas developed in a line of work [HJP95, PPZ99, MW19, GRSS24] which previously established top-down lower bounds for circuits of depth 3 and 4. We first present a proof of a lower bound $\exp(n^{3^{-d}})$. In this case, nearly all of the relevant combinatorial ideas necessary for the proof are already present in some form in [GRSS24]. We then present two extensions of this argument, the first achieving a lower bound $\exp(\epsilon_d n^{1/(2d-2)})$ for some $\epsilon_d > 0$ depending only on d , and the second achieving the essentially tight lower bound $\exp(\epsilon_d n^{1/(d-1)})$. These improved results each hinge on establishing a key lemma which quantifies the extent to which a high entropy random variable in $\{0,1\}^n$ will look close to uniform after projecting it onto a random small set of coordinates $R \subseteq [n]$.

AI Usage: The result presented in Section 2, giving the first top-down lower bounds for arbitrary depth $\mathbf{AC}^0$ circuits, was generated autonomously by GPT-6 Astra. This initial proof gave a lower bound of the form $\exp(n^{3^{-d}})$ for depth d circuits. Subsequently, the author extended these methods to give a top-down proof of the stronger lower bound $\exp(n^{1/(2d-2)})$; this human-generated result is presented in Section 3. The author then determined a combinatorial conjecture which would suffice to extend the arguments in Section 3 to the optimal $\exp(n^{1/(d-1)})$ bound, and asked the machine to prove it; the machine succeeded with some high-level human direction and the final tight lower bound is presented in Section 4. A more detailed description of the AI methodology is given at the end of the introduction (1.4). All contents of this document were written solely by the human author, who takes full responsibility for their correctness.

1 Introduction

A highly influential work of Karchmer and Wigderson [KW90] showed that lower bounds on the depth of circuits computing a function f can be rephrased equivalently as lower bounds on the *communication complexity* of a certain search problem associated with the function f now called the “KW game for f .” A variant of this correspondence was shown to hold also for *monotone computation*, and was immediately applied in [KW90] to obtain tight lower bounds on the depth of monotone circuits deciding undirected connectivity. Since [KW90], a rich theory centered around *lifting* has developed which supplies a very modular and general recipe for proving communication lower bounds for monotone KW games, and hence depth lower bounds for monotone circuits (the framework also yields strong lower bounds on monotone circuit size) [RM99, dRNV16, PR17, GP18, GPW18, GKRS19, GGKS20, dRMNPRV20, LMMPZ22, dRV25]. In this framework, a lower bound is first proven for the monotone KW game in a simplified “decision tree model,” and then that simple lower bound is “lifted” to the more general communication-based model.

*Institute for Advanced Study. Email: korten@ias.edu. This material is based upon work supported by the National Science Foundation under Grant No. DMS-2424441.

It has been a longstanding open problem to apply communication based arguments to obtain interesting lower bounds for *non-monotone* circuits. In particular, it was pointed out in [HJP95] that at the time we did not even know how to reproduce *known non-monotone lower bounds* in this framework. The authors of [HJP95] initiated the project of reproving the celebrated result of [FSS84, Ajt83] that Parity $\notin \mathbf{AC}^0$ via communication-complexity arguments, which they refer to as “top-down proofs” (we will explain the meaning of “top-down” later on in the introduction). They proceeded to solve this problem in the particular case of depth 3 circuits. A few works have continued to study this question in the ensuing decades [PPZ99, MW19, GRSS24]; in particular, exponential lower bounds for depth 4 were obtained in [GRSS24]. The problem has remained open for any depth larger than 4. The current work completes this program, establishing exponential lower bounds for $\mathbf{AC}^0$ circuits of any fixed depth computing parity by a proof which is completely top-down in the sense intended in [HJP95].

The value in proving an old result by a new method can be hard to judge on arrival. Our hope, and the hope of those who’ve previously studied this problem [HJP95, MW19, GRSS24], is that such a proof will offer a bridge over which ideas from the highly-developed theory of monotone circuit complexity may start to be transferred over to the study of non-monotone circuit complexity, perhaps in the eventual service of making progress on the KRW conjecture [KRW95] and $\mathbf{P} \neq \mathbf{NC}^1$. Another motivation, highlighted in [GRSS24], is the well-known problem of proving lower bounds for $\mathbf{PH}^{\text{cc}}$; this is a class extending $\mathbf{AC}^0$ for which random-restriction based arguments appear completely unavailable, and the authors in [GRSS24] argue that a top-down lower bound for $\mathbf{AC}^0$ may point us in the direction of such a lower bound. Currently we have no clue as to whether the ideas in this work will play any role in future lower bounds of these kinds. At the very least, we believe the arguments here will offer some richer understanding of bounded depth circuits beyond what can be achieved by the method of random restrictions [FSS84, Ajt83, Yao85, Has86] or the method of polynomial approximation [Raz87, Smo87].

1.1 KW Games and Adversaries:

We start by defining Karchmer-Wigderson communication problems. For a pair of sets $X, Y \subseteq \{0, 1\}^n$, $X \cap Y = \emptyset$, the communication problem $\text{KW}(X, Y)$ is defined as follows. Two parties Alice and Bob receive inputs $x \in X$ and $y \in Y$ respectively. Their goal is to determine some $i \in [n]$ such that $x_i \neq y_i$, by communicating with one another over the course of several rounds. We say that $\text{KW}(X, Y)$ has a d -round protocol of cost m , if there is a deterministic protocol in which the players exchange messages for at most d rounds, sending at most m bits in every round, such that at the end of the protocol both parties agree on a valid output for $\text{KW}(X, Y)$. In a given round, the next speaker is determined arbitrarily by the prior communication transcript. For a Boolean function f , we use the shorthand $\text{KW}(f) := \text{KW}(f^{-1}(0), f^{-1}(1))$.

Lemma 1 ([KW90]). *A function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ has a depth $\leq d$ unbounded fan-in De Morgan circuit in which each $\vee / \wedge$ node has $\leq 2^m$ children if and only if $\text{KW}(f)$ has a d -round protocol of cost m .*

As is well known, a communication protocol for a problem in which Alice receives inputs in X , and Bob in Y , corresponds to a recursive decomposition of $X \times Y$ into rectangles $X' \times Y'$, $X' \subseteq X, Y' \subseteq Y$ (see [KN97] for a comprehensive textbook on communication complexity). In a given round i of a d -round protocol for $\text{KW}(X, Y)$, the set of inputs consistent with the current communication transcript is a rectangle $X' \times Y'$. If the next player to speak is Alice, her m -bit message performs an arbitrary partition of X' into at most 2^m parts $X' = \bigcup_{j \leq 2^m} X'_j$. The remaining protocol can then be seen as giving, separately for each subrectangle $X'_j \times Y'$, a $(d-i)$ -round protocol

of cost m for $\text{KW}(X'_j, Y')$. The existence of a zero-round protocol for some X', Y' is equivalent to the existence of some $i \in [n]$ such that $x_i \neq y_i$ for all $x \in X', y \in Y'$.

With this perspective in mind, it is possible to prove that no d -round, cost m protocol exists for $\text{KW}(X, Y)$ by a kind of *adversary argument*. Say that we define some combinatorial invariants $I_1, \dots, I_{d+1}$ on rectangles in $X \times Y$, where invariant I_i is meant to be satisfied prior to the i^{th} round of communication, and I_{d+1} to be satisfied after all communication is complete. If we can prove:

1. (Base Case): The initial rectangle $X \times Y$ satisfies the initial round 1 invariant I_1 .
2. (Induction Step): If $X' \times Y'$ is any subrectangle satisfying the invariant I_i for round i , and we partition one of the sides into at most 2^m parts e.g. $X' = \bigcup_{j \leq 2^m} X'_j$, then there exists some part X'_j such that $X'_j \times Y'$ satisfies the invariant I_{i+1} for round $i + 1$.
3. (Terminal Case): If $X' \times Y'$ satisfies the final invariant I_{d+1} , then for all $i \in [n]$ there exists $x \in X', y \in Y'$ such that $x_i = y_i$.

then it follows that $\text{KW}(X, Y)$ does not have a d -round protocol of cost m . We refer to this as an adversary argument, since we imagine the choice of part X'_j in each inductive step as being selected by an adversary aiming to focus attention on the inputs in which the protocol fails to make sufficient progress.

1.2 What Exactly Constitutes a “Top-Down” Proof?

A “top-down proof” of a lower bound refers, informally speaking, to any communication lower bound for $\text{KW}(f)$ which is proven by giving an *explicit communication adversary* for $\text{KW}(f)$. This is called top-down since tracing the evolution of a rectangle across the communication protocol corresponds precisely to tracing the set of inputs rejected/accepted by a gate in the circuit, starting from the output gate and following a downward path to some leaf computing a literal. This is in contrast to a “bottom-up” proof such as the random-restriction method, which eliminates gates of the circuit starting from the bottom and continuing upwards towards the output.

The substance of the term “top-down proof” rests entirely on what we mean by an “explicit communication adversary.” Indeed, the KW correspondence in Lemma 1 is an exact equivalence, so *any proof whatsoever* of $\text{Parity} \notin \mathbf{AC}^0$ must in fact yield a lower bound on the communication complexity of $\text{KW}(\text{Parity})$, and hence implies the existence of *some* adversary. However, if one takes any of the previously known proofs of $\text{Parity} \notin \mathbf{AC}^0$ and works backwards through the correspondence, they will end up building an adversary whose next move is guaranteed only by a counterfactual, nonconstructive argument: assuming no next move for the adversary exists, we work backwards through the KW correspondence to deduce the existence of a circuit computing the function and contradict the known circuit lower bound. The reader may refer to [KR26] for an interesting analysis of such a “non-explicit” adversary arising from the approximation method¹. In contrast, our proof maintains a simple combinatorial invariant on the current rectangle, and directly supplies the adversary’s next move based on this invariant.

1.3 Notation:

We introduce most notation as needed but start with some basic conventions that will be used throughout. All logarithms are base 2. For a finite coordinate set I and $p \in [0, 1]$, we use $P \subseteq_p I$

¹The work in [KR26] does not merely apply the KW correspondence generically to obtain an adversary in the way we describe, but proves that the approximation method yields an adversary with some particular features of interest. Nonetheless the adversary and its analysis remain non-constructive in roughly the same sense discussed here.

to denote a random $P \subseteq I$ which is sampled by including each $i \in I$ in P independently with probability p . For a finite set X , we use $x \sim X$ to denote a uniformly random sample from X . For a random variable $\mathbf{X}$ with finite support, we use $H(\mathbf{X})$, $H_\infty(\mathbf{X})$ to refer to its Shannon entropy and min-entropy respectively. If $\mathbf{X}$ is supported on $\{0, 1\}^n$, we use $D_\infty(\mathbf{X}) := n - H_\infty(\mathbf{X})$ to refer to its *min-entropy deficit*. For $X \subseteq \{0, 1\}^n$, we use $D_\infty(X) := n - \log |X|$; this is exactly the min-entropy deficit of the random variable $\mathbf{X}$ which is uniform on X , and hence we refer to this quantity also as the min-entropy deficit of the set X itself.

1.4 Roadmap and AI Methodology

In Sections 2, 3, and 4, we present top-down depth- d circuit lower bounds of increasing quantitative strength: first $\exp(n^{3^{-d}})$, then $\exp(\epsilon_d n^{1/(2d-2)})$, and finally $\exp(\epsilon_d n^{1/(d-1)})$ which is tight up to the constant $\epsilon_d > 0$ (which in [Has86] can be replaced by an absolute constant $\epsilon > 0$). We describe here in a bit more detail the respective roles of the author and the machine assistant over the course of this work.

Initially, the result in Section 2 was obtained completely autonomously by GPT-6 Astra. The author made some substantial effort to reorganize the proof in an intelligible way (the most substantive change being the introduction of the guiding function and the random variable $\mathbf{Y}^g$ to construct the mirror set, in place of a more inscrutable Markov argument in the original proof) but the underlying technical ingredients are essentially the same here as they appeared in the original machine-generated proof. In fact, the original proof found by Astra gave a lower bound of $\Omega_d(n^{1/F_d})$ where F_d is the d^{th} Fibonacci number (F_d grows as $\Theta(\alpha^d)$ for some constant $\alpha \approx 1.6$); we chose to simplify the bound to $n^{3^{-d}}$ for the sake of a more straightforward exposition.

Subsequent to writing up the machine-generated results in Section 2, the author found a way to extend these methods to obtain the improved lower bound in Section 3; the introduction of the notion of (p, k) -limit, the reduction to the “light patterns lemma,” and the proof of the first form of the light patterns lemma via Fourier analysis were found by the author. GPT-6 Astra was used at various points to optimize parameters and simplify the presentation of the proof.

After the work in Section 3, the author had determined that a sufficient strengthening of the “light patterns lemma” would be sufficient to obtain the near-optimal $\exp(\epsilon_d n^{1/(d-1)})$ lower bound using the same proof structure introduced by the author in Section 3. With some high level steering by the author, the machine ultimately found a proof of this strengthened lemma. The author’s contribution here was to suggest reusing the “transference principle” for downward closed sets (applied originally in Section 2 to prove the shattering lemma) and then to look for a more analytic analogue of the inductive proof of Pajor’s lemma. Eventually the machine found a proof of the light patterns lemma which involved deriving a specialized reverse-hypercontractive inequality for decreasing functions on the $1/3$ -biased hypercube. The author found this original proof to be conceptually opaque and worked for some time with the machine to simplify it, eventually arriving at the current argument in Section 4 involving the “harmonic mean transform.”

2 The Initial Lower Bound

In this section we present the initial proof generated by GPT-6 Astra of the lower bound $m \geq n^{3^{-d}}$ on the communication cost of d -round protocols for $\text{KW}(X, Y)$ where X, Y are the 0 and 1-inputs of the parity function on n bits.

2.1 The Adversary

We start by describing the adversary and reduce its correctness to a single key combinatorial lemma (the “mirror set lemma”). The following subsection is dedicated to a proof of the mirror set lemma.

Theorem 1. *For any $n, d, m \in \mathbb{N}$, $\text{KW}(\text{Parity}_n)$ does not have a d -round communication protocol of cost m when $m \leq n^{3^{-d}}$, $n \geq 64^{3^d}$. Consequently any depth d circuit computing Parity_n must have $> 2^{n^{3^{-d}}}$ wires whenever $n \geq 64^{3^d}$.*

The proof will maintain a certain invariant on the rectangle $X \times Y$ which intuitively says it is hard to tell the two sets apart by looking at a random small set of coordinates. We make a note here that in the generic description of a communication adversary given in Subsection 1.1, we may without loss of generality allow the adversary to restrict a selected rectangle $X_i \times Y$ to some nonempty subrectangle $X'_i \times Y$, where $X'_i \subseteq X_i$ (and symmetrically on the Y side), and maintain some invariant defined by that subrectangle. We will use this convention as a matter of notational convenience. The crux of the proof is to show that, if the current rectangle satisfies the invariant and the next protocol message has small cost, then the adversary can select a subrectangle which satisfies a quantitatively weaker version of the invariant. In the following, for $x \in \{0, 1\}^n$, $P \subseteq [n]$ we use $[x]_P$ to denote the subcube $\{z \in \{0, 1\}^n \mid z_i = x_i \ \forall i \notin P\}$ and $x_P \in \{0, 1\}^P$ to denote the projection of x onto the coordinates in P .

Definition 1 (p -limit Condition). *Say that $x \in \{0, 1\}^n$ is a p -limit of $Y \subseteq \{0, 1\}^n$ if*

$$\Pr_{P \subseteq_p [n]} [Y \cap [x]_P \neq \emptyset] \geq \frac{3}{4}$$

For sets $X, Y \subseteq \{0, 1\}^n$, we say that the rectangle $X \times Y$ satisfies the p -limit condition on the left (resp. right) if every element $x \in X$ is a p -limit of Y (resp. every $y \in Y$ is a p -limit of X).

We note that the threshold $\frac{3}{4}$ is somewhat arbitrary; changing the threshold to any constant in $(0, 1)$ would suffice for the main result. The notion of a “ p -limit” is a minor variant of a definition used in [GRSS24] (see Section 2.2, equation (2) of [GRSS24]; their definition involves a uniform random k -set rather than an independent Bernoulli set). This in turn is a variant of the “ k -limit” used earlier in [HJP95] which requires the stated condition to hold *for all sets* P of a given size; the authors of [HJP95] credit Sipser [Sip84] with the introduction of the concept.

The invariant we will maintain on the rectangle $X \times Y$ is that (1) both X and Y are suitably large and (2) the rectangle satisfies either the left or right p -limit condition (for some largeness thresholds and p values which decay as the rounds progress). As mentioned above, the key step is to show that after each protocol message, the adversary can select a subrectangle satisfying the next invariant. We make a few observations. First note that it is trivial to maintain condition (1), losing a factor 2^{-m} in largeness at each step, by choosing the largest component of the decomposition. Second, note that if the p -limit condition is satisfied on the left (resp. right) and the next message partitions X (resp. Y) then the p -limit condition is trivially maintained, since it involves a universal quantifier over elements of X (resp. Y). Hence, if it were possible to show that the left p -limit condition implies the right q -limit condition (possibly after passing to subsets of X, Y) and vice versa for some reasonable value of q , we would be done. This is exactly what the following “mirror set lemma” will accomplish.

Lemma 2 (Mirror Set Lemma). *Suppose $D_\infty(X) \leq k$, where $k \geq 1$, $p > 0$, and the rectangle $X \times Y$ satisfies the p -limit condition on the left. If $q = 64kp \leq 1$ then there exists a “mirror set” $Y' \subseteq Y$ such that the rectangle $X \times Y'$ satisfies the q -limit condition on the right, and $D_\infty(Y') \leq k + 2pn + 2$.*

Note that the statement of this lemma immediately implies the symmetric form in which we swap the roles of X, Y , and so we use both forms freely. We are now ready to prove Theorem 1:

Proof of Theorem 1. Let $X_0 = \text{Parity}^{-1}(0)$, $Y_0 = \text{Parity}^{-1}(1)$. For $i \in \{0, \dots, d\}$ let $k_i = n^{3^{i-d}}$, $p_i = \frac{k_i}{4n}$. We observe the following:

$$64 \leq k_0 \leq k_i \leq n \text{ for all } 0 \leq i \leq d, \quad m \leq k_0, \quad k_{i+1} = k_i^3 \text{ for all } 0 \leq i < d$$

and from these it follows that, for all $0 \leq i < d$

$$1. \quad k_i + 2p_i n + 2 + m \leq 3k_i + 2 \leq k_{i+1}, \quad 2. \quad 64k_i p_i \leq \frac{k_i^3}{4n} \leq p_{i+1} \leq \frac{1}{4}$$

After i rounds we maintain a rectangle $X_i \times Y_i$ such that the p_i -limit condition is satisfied (either on the left or right) and $D_\infty(X_i), D_\infty(Y_i) \leq k_i$. At the start, $k_0 = n^{3^{-d}} \geq 1 = D_\infty(X_0) = D_\infty(Y_0)$ so the deficit condition is satisfied. On the other hand it may be verified that for our parameters, a p_0 -random set P has $|P| > 0$ with probability $\geq \frac{3}{4}$; this immediately implies that the p_0 -limit property is satisfied on both the left and right by the rectangle $X_0 \times Y_0$ since for any string z and any $|P| > 0$, $[z]_P$ will contain both even and odd parity strings. At the end we have $p_d \leq \frac{1}{4}$ and $k_d \leq n$, which implies that the final rectangle $X_d \times Y_d$ does not admit a zero-round protocol: in this case the p_d -limit is satisfied on the left or right (say left) and both X_d, Y_d are nonempty. For every i , a p_d -random set will contain i with probability $\leq \frac{1}{4}$, and hence for any $x \in X_d$ and every $i \leq n$ we can find some $y \in Y_d$ agreeing with x at index i .

It remains to show that if the rectangle $X_i \times Y_i$ satisfies the key invariant then we can reestablish this for $X_{i+1} \times Y_{i+1}$. Suppose the next message partitions $X_i = X_{i,1} \cup \dots \cup X_{i,2^m}$ (the other case is handled symmetrically). If the p_i -limit condition holds for $X_i \times Y_i$ on the left then it trivially holds for every subrectangle $X_{i,j} \times Y_i$ (and hence so does the p_{i+1} -limit condition since $p_{i+1} \geq p_i$); we may take j maximizing $|X_{i,j}|$ and set $X_{i+1} = X_{i,j}, Y_{i+1} = Y_i$. We have $D_\infty(X_{i+1}) \leq D_\infty(X_i) + m \leq k_{i+1}$ and $D_\infty(Y_{i+1}) = D_\infty(Y_i) \leq k_{i+1}$ so we are done. In the interesting case, the p_i -limit condition holds on the right, and we apply the mirror set lemma. We pass to some $X'_i \subseteq X_i$ so that $D_\infty(X'_i) \leq k_i + 2p_i n + 2$ and the rectangle $X'_i \times Y_i$ satisfies the $64k_i p_i$ -limit condition on the left (we are able to apply this lemma since $64k_i p_i \leq \frac{1}{4}$ by (2)). We then consider the decomposition $X'_i = X'_{i,1} \cup \dots \cup X'_{i,2^m}$ with $X'_{i,j} = X'_i \cap X_{i,j}$ and set X_{i+1} to equal the largest component (and $Y_{i+1} = Y_i$). The deficit of X_{i+1} is bounded by $k_i + 2p_i n + 2 + m$ and the rectangle $X_{i+1} \times Y_{i+1}$ satisfies the $64k_i p_i$ -limit condition. Since $k_{i+1} \geq k_i + 2p_i n + 2 + m$ and $p_{i+1} \geq 64k_i p_i$ by (1)/(2) we are done. $\square$

It is instructive to compare the evolution of the random coordinate sets in this proof with that of the free coordinate sets in the iterative random restriction proof of Håstad [Has86]. The two arguments traverse the circuit in opposite directions. Here we need to choose random coordinate sets of expected size $\Omega(m)$, and allow their expected size to grow cubically at each stage, ending at $n/4$. Read backwards, the size of the random sets starts close to n and shrinks cubically at each step. In the random-restriction proof, by contrast, each random restriction shrinks the number of free coordinates by a factor $\Theta(m)$. This more favorable parameter evolution yields the tight communication-cost lower bound $m = \Omega(n^{1/(d-1)})$; we will be able to (almost) match this parameter sequence later on in Sections 3 (growing the sets by a factor $\Theta_d(m^2)$) and 4 (growing them by a factor $\Theta_d(m)$).

2.2 Proof of the Mirror Set Lemma

To prove the mirror set lemma we need two combinatorial lemmas which occur either directly, or in a slightly altered form, in [GRSS24]. The first appears roughly as Lemma 9 in [GRSS24]. We give

a proof here which is noticeably simpler and yields a stronger bound. Lemma 9 in [GRSS24] is also sufficient to obtain the main result in this section, and indeed the original machine-generated proof used Lemma 9 of [GRSS24] as a black-box rather than our Lemma 3 below. In the following, we say that a set $X \subseteq \{0, 1\}^n$ “shatters” $R \subseteq [n]$ if $\{x_R \mid x \in X\} = \{0, 1\}^R$. We require a well-known result of Pajor which strengthens the classical Sauer-Shelah lemma [Sau72, She72]:

Lemma (Pajor’s Lemma [Paj85]). *Any $X \subseteq \{0, 1\}^n$ shatters at least $|X|$ distinct sets.*

Lemma 3 (Shattering Lemma). *If $X \neq \emptyset$ and $0 \leq r \leq 1/2$, then*

$$\Pr_{R \subseteq_r [n]}[X \text{ does not shatter } R] \leq 4rD_\infty(X).$$

Proof. Clearly we may assume $r > 0$. Below we reproduce a proof of a standard inequality [Zha24, Lemma 4.3.7]: for any downward-closed family $\mathcal{A} \subseteq 2^{[n]}$, $\Pr_{R \subseteq_r [n]}[R \in \mathcal{A}] \geq (\frac{|\mathcal{A}|}{2^n})^{4r}$. Apply this to the family $\mathcal{A}$ of sets shattered by X ; by Pajor’s Lemma [Paj85], we have $|\mathcal{A}| \geq |X|$, hence:

$$\Pr_R[R \notin \mathcal{A}] \leq 1 - \left(\frac{|\mathcal{A}|}{2^n}\right)^{4r} \leq 1 - \left(\frac{|X|}{2^n}\right)^{4r} \leq 1 - 2^{-4rD_\infty(X)} \leq 4rD_\infty(X)$$

where we apply the bound $1 - 2^{-t} \leq t$ valid for all $t \geq 0$. It remains to prove the general inequality for downward-closed families. Let $h = \lfloor \frac{1}{2r} \rfloor$, so $h \geq \frac{1}{4r}$, and sample h independent sets $R_1, \dots, R_h \subseteq_r [n]$. Their union has the same distribution as a sample $A \subseteq_\theta [n]$, where $\theta := 1 - (1 - r)^h \leq hr \leq \frac{1}{2}$. Note that the quantity $\Pr_{A \subseteq_\alpha [n]}[A \in \mathcal{A}]$ is decreasing in α since $\mathcal{A}$ is downward closed. We thus have:

$$\frac{|\mathcal{A}|}{2^n} = \Pr_{A \subseteq_{\frac{1}{2}} [n]}[A \in \mathcal{A}] \leq \Pr_{A \subseteq_\theta [n]}[A \in \mathcal{A}] = \Pr\left[\bigcup_i R_i \in \mathcal{A}\right] \leq \Pr\left[\bigwedge_i (R_i \in \mathcal{A})\right] = \Pr[R_1 \in \mathcal{A}]^h$$

□

We also need the following well-known fact about min-entropy, whose proof can be found in [GRSS24] as a subclaim in the proof of Lemma 6. This “entropy lemma” says that if we condition on the value a random variable in $\{0, 1\}^n$ takes on the coordinates outside a set $S \subseteq [n]$, the entropy deficit on the remaining coordinates in S will be at most the entropy deficit of the original random variable (on average).

Lemma 4 (Entropy Lemma). *For $S \subseteq [n]$ and $x \in X \subseteq \{0, 1\}^n$, let $X_{x,S} = X \cap [x]_S$, viewed inside its $|S|$ -dimensional cube (in particular its min-entropy deficit is defined with respect to this subcube). Then*

$$\mathbb{E}_{x \sim X} D_\infty(X_{x,S}) \leq D_\infty(X).$$

Proof. Let $\mathbf{X}$ be uniform on X . Then

$$\mathbb{E}_{x \sim X} D_\infty(X_{x,S}) = |S| - H(\mathbf{X}_S \mid \mathbf{X}_{\bar{S}}) = |S| - H(\mathbf{X}) + H(\mathbf{X}_{\bar{S}}) \leq |S| - (n - D_\infty(X)) + |\bar{S}| = D_\infty(X)$$

where the first equality uses the agreement of the entropies H_∞, H on uniform distributions, and the second equality uses the chain rule for Shannon entropy. □

At this point we are ready to prove the mirror set lemma, whose statement we reproduce for the pleasure of the reader:

Lemma (Mirror Set Lemma (Lemma 2), Restated). *Suppose $D_\infty(X) \leq k$, where $k \geq 1$, $p > 0$, and the rectangle $X \times Y$ satisfies the p -limit condition on the left. If $q = 64kp \leq 1$ then there exists a “mirror set” $Y' \subseteq Y$ such that the rectangle $X \times Y'$ satisfies the q -limit condition on the right, and $D_\infty(Y') \leq k + 2pn + 2$.*

We describe the high-level structure of the proof before beginning formally. Say that $x, y \in \{0, 1\}^n$ are “ P -neighbors” if $y \in [x]_P$ (equivalently, $x \in [y]_P$). When x is a p -limit of Y , this means that the elements of Y “surround” x in the following geometric sense: with high probability over $P \subseteq_p [n]$, x has some P -neighbor contained in Y . We aim to show that, assuming this holds and X is suitably large, many points in Y will also be surrounded by X (where in this case surrounded = q -limit). It is natural to try to prove this by a simple random process: sample $x \sim X$, $P \subseteq_p [n]$, and choose y to be a random P -neighbor of x . If we could prove that y was typically a q -limit of X , we would be part of the way there, but at this point we’d have no guarantee that this random process will ever generate some $y \in Y$. However, with an extra “guiding step” we can nudge this random step from X so that it frequently falls into Y . Define a *guiding function* $g : X \times 2^{[n]} \rightarrow \{0, 1\}^n$ to be any function such that $g(x, P) \in [x]_P$ for all x, P . Consider choosing a guiding function g which selects $g(x, P)$ to lie in Y whenever $[x]_P \cap Y$ is nonempty, and which is set to some arbitrary element of $[x]_P$ otherwise. We now consider the *guided random step*: sample x, P and walk to $g(x, P) \in [x]_P$. Since every element of X is a p -limit of Y , this random step will land in Y with probability $\geq \frac{3}{4}$. If we can show that this guided random output y is usually a q -limit of X , then we will have established the existence of some $y \in Y$ which is a q -limit of X . Finally, if we can prove that this guided random step has sufficiently high min-entropy, we can conclude that *many* elements of Y are q -limits of X and the lemma will be proven.

At this point, we have reduced the lemma to proving a statement *purely about X* . Namely, assuming X is large, if we take any guiding function g and sample $x \sim X, P \subseteq_p [n]$, with high probability we have that the guided random step $g(x, P)$ is a q -limit of X . It suffices to show that with high probability over x, P , *every element of $[x]_P$* is a q -limit of X ; this will handle all possible guiding functions. Moreover we need to establish that any such guided random step has suitably high entropy (so that we may obtain many different q -limits of X in Y in the end). The latter entropic claim follows from a simple double-counting argument. The crux of the entire proof is showing the former claim, that every P -neighbor of x is a q -limit of X with high probability over $x \sim X, P \subseteq_p [n]$.

Proof of Mirror Set Lemma. We follow the outline above. Fix any guiding function $g(x, P) \in [x]_P$ and consider the random variable $\mathbf{Y}^g$ defined as follows: sample $x \sim X, P \subseteq_p [n]$, and output $y = g(x, P)$. We establish the following two key claims for all g :

1. $\Pr[\mathbf{Y}^g \text{ is a } q\text{-limit of } X] \geq \frac{1}{2}$
2. $H_\infty(\mathbf{Y}^g) \geq n - k - 2pn$

Given these the lemma follows directly. Set $g(x, P)$ to be any element of $Y \cap [x]_P$ if $Y \cap [x]_P \neq \emptyset$, otherwise define it to be some arbitrary element of $[x]_P$. By the p -limit condition for the rectangle $X \times Y$ on the left, we have that $\Pr[\mathbf{Y}^g \in Y] \geq \frac{3}{4}$. Combining this with (1), we have that $\Pr[\mathbf{Y}^g \in Y \wedge \mathbf{Y}^g \text{ is a } q\text{-limit of } X] \geq \frac{1}{4}$. Applying (2), $H_\infty(\mathbf{Y}^g \mid \mathbf{Y}^g \in Y \wedge \mathbf{Y}^g \text{ is a } q\text{-limit of } X) \geq n - k - 2pn - 2$. If we now take Y' to be those elements y in the support of $\mathbf{Y}^g$ satisfying “ $y \in Y \wedge y$ is a q -limit of X ”, then we must have $D_\infty(Y') \leq k + 2pn + 2$ and we are done.

We establish (2) first since it’s simpler. Let $y \in \{0, 1\}^n$. The key point is that, conditioned on having sampled a particular set P during the generation of $\mathbf{Y}^g$, there are at most $2^{|P|}$ possible values

of x that could have produced a given output y . Hence

$$\Pr[\mathbf{Y}^g = y] = \mathbb{E}_P \Pr[\mathbf{Y}^g = y \mid P \text{ is sampled}] \leq \frac{1}{|X|} \mathbb{E}_P 2^{|P|} = \frac{1}{|X|} \prod_{i \leq n} (\mathbb{E}_{b \sim \text{ber}(p)} 2^b) = \frac{(1+p)^n}{|X|}$$

so $H_\infty(\mathbf{Y}^g) \geq \log |X| - n \log(1+p) \geq n - k - 2pn$.

We now establish (1). By Markov, it suffices to show

$$\Pr_{x,P,Q} [g(x,P)]_Q \cap X \neq \emptyset \geq \frac{15}{16}, \text{ which follows from } \Pr_{x,P,Q} [\forall y \in [x]_P, [y]_Q \cap X \neq \emptyset] \geq \frac{15}{16}$$

where $x \sim X, P \subseteq_p [n], Q \subseteq_q [n]$ are sampled independently in both of the above expressions. We interpret the inequality on the right above. Let $\mathbf{X}$ be the random variable distributed uniformly on X . We sample two random sets $P \subseteq_p [n], Q \subseteq_q [n], |Q| \gg |P|$ (typically), and want to show that if we reveal the coordinates of $\mathbf{X}$ *outside* $P \cup Q$, then for every possible modification of the coordinates in P , there is some subsequent modification to the coordinates in Q which leads us back into the support of $\mathbf{X}$. Since the subsequent modifications to Q allow us to overwrite any initial modification to $P \cap Q$, we may assume that the first step only modifies $P \setminus Q$. Then in terms of the random variable $\mathbf{X}$, this is saying that if we learn the bits outside of $P \cup Q$, with high probability we will not learn *anything definitive* about the possible values of the random variable on the coordinates in $P \setminus Q$: every candidate value in $\{0,1\}^{P \setminus Q}$ will still have positive probability. We can phrase this directly in the language of shattering: if there exists $y \in [x]_P$ such that $[y]_Q \cap X = \emptyset$, this means that the set $X_{x,P \cup Q}$ does not shatter $P \setminus Q$, where $X_{x,P \cup Q} = X \cap [x]_{P \cup Q}$ is defined as in the statement of the entropy lemma. So our goal is precisely to prove:

$$\Pr_{x,P,Q} [X_{x,P \cup Q} \text{ does not shatter } P \setminus Q] \leq \frac{1}{16}$$

Now, choose some $x \in X$ and $S \subseteq [n]$ and consider the above probability when we fix this value of x and condition on $P \cup Q = S$. Note that, conditioned on $P \cup Q = S$, $P \setminus Q$ has the same distribution as a sample $R \subseteq_r S$ where $r := \frac{p(1-q)}{p+q-pq} \leq \frac{p}{q} \leq \frac{1}{64k}$ (remember $q = 64kp$ in the statement of the lemma). Now, we apply the shattering lemma to $X_{x,S}$ (considered to lie inside $\{0,1\}^S$). Since $k \geq 1$ we have $r < \frac{1}{2}$ so we may soundly apply it and conclude

$$\Pr_{P,Q} [X_{x,P \cup Q} \text{ does not shatter } P \setminus Q \mid P \cup Q = S] = \Pr_{R \subseteq_r S} [X_{x,S} \text{ does not shatter } R] \leq 4r D_\infty(X_{x,S})$$

We then use the entropy lemma to reason that $D_\infty(X_{x,S})$ is bounded by $D_\infty(X)$ on average. We have:

$$\begin{aligned} \mathbb{E}_S \Pr_{x,P,Q} [X_{x,P \cup Q} \text{ does not shatter } P \setminus Q \mid P \cup Q = S] \\ \leq 4r \mathbb{E}_{x,S} D_\infty(X_{x,S}) \leq 4r D_\infty(X) \leq 4kr \leq \frac{4k}{64k} \leq \frac{1}{16} \end{aligned}$$

□

3 A Stronger Lower Bound

The primary source of the suboptimal parameters in the lower bound in the previous section is the additive pn penalty in the deficit of the mirror set Y' in the mirror set lemma. This required us to take each $p_i n$ to be at least $p_{i-2} n \cdot p_{i-1} n$ and led to a lower bound $m \geq n^{\exp(-\Theta(d))}$ on communication

cost. Indeed, if it were possible to prove an extension of the mirror set lemma, where all aspects of the lemma statement are the same, but the lower bound on the size of the mirror set is improved to $D_\infty(Y') \leq O(k)$, this would be sufficient to obtain an $m \geq \Omega_d(n^{1/d})$ lower bound on communication. In this case, during the inductive depth d argument, we could take $k_{i+1} \leq O(k_i) + m$ to be the current entropy loss at round $i + 1$, and hence maintain $k_i = O_d(m)$ throughout. Since each application of our hypothetical extended mirror set lemma would merely require $p_{i+1} = O(k_i p_i)$, starting with $p_0 = \frac{2}{n}$ we would have $p_d \leq O_d(\frac{m^d}{n})$, and hence $p_d \leq \frac{1}{4}$ whenever $m \leq \epsilon_d n^{1/d}$ for a suitable $\epsilon_d > 0$. With slightly more care (observing that the limit condition is satisfied on both sides at the start) this argument could be improved to $m \geq \Omega_d(n^{1/(d-1)})$, which would be optimal up to the hidden constant in $\Omega_d(\cdot)$ depending on d .

Unfortunately this proposed extension of the mirror set lemma is false for a quite basic reason: under the assumptions of the lemma (namely, $D_\infty(X) \leq k$ and everything in X is a p -limit of Y), it is not possible to bound $D_\infty(Y)$ by any function of k . We thus can't even get the desired lower bound on the density of Y itself, let alone the density of the elements in Y which are q -limits of X . An example is as follows: let $[n]$ be partitioned into $\sqrt{n}$ blocks of length $\sqrt{n}$, let Y consist of the strings whose parity in every block is even, and let X be its complement. Then $D_\infty(X) \leq 1$, and for $p = \frac{\log n}{\sqrt{n}}$, the p -limit condition for $X \times Y$ on the left is satisfied: a random set $P \subseteq_p [n]$ will intersect every block with probability $\geq 1 - \sqrt{n}(1 - \frac{\log n}{\sqrt{n}})^{\sqrt{n}} \geq \frac{3}{4}$. On the other hand, $D_\infty(Y) = \sqrt{n} \gg 1$.

In the above counterexample, we had a set Y which was both (1) sparse and (2) had many p -limits. Upon closer inspection, for these p -limits x , we can observe that for most $P \subseteq_p [n]$, only a very small fraction of $y \in [x]_P$ lie in Y . In order to get around this kind of example, it might therefore suffice to change the requirement on a limit point x and demand that for a typical P , $Y \cap [x]_P$ contains a decently large fraction of the subcube $[x]_P$. This motivates the following definition:

Definition 2. For $x \in \{0, 1\}^n$, $Y \subseteq \{0, 1\}^n$, we say that x is a (p, k) -limit of Y if

$$\Pr_{P \subseteq_p [n]} [|x]_P \cap Y| \geq 2^{|P|-k}] \geq \frac{3}{4}$$

The left and right (p, k) -limit conditions for a rectangle $X \times Y$ are defined exactly as before. The reader may guess by our choice of variable names that we will consider the (p, k) -limit condition for sets with min-entropy deficit $\approx k$. As a sanity check, we can at least confirm that the previous kind of counterexample no longer exists: if $D_\infty(X) \leq k$, and $X \times Y$ satisfies the (p, k) -limit condition on the left, then $D_\infty(Y) \leq O(k)$. In particular:

$$\frac{|Y|}{2^n} = \mathbb{E}_{\substack{z \sim \{0, 1\}^n \\ P \subseteq_p [n]}} \frac{|Y \cap [z]_P|}{2^{|P|}} \geq \frac{3}{4} \frac{|X|}{2^n} 2^{-k} \geq \frac{3}{4} 2^{-2k}$$

So at the very least, we have ruled out the possibility that a set Y can have deficit much larger than k while having at least 2^{n-k} many (p, k) -limits. It remains to show that in fact, we can obtain some $Y' \subseteq Y$ of comparable size consisting entirely of $(O(kp), O(k))$ -limits of X . In this section we will not be able to achieve this exactly, but instead obtain a subset of Y with deficit $O(k)$, consisting of $(O(k^2p), O(k))$ -limits of X ; this will lead to a bound of the form $\exp(n^{1/(2d-2)})$ rather than $\exp(n^{1/(d-1)})$.

Lemma 5 (Strong Mirror Set Lemma). *Let $p > 0, k \geq 1$. Say that $X, Y \subseteq \{0, 1\}^n$, $X \times Y$ satisfies the (p, k) -limit condition on the left, and $D_\infty(X) \leq k$. Assume $q = Ck^2p \leq \frac{1}{2}$, where C is some universal constant. Then there exists $Y' \subseteq Y$, $D_\infty(Y') \leq 2k + 2$, such that $X \times Y'$ satisfies the $(q, 65k)$ -limit condition on the right.*

The high level structure of the argument will be the same; the main difference in the proof is that we must replace the shattering lemma (Lemma 3) with the following:

Lemma 6 (Light Patterns). *Let $k \geq 1, r \in [0, 1]$, and let $\mathbf{X}$ be a random variable in $\{0, 1\}^n$, $D_\infty(\mathbf{X}) \leq k$. There is a universal constant C such that the following holds whenever $r \leq (Ck)^{-2}$: with probability $\geq \frac{63}{64}$ over $R \subseteq_r [n]$ we have:*

$$|\{z \in \{0, 1\}^R \mid \Pr[\mathbf{X}_R = z] \leq 2^{-|R|-1}\}| \leq 2^{|R|-k}$$

We will defer the proof of this lemma to the following subsection, and proceed now to the proof of the strong mirror set lemma under the assumption that it is true.

Proof of the Strong Mirror Set Lemma. Define a “ k -guiding function” to be a function which selects, for each $x \in X$, $P \subseteq [n]$, a set $G(x, P) \subseteq [x]_P$ of size $|G(x, P)| = \lceil 2^{|P|-k} \rceil$ (compare to the guiding function from the proof of the original mirror set lemma, which selected a single element of $[x]_P$). We now consider the random variable $\mathbf{Y}^G$ defined as follows: sample a uniform $x \sim X$, $P \subseteq_p [n]$, a uniform $y \sim G(x, P)$, and output y . By the left (p, k) -limit condition on $X \times Y$, we know that for an appropriate choice of G , we have $\Pr[\mathbf{Y}^G \in Y] \geq \frac{3}{4}$. We want to show that (1) $H_\infty(\mathbf{Y}^G) \geq n - 2k$ and (2) with probability $\geq \frac{1}{2}$, $\mathbf{Y}^G$ is a $(q, 65k)$ -limit of X . If we have both of these, we may take Y' to be the set of $(q, 65k)$ -limits of X lying in Y ; combining (1)/(2) and the fact that $\Pr[\mathbf{Y}^G \in Y] \geq \frac{3}{4}$ guarantees that $D_\infty(Y') \leq 2k + 2$ and we will be done.

The proof of (1) was already sketched in the beginning of this section: for each $y \in \{0, 1\}^n$ we have:

$$\Pr[\mathbf{Y}^G = y] = \mathbb{E}_P \frac{1}{|X|} \sum_{x \in X} \frac{\mathbf{1}\{y \in G(x, P)\}}{|G(x, P)|} \leq \mathbb{E}_P \frac{2^{|P|}}{|X|} 2^{k-|P|} \leq \frac{2^k}{|X|} \leq 2^{2k-n}$$

Again following the proof of the mirror set lemma we apply a Markov argument and reduce (2) to showing

$$\Pr_{x, P, Q, y \sim G(x, P)} \left[|[y]_Q \cap X| \geq 2^{|Q|-65k} \right] \geq \frac{15}{16} \quad (3)$$

where $x \sim X, P \subseteq_p [n], Q \subseteq_q [n]$ are sampled independently. We claim that (3) in turn follows from showing that, for the set $A(x, P, Q) := \{y \in [x]_P \mid |[y]_Q \cap X| < 2^{|Q|-65k}\}$, we have

$$\Pr_{x, P, Q} \left[|A(x, P, Q)| \leq 2^{|P|-k-6} \right] \geq \frac{31}{32} \quad (4)$$

Indeed, if (4) holds, then with probability at least $\frac{31}{32}$ over x, P, Q , we will have that a random $y \sim G(x, P)$ lands in $A(x, P, Q)$ with probability at most 2^{-6} , and hence the overall probability over $x, P, Q, y \sim G(x, P)$ that $|[y]_Q \cap X| < 2^{|Q|-65k}$ is bounded by $\frac{1}{32} + 2^{-6} \leq \frac{1}{16}$.

We interpret the quantity in (4). We want to show that, with probability $\geq \frac{31}{32}$ over the choice of x, P, Q the following holds: for at least a $(1 - 2^{-k-6})$ fraction of the possible modifications we can make to x on the coordinates in P , at least a 2^{-65k} fraction of *subsequent* modifications to the variables in Q will cause us to end up in X . To simplify the picture a bit, we observe that in the first phase we may instead count the fraction of modifications made to variables in $P \setminus Q$, since we may freely reassign the variables in $P \cap Q$ in the second stage, and every assignment in $P \setminus Q$ has exactly $2^{|P \cap Q|}$ extensions along the variables in P . Let $R = P \setminus Q$ and $S = P \cup Q$ in what follows. Say that in the first step we modified the coordinates in R to some pattern $w \in \{0, 1\}^R$. We then want to lower bound the fraction of patterns $z \in \{0, 1\}^Q$ such that the combined string using z on

the bits in Q , w on the bits in R , and $x_{\bar{S}}$ on the bits outside S lies in X ; in particular we want to lower bound this fraction by 2^{-65k} . For the set $X_{x,S} = X \cap [x]_S$, we can observe that this “fraction of Q -patterns leading us to X ” is given by:

$$\frac{|\{x' \in X_{x,S} \mid x'_R = w\}|}{2^{|Q|}} = \frac{|X_{x,S}|}{2^{|Q|}} \Pr[(\mathbf{X}_{x,S})_R = w] \quad (5)$$

$$= 2^{|S| - D_\infty(X_{x,S}) - |Q|} \Pr[(\mathbf{X}_{x,S})_R = w] = 2^{|R| - D_\infty(X_{x,S})} \Pr[(\mathbf{X}_{x,S})_R = w] \quad (6)$$

where $\mathbf{X}_{x,S}$ is uniform on $X_{x,S}$.

At this point, we proceed as in the mirror set lemma, considering the quantity in (4) when we fix a particular value of x and S and condition on $P \cup Q = S$. Say that (x, S) are “good” if $D_\infty(\mathbf{X}_{x,S}) \leq 64k$. Under this conditioning, $R = P \setminus Q$ is sampled as $R \subseteq_r S$ for some $r \leq \frac{p}{q}$ (this is the same calculation from the original mirror set lemma). Under the assumption that (x, S) are good we may apply the “light patterns lemma” (Lemma 6) with min-entropy deficit parameter $64k$ to the random variable $\mathbf{X}_{x,S}$ in $\{0, 1\}^S$, and conclude: with probability $\geq \frac{63}{64}$ over R , for all but a $2^{-64k} \leq 2^{-k-6}$ fraction of the possible assignments $w \in \{0, 1\}^R$, $\Pr[(\mathbf{X}_{x,S})_R = w] \geq 2^{-|R|-1}$. We may safely apply this lemma provided $r \leq (C64k)^{-2}$ for a suitably large constant C , which the assumptions of the lemma guarantee. Now, overall we get that whenever (x, S) are good, with probability $\geq \frac{63}{64}$ over R , for all but a 2^{-k-6} fraction of patterns $w \in \{0, 1\}^R$, the fraction of Q patterns leading us into X is at least

$$2^{|R| - D_\infty(X_{x,S})} 2^{-|R|-1} \geq 2^{-64k-1} \geq 2^{-65k}$$

We conclude with an application of the entropy lemma (Lemma 4), which (using an additional Markov argument) tells us that (x, S) are good with probability at least $\frac{63}{64}$. Overall we determine that the probability in (4) is lower bounded by $\frac{63}{64}$ for any fixing of (x, S) which are good, and the probability that (x, S) are good is at least $\frac{63}{64}$, from which (4) follows by a union bound. $\square$

From this we immediately obtain our improved communication adversary/circuit lower bound:

Theorem 2. *For each $d \geq 2$ there is $\epsilon_d > 0$ such that $\text{KW}(\text{Parity}_n)$ does not have a d -round communication protocol of cost m when $m \leq \epsilon_d n^{1/(2d-2)}$, provided n is sufficiently large. Consequently any depth d circuit computing Parity_n must have $> 2^{\epsilon_d n^{1/(2d-2)}}$ wires.*

Proof Sketch. The proof is essentially identical to that of Theorem 1, using our new limit condition and the strong mirror set lemma in place of the old condition and old lemma. We indicate merely how the parameters must be set. Let $p_0 = p_1 = \frac{2}{n}$, $k_i = 65^i m$, and for $i > 0$, $p_{i+1} = C k_i^2 p_i$ for a suitable constant C . We maintain that, if $X_i \times Y_i$ is the current rectangle after round i , then it satisfies the (p, k) -limit condition (either on the left or right) for some $p \leq p_i, k \leq k_i$, and that $D_\infty(X_i), D_\infty(Y_i) \leq k_i$. Initially, $X_0 \times Y_0$ satisfies the limit condition on both sides, hence we can get through the first round without applying the mirror set lemma which allows us to take $p_0 = p_1$. $\square$

3.1 Random Projections of a Dense Set Are Close to Uniform

The shattering lemma from Section 2 tells us the following: if $\mathbf{X}$ is a high min-entropy variable in $\{0, 1\}^n$, and R is a random small set of coordinates, then the random variable $\mathbf{X}_R$ (the projection of $\mathbf{X}$ onto R) will be close to the uniform distribution with high probability over R , in the following weak sense: the support of $\mathbf{X}_R$ equals that of the uniform distribution (namely, all of $\{0, 1\}^R$). In particular, the quantitative statement says that if $R \subseteq_r [n]$, then $\mathbf{X}_R$ “looks uniform” in this sense

with probability $\geq 1 - O(rD_\infty(\mathbf{X}))$. Note that this is essentially tight and matches what will occur when $\mathbf{X}$ is uniform on a subcube with, say, the first k coordinates fixed to 0 (so $D_\infty(\mathbf{X}) = k$). In this case we will have that $\mathbf{X}_R$ is exactly uniform whenever $R \cap \{1, \dots, k\} = \emptyset$, which happens with probability $\geq 1 - O(kr)$. Another more basic lemma of the same flavor is Shearer’s lemma (which was in fact used in the proof of the weaker shattering lemma in [GRSS24]), which once again says that under the same conditions, $\mathbf{X}_R$ will be close to uniform for most R ; in this case, “close to uniform” is quantified by Shannon entropy. In particular, we can conclude that with probability $\geq 1 - O(rD_\infty(\mathbf{X}))$ over R , $H(\mathbf{X}_R) \geq |R| - 0.1$, in other words we’ve only learned a fraction of a bit of Shannon information about $\mathbf{X}_R$.

In this section we present two more results of the same form, using different notions of closeness to uniformity. Our goal is to prove the “light patterns” lemma (Lemma 6), which says that aside from a very small number of exceptions, every pattern in $\{0, 1\}^R$ will occur in $\mathbf{X}_R$ with at least half its probability mass under the uniform distribution. While the ideal version of this result would require only that $rD_\infty(\mathbf{X}) \ll 1$ (matching the case of subcubes, and the shattering/Shearer lemmas), in this section we are only able to establish it in the regime $rD_\infty(\mathbf{X})^2 \ll 1$. We will prove Lemma 6 as a simple corollary of a more general result that we call the *uniformity lemma* which says that with high probability over R , the $\|\cdot\|_t$ distance of the (appropriately normalized) probability mass function of $\mathbf{X}_R$ from uniform is small. Our proof is inspired by the Fourier-analytic proof of the *uniform marginals lemma* in [GPW20] and its subsequent simplification using the level- ℓ inequality in [Wu18]. A key step in the proof of the uniform marginals lemma is to show that, for high entropy $\mathbf{X}$ and a random set R , $\mathbf{X}_R$ is close to uniform in L_∞ distance *when the distribution is averaged over R* . This is incomparable to our task; on the one hand we require a stronger statement about the distribution of $\mathbf{X}_R$ for most *fixed values* of R , but on the other hand our notion of distributional closeness is weaker. More precisely, in the terminology introduced below, [GPW20, Wu18] aim to bound $\|\mathbb{E}_R f_R - 1\|_\infty$ ², while our goal is to bound $\|f_R - 1\|_t$ for most R , with $t < \infty$.

We begin with some standard definitions from the analysis of Boolean functions; see [O’D14] for a comprehensive textbook on the subject.

Definition 3. For $S \subseteq [m]$, the S^{th} Fourier character $\chi_S : \{0, 1\}^m \rightarrow \{-1, 1\}$ is given by $\chi_S(z) = (-1)^{\sum_{i \in S} z_i}$. For a function $f : \{0, 1\}^m \rightarrow \mathbb{R}$, the S^{th} Fourier coefficient of f is given by $\hat{f}(S) = \mathbb{E}_{z \sim \{0, 1\}^m} f(z) \chi_S(z)$. We use $\|f\|_t := (\mathbb{E}_{z \sim \{0, 1\}^m} |f(z)|^t)^{1/t}$ to denote the (normalized) L_t norm of f .

As in the case of Shearer’s lemma, our results will apply to a wider family of distributions over the random set R than just the law of $R \subseteq_r [n]$; they will hold when R is drawn from any *r-spread* distribution.

Definition 4. A distribution $\mathcal{R}$ over subsets of $[n]$ is *r-spread* if, for every $S \subseteq [n]$, we have $\Pr_{R \sim \mathcal{R}}[R \supseteq S] \leq r^{|S|}$.

We will only need our lemma in the particular case where $R \subseteq_r [n]$ but we state it in this general form since we believe it might be of some wider interest. In particular, we emphasize that the uniform marginals lemma [GPW20, Wu18] mentioned above applies in this same setting (where the random set R is only guaranteed to be *r-spread* rather than having a particular distribution). In the literature on lifting, spreadness is instead referred to as “blockwise density.” Spread sets arise also in the context of the sunflower lemma [ALWZ21] and the fractional expectation threshold conjecture [Tal10, FKNP21] (see [LMMPZ22] for work connecting the sunflower lemma to lifting).

²In their setting, the random set R has a fixed size m and there is a canonical identification of each projection $x \mapsto x_R$ with a map $\{0, 1\}^n \rightarrow \{0, 1\}^m$; this makes $\mathbb{E}_R f_R$ well-defined.

Lemma 7 (Uniformity Lemma). *Let $k \geq 1, t \geq 2, r \in [0, 1]$ be given. Say that $\mathbf{X}$ is a random variable in $\{0, 1\}^n$, $D_\infty(\mathbf{X}) \leq k$, and $\mathcal{R}$ is an r -spread distribution on subsets of $[n]$. Let $f(x) = 2^n \Pr[\mathbf{X} = x]$ be the renormalized probability mass function of $\mathbf{X}$. For $R \subseteq [n]$, let $f_R : \{0, 1\}^R \rightarrow \mathbb{R}$ be the renormalized probability mass function of $\mathbf{X}_R$ given by:*

$$f_R(z) = 2^{|R|} \Pr[\mathbf{X}_R = z] = \mathbb{E}_{x \sim \{0, 1\}^n} [f(x) \mid x_R = z]$$

Then for any $\eta \in (0, 1)$,

$$\Pr_{R \sim \mathcal{R}} \left[\|f_R - 1\|_t > \sqrt{\frac{Crk(t-1)}{\eta}} \right] \leq \eta$$

provided $rk(t-1) < \epsilon$, where $C \in \mathbb{N}, \epsilon > 0$ are universal constants.

Note that the constant function 1 on $\{0, 1\}^R$ is the renormalized probability mass function of the uniform distribution on $\{0, 1\}^R$, so this lemma is giving a bound on the deviation of $\mathbf{X}_R$ from the uniform distribution. To prove this lemma we need two standard consequences of the hypercontractivity theorem of Bonami [Bon70] (see [O'D14]).

Lemma 8 ($L_2 \rightarrow L_t$ Inequality). *For any $f : \{0, 1\}^m \rightarrow \mathbb{R}$, $t \geq 2$,*

$$\|f\|_t^2 \leq \sum_{S \subseteq [m]} (t-1)^{|S|} \widehat{f}(S)^2$$

Lemma 9 (Level- ℓ Inequality). *Let $k \geq 1$, and say that $\mathbf{X}$ is a random variable in $\{0, 1\}^n$ with $D_\infty(\mathbf{X}) \leq k$. For $S \subseteq [n]$, let $b_{\mathbf{X}}(S) := \mathbb{E}[\chi_S(\mathbf{X})]$ be the bias of the S^{th} Fourier character on $\mathbf{X}$. Then for any $\ell \in \{1, \dots, n\}$,*

$$\sum_{|S|=\ell} b_{\mathbf{X}}(S)^2 \leq (Ck)^\ell$$

where C is a universal constant.

Proof of the Uniformity Lemma. We will show

$$\mathbb{E}_{R \sim \mathcal{R}} \|f_R - 1\|_t^2 \leq 2Crk(t-1)$$

where C is the constant from the Level- ℓ inequality which will yield the lemma via Markov. Observe that for any $S \subseteq R$, $\widehat{f_R - 1}(S) = b_{\mathbf{X}}(S)$ when $S \neq \emptyset$, and $\widehat{f_R - 1}(\emptyset) = 0$. Applying first the $L_2 \rightarrow L_t$ inequality, then the r -spreadness condition, then the Level- ℓ inequality (with basic rearrangements in between) we have:

$$\begin{aligned} \mathbb{E}_{R \sim \mathcal{R}} \|f_R - 1\|_t^2 &\leq \mathbb{E}_R \sum_{S \subseteq R} (t-1)^{|S|} \widehat{f_R - 1}(S)^2 = \mathbb{E}_R \sum_{\emptyset \neq S \subseteq R} (t-1)^{|S|} b_{\mathbf{X}}(S)^2 \\ &= \sum_{\emptyset \neq S \subseteq [n]} \Pr_R[R \supseteq S] \cdot (t-1)^{|S|} b_{\mathbf{X}}(S)^2 \leq \sum_{\emptyset \neq S \subseteq [n]} r^{|S|} (t-1)^{|S|} b_{\mathbf{X}}(S)^2 \\ &= \sum_{\ell=1}^n (r(t-1))^\ell \sum_{S \subseteq [n], |S|=\ell} b_{\mathbf{X}}(S)^2 \leq \sum_{\ell=1}^\infty (Crk(t-1))^\ell \leq 2Crk(t-1) \end{aligned}$$

where in the final inequality we require $Crk(t-1) \leq \frac{1}{2}$, which follows from our assumption that $rk(t-1)$ is sufficiently small. $\square$

We can now prove Lemma 6. We will state here the strengthening to r -spread sets for the sake of completeness, although we only need the case $R \subseteq_r [n]$ in our main results.

Lemma (Light Patterns (Lemma 6), Generalized to Spread Sets). *Let $k \geq 1, r \in [0, 1]$, let $\mathcal{R}$ be an r -spread distribution on $2^{[n]}$, and let $\mathbf{X}$ be a random variable in $\{0, 1\}^n$ with $D_\infty(\mathbf{X}) \leq k$. There is a universal constant C such that the following holds whenever $r \leq (Ck)^{-2}$: with probability $\geq \frac{63}{64}$ over $R \sim \mathcal{R}$ we have:*

$$|\{z \in \{0, 1\}^R \mid \Pr[\mathbf{X}_R = z] \leq 2^{-|R|-1}\}| \leq 2^{|R|-k}$$

Proof. For a fixed R , let $Z_R = \{z \in \{0, 1\}^R \mid \Pr[\mathbf{X}_R = z] \leq 2^{-|R|-1}\}$. Set $t = \max\{k, 2\}$. Then $|f_R(z) - 1|^t \geq 2^{-t}$ for all $z \in Z_R$. Say that $|Z_R| \geq 2^{|R|-k}$. Then we have

$$\|f_R - 1\|_t \geq (2^{-t} \Pr_{z \sim \{0, 1\}^R}[z \in Z_R])^{1/t} \geq 2^{-\frac{k}{t}-1} \geq \frac{1}{4}$$

Now, applying the uniformity lemma, we have $\Pr_{\mathcal{R}}[|Z_R| \geq 2^{|R|-k}] \leq \frac{1}{64}$ provided that $\sqrt{Ckr(t-1)} < \frac{1}{4}$ for some universal constant C . This holds provided $r \leq (C'k)^{-2}$ for some universal constant C' . $\square$

4 A Nearly Tight Lower Bound

Finally in this section we prove:

Theorem 3. *For each $d \geq 2$ there is $\epsilon_d > 0$ such that $\text{KW}(\text{Parity}_n)$ does not have a d -round communication protocol of cost m when $m \leq \epsilon_d n^{1/(d-1)}$, provided n is sufficiently large. Consequently any depth d circuit computing Parity_n must have $> 2^{\epsilon_d n^{1/(d-1)}}$ wires.*

Theorem 3 will be a consequence of the following improvement to Lemma 6:

Lemma 10 (Improved Light Patterns). *Let $k \geq 1, 0 \leq r \leq (512k)^{-1}$, and let $\mathbf{X}$ be a random variable in $\{0, 1\}^n$, $D_\infty(\mathbf{X}) \leq k$. With probability $\geq \frac{63}{64}$ over $R \subseteq_r [n]$, we have*

$$|\{z \in \{0, 1\}^R \mid \Pr[\mathbf{X}_R = z] \leq 2^{-|R|-2k-2}\}| \leq 2^{|R|-k}$$

Formally this is incomparable to Lemma 6; the important sense in which it is stronger is the weakened requirement $r \leq (Ck)^{-1}$ for a universal constant C , whereas Lemma 6 required $r \leq (Ck)^{-2}$. We do however have to pay for this by restricting quantitatively the definition of a “light pattern” (the kinds of patterns which we aim to bound the number of): in Lemma 6 these were patterns whose probabilities decayed by a factor $\leq \frac{1}{2}$ compared to uniform, whereas here they are patterns whose probabilities decayed by $\leq 2^{-2k-2}$. Fortunately, using this more restrictive definition of a light pattern comes at no cost to us in our application to Theorem 3. Theorem 3 is derived from Lemma 10 in exactly the same way that Theorem 2 is derived from Lemma 6 (except we now achieve $q = O(kp)$ in our mirror set lemma) and so we will not repeat the argument here. The remainder of this section will be dedicated to proving Lemma 10.

Let $f_R : \{0, 1\}^R \rightarrow \mathbb{R}$ be the renormalized distribution function of $\mathbf{X}_R$ as in Section 3, given by $f_R(z) = 2^{|R|} \Pr[\mathbf{X}_R = z]$. Our approach will be to show that with probability at least $63/64$ over $R \subseteq_r [n]$,

$$\mathbb{E}_{z \sim \{0, 1\}^R} f_R(z)^{-1} \leq 4 \cdot 2^k \tag{7}$$

with the convention $0^{-1} = +\infty$. This immediately yields the bound in Lemma 10 by Markov. We will derive this from some general analytic inequalities involving a transform on functions over the hypercube which we will call the “harmonic mean transform:”

Definition 5 (Harmonic Mean Transform). For a function $f : \{0, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$ and any $R \subseteq [n]$, define $f_R : \{0, 1\}^R \rightarrow \mathbb{R}$ by

$$f_R(z) = \mathbb{E}_{x \sim \{0, 1\}^n} [f(x) \mid x_R = z]$$

We then define $\check{f} : 2^{[n]} \rightarrow \mathbb{R}_{\geq 0}$, the “harmonic mean transform of f ,” by

$$\check{f}(R) = \left(\mathbb{E}_{z \sim \{0, 1\}^R} f_R(z)^{-1} \right)^{-1}$$

where we use the conventions $0^{-1} = +\infty$, $(+\infty)^{-1} = 0$.

Our desired bound in (7) can be immediately recast in terms of the harmonic mean transform of the renormalized distribution function of $\mathbf{X}$: for $f(x) = 2^n \Pr[\mathbf{X} = x]$, bound (7) becomes $\check{f}(R) \geq 2^{-k-2}$. In some very rough sense this transform will play a role similar to that of the Fourier transform in the proof of the uniformity lemma in Section 3. To be precise, we believe the most accurate analogy is between $\check{f}(R)$ and the quantity $\sum_{\emptyset \neq S \subseteq R} (t-1)^{|S|} \hat{f}(S)^2$ from the uniformity lemma’s proof, rather than between $\check{f}(R)$ and $\hat{f}(R)$. We start with presenting an alternate variational formula for the harmonic mean transform:

Lemma 11 (Variational Formula). For all $f : \{0, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$, $R \subseteq [n]$,

$$\check{f}(R) = \inf_{\substack{w : \{0, 1\}^R \rightarrow \mathbb{R}_{\geq 0} \\ \mathbb{E}_z w(z) = 1}} \mathbb{E}_z [f_R(z) w(z)^2]$$

where each expectation over z is uniform on $\{0, 1\}^R$.

Proof. We may assume $f_R(z) > 0$ for all $z \in \{0, 1\}^R$, otherwise both sides of the identity must be zero. It is easy to see that the right hand side is at most the left by taking $w(z) = \check{f}(R) f_R(z)^{-1}$. For the interesting direction, take any admissible w and apply Cauchy-Schwarz:

$$1 = (\mathbb{E}_z w(z))^2 = (\mathbb{E}_z (\sqrt{f_R(z)} w(z)) f_R(z)^{-1/2})^2 \leq \frac{\mathbb{E}_z [f_R(z) w(z)^2]}{\check{f}(R)}$$

□

Three useful corollaries follow immediately:

Corollary 1. For all $f, g, h : \{0, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$ we have:

1. The map $f \mapsto \check{f}$ is concave: if $f = \lambda g + (1 - \lambda)h$ for some $\lambda \in [0, 1]$, then $\check{f} \geq \lambda \check{g} + (1 - \lambda) \check{h}$ pointwise.
2. $\check{f}(R) \geq \check{f}(S)$ whenever $R \subseteq S$.
3. $\check{f}(R) \leq \mathbb{E} f = \check{f}(\emptyset)$.

Proof. The first follows since the variational formula expresses $f \mapsto \check{f}(R)$ as an infimum over linear functions of f . The second follows since any admissible $w : \{0, 1\}^R \rightarrow \mathbb{R}_{\geq 0}$ in the variational form for $\check{f}(R)$ extends to an admissible $w' : \{0, 1\}^S \rightarrow \mathbb{R}_{\geq 0}$ in the variational form for $\check{f}(S)$ achieving the same value: take $w'(z) = w(z_R)$. The third is immediate from the second. □

We now arrive at the central analytic inequality we use for the harmonic mean transform:

Lemma 12. For any $f : \{0, 1\}^n \rightarrow \mathbb{R}_{\geq 0}$,

$$\mathbb{E}_{R \subseteq_{1/4} [n]} \sqrt{\check{f}(R)} \geq \mathbb{E}_{x \sim \{0,1\}^n} \sqrt{f(x)}$$

Proof. We prove by induction on n . For $n = 1$, let $a = f(0), b = f(1)$. In this case

$$\check{f}(\emptyset) = \mathbb{E}f = \frac{a+b}{2}, \quad \check{f}(\{1\}) = \frac{2ab}{a+b}$$

so we aim to prove:

$$\frac{3}{4} \sqrt{\frac{a+b}{2}} + \frac{1}{4} \sqrt{\frac{2ab}{a+b}} \geq \frac{\sqrt{a} + \sqrt{b}}{2}$$

with the convention here and below that $\frac{2ab}{a+b} = 0$ when $a = b = 0$. The case $a = b = 0$ is immediate, so we assume $a + b > 0$. To verify the inequality, set $u = \sqrt{a}, v = \sqrt{b}, c = \sqrt{(u^2 + v^2)/2}$, $d = (u + v)/2$, and rewrite it as:

$$\frac{3}{4}c + \frac{uv}{4c} \geq d$$

Since $uv = 2d^2 - c^2$, $c > 0$, we have:

$$\frac{3}{4}c + \frac{uv}{4c} - d = \frac{3}{4}c + \frac{2d^2 - c^2}{4c} - d = \frac{(c - d)^2}{2c} \geq 0$$

For the inductive step, let $n \geq 2$ and define

$$f_0(x) = f(x, 0), \quad f_1(x) = f(x, 1), \quad f_{\text{avg}} = \frac{f_0 + f_1}{2}.$$

Observe that $\check{f}_{\text{avg}} \geq \frac{1}{2}(\check{f}_0 + \check{f}_1)$ pointwise by concavity. For every $S \subseteq [n-1]$ we have:

$$\check{f}(S) = \check{f}_{\text{avg}}(S), \quad \check{f}(S \cup \{n\}) = \frac{2\check{f}_0(S)\check{f}_1(S)}{\check{f}_0(S) + \check{f}_1(S)},$$

where the fraction on the right is interpreted as zero when its denominator is zero. Sampling $S \subseteq_{1/4} [n-1]$ and conditioning on whether $n \in R$, we obtain

$$\begin{aligned} \mathbb{E}_{R \subseteq_{1/4} [n]} \sqrt{\check{f}(R)} &= \mathbb{E}_S \left[\frac{3}{4} \sqrt{\check{f}_{\text{avg}}(S)} + \frac{1}{4} \sqrt{\frac{2\check{f}_0(S)\check{f}_1(S)}{\check{f}_0(S) + \check{f}_1(S)}} \right] \\ &\geq \mathbb{E}_S \left[\frac{3}{4} \sqrt{\frac{\check{f}_0(S) + \check{f}_1(S)}{2}} + \frac{1}{4} \sqrt{\frac{2\check{f}_0(S)\check{f}_1(S)}{\check{f}_0(S) + \check{f}_1(S)}} \right] \\ &\geq \frac{1}{2} \mathbb{E}_S \sqrt{\check{f}_0(S)} + \frac{1}{2} \mathbb{E}_S \sqrt{\check{f}_1(S)} \\ &\geq \frac{1}{2} \mathbb{E}_{x \sim \{0,1\}^{n-1}} \sqrt{f_0(x)} + \frac{1}{2} \mathbb{E}_{x \sim \{0,1\}^{n-1}} \sqrt{f_1(x)} \\ &= \mathbb{E}_{x \sim \{0,1\}^n} \sqrt{f(x)}. \end{aligned}$$

where the first inequality uses concavity, the second applies the $n = 1$ case, and the third applies the inductive hypothesis. $\square$

To apply this inequality to study a sparser random set $R \subseteq_r [n]$, we apply an altered form of the same standard transfer principle for downward closed families used in the proof of the shattering lemma [Zha24, Lemma 4.3.7]: for any nonempty downward-closed family of sets $\mathcal{A} \subseteq 2^{[n]}$ and $0 \leq r \leq 1/20$,

$$\Pr_{R \subseteq_r [n]} [R \in \mathcal{A}] \geq \left(\Pr_{S \subseteq_{1/4} [n]} [S \in \mathcal{A}] \right)^{5r}.$$

The proof is exactly the same as the case we presented during the proof of the shattering lemma (in that case we used $1/2$ in place of $1/4$). At this point we are ready to prove Lemma 10:

Proof of Lemma 10. As per the preceding discussion, it suffices to establish that, with probability $\geq \frac{63}{64}$ over $R \subseteq_r [n]$, we have $\check{f}(R) \geq 2^{-k-2}$, where f is the renormalized distribution function of $\mathbf{X}$. Let $\mathcal{A} \subseteq 2^{[n]}$ be the family of sets R such that $\check{f}(R) \geq 2^{-k-2}$. By the second point in Corollary 1, we know that $\mathcal{A}$ is a downward-closed family. We will prove that $\Pr_{R \subseteq_{1/4} [n]} [R \in \mathcal{A}] \geq 2^{-k/2-1}$. Applying the above transfer principle and the standard estimate $1 - 2^{-t} \leq t$ valid for all $t \geq 0$ we have

$$\Pr_{R \subseteq_r [n]} [R \notin \mathcal{A}] \leq 1 - (2^{-k/2-1})^{5r} \leq 5r(k/2 + 1) \leq \frac{1}{64}$$

where the final inequality simply applies our assumptions $r \leq (512k)^{-1}, k \geq 1$. It remains only to prove $\Pr_{R \subseteq_{1/4} [n]} [R \in \mathcal{A}] \geq 2^{-k/2-1}$. Applying Lemma 12:

$$\mathbb{E}_{R \subseteq_{1/4} [n]} \sqrt{\check{f}(R)} \geq \mathbb{E}_{x \sim \{0,1\}^n} \sqrt{f(x)} \geq 2^{-D_\infty(\mathbf{X})/2} \mathbb{E}_x f(x) \geq 2^{-k/2}$$

Hence, by Markov,

$$\Pr_{R \subseteq_{1/4} [n]} [\sqrt{\check{f}(R)} < \frac{1}{2} 2^{-k/2}] \leq 1 - 2^{-k/2-1}$$

using the fact (point 3 of Corollary 1) that $\sqrt{\check{f}(R)} \leq 1 = \mathbb{E}f$ for all R . □

Acknowledgments

The author thanks Mika Göös, Toniann Pitassi, Artur Riazanov, and Avi Wigderson for their comments on an initial draft of this manuscript.

References

- [Paj85] Alain Pajor. *Sous-espaces ℓ_1^n des espaces de Banach*. Travaux en cours, Hermann, Paris, 1985.
- [Sau72] Norbert Sauer. *On the Density of Families of Sets*. Journal of Combinatorial Theory, Series A, 13(1):145–147, 1972. [https://doi.org/10.1016/0097-3165\(72\)90019-2](https://doi.org/10.1016/0097-3165(72)90019-2).
- [She72] Saharon Shelah. *A Combinatorial Problem; Stability and Order for Models and Theories in Infinitary Languages*. Pacific Journal of Mathematics, 41(1):247–261, 1972. <https://doi.org/10.2140/pjm.1972.41.247>.
- [Zha24] Yufei Zhao. *Probabilistic Methods in Combinatorics*. MIT 18.226 lecture notes, Fall 2022. Updated June 18, 2024. https://yufeizhao.com/pm/probmethod_notes.pdf.
- [GRSS24] Mika Göös, Artur Riazanov, Anastasia Sofronova, and Dmitry Sokolov. *Top-Down Lower Bounds for Depth-Four Circuits*. arXiv:2304.02555v2, 2024. <https://arxiv.org/abs/2304.02555>.
- [KW90] Mauricio Karchmer and Avi Wigderson. *Monotone Circuits for Connectivity Require Super-Logarithmic Depth*. SIAM Journal on Discrete Mathematics, 3(2):255–265, 1990. <https://doi.org/10.1137/0403021>.

- [KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. *Super-Logarithmic Depth Lower Bounds via the Direct Sum in Communication Complexity*. Computational Complexity, 5(3–4):191–204, 1995. <https://doi.org/10.1007/BF01206317>.
- [HJP95] Johan Håstad, Stasys Jukna, and Pavel Pudlák. *Top-Down Lower Bounds for Depth-Three Circuits*. Computational Complexity, 5(2):99–112, 1995. <https://doi.org/10.1007/BF01268140>.
- [PPZ99] Ramamohan Paturi, Pavel Pudlák, and Francis Zane. *Satisfiability Coding Lemma*. Chicago Journal of Theoretical Computer Science, 1999, Article 11, pages 1–19. <https://doi.org/10.4086/cjtc.1999.011>.
- [MW19] Or Meir and Avi Wigderson. *Prediction from Partial Information and Hindsight, with Application to Circuit Lower Bounds*. Computational Complexity, 28(2):145–183, 2019. <https://doi.org/10.1007/s00037-019-00177-4>.
- [FSS84] Merrick Furst, James B. Saxe, and Michael Sipser. *Parity, Circuits, and the Polynomial-Time Hierarchy*. Mathematical Systems Theory, 17:13–27, 1984. <https://doi.org/10.1007/BF01744431>.
- [Ajt83] Miklós Ajtai. Σ_1^1 -Formulae on Finite Structures. Annals of Pure and Applied Logic, 24(1):1–48, 1983. [https://doi.org/10.1016/0168-0072\(83\)90038-6](https://doi.org/10.1016/0168-0072(83)90038-6).
- [Yao85] Andrew Chi-Chih Yao. *Separating the Polynomial-Time Hierarchy by Oracles*. In Proceedings of the 26th Annual Symposium on Foundations of Computer Science (FOCS), pages 1–10, 1985. <https://doi.org/10.1109/SFCS.1985.49>.
- [Has86] Johan Håstad. *Almost Optimal Lower Bounds for Small Depth Circuits*. In Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC), pages 6–20, 1986. <https://doi.org/10.1145/12130.12132>.
- [Raz87] Alexander A. Razborov. *Lower Bounds on the Size of Bounded Depth Circuits over a Complete Basis with Logical Addition*. Mathematical Notes of the Academy of Sciences of the USSR, 41(4):333–338, 1987. <https://doi.org/10.1007/BF01137685>.
- [Smo87] Roman Smolensky. *Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity*. In Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC), pages 77–82, 1987. <https://doi.org/10.1145/28395.28404>.
- [Bon70] Aline Bonami. *Étude des coefficients de Fourier des fonctions de $L^p(G)$* . Annales de l’Institut Fourier, 20(2):335–402, 1970. <https://doi.org/10.5802/aif.357>.
- [O’D14] Ryan O’Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. <https://doi.org/10.1017/CB09781139814782>.
- [FKNP21] Keith Frankston, Jeff Kahn, Bhargav Narayanan, and Jinyoung Park. *Thresholds versus Fractional Expectation-Thresholds*. Annals of Mathematics, 194(2):475–495, 2021. <https://doi.org/10.4007/annals.2021.194.2.2>.
- [Wu18] Xinyu Wu. *The uniform marginals lemma in [GPW17]*. Expository note, 2018. <https://www.contrib.andrew.cmu.edu/~xinyuw1/papers/uniform-marginals-lemma.pdf>.
- [GPW20] Mika Göös, Toniann Pitassi, and Thomas Watson. *Query-to-Communication Lifting for BPP*. SIAM Journal on Computing, 2020. Preliminary version in FOCS 2017. https://courses.cs.washington.edu/courses/cse599i/24sp/papers/GoosPitassiWatson_2020_journal.pdf.
- [Tal10] Michel Talagrand. *Are Many Small Sets Explicitly Small?* In Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC), pages 13–36, 2010. <https://doi.org/10.1145/1806689.1806693>.
- [ALWZ21] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang. *Improved Bounds for the Sunflower Lemma*. Annals of Mathematics, 194(3):795–815, 2021. <https://doi.org/10.4007/annals.2021.194.3.5>.
- [LMMPZ22] Shachar Lovett, Raghu Meka, Ian Mertz, Toniann Pitassi, and Jiapeng Zhang. *Lifting with Sunflowers*. In 13th Innovations in Theoretical Computer Science Conference (ITCS), LIPIcs 215, pages 104:1–104:24, 2022. <https://doi.org/10.4230/LIPIcs.ITCS.2022.104>.
- [KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997. <https://doi.org/10.1017/CB09780511574948>.
- [RM99] Ran Raz and Pierre McKenzie. *Separation of the Monotone NC Hierarchy*. Combinatorica, 19(3):403–435, 1999. Preliminary version in FOCS 1997. <https://doi.org/10.1007/s004930050062>.
- [GPW18] Mika Göös, Toniann Pitassi, and Thomas Watson. *Deterministic Communication vs. Partition Number*. SIAM Journal on Computing, 47(6):2435–2450, 2018. Preliminary version in FOCS 2015. <https://doi.org/10.1137/16M1059369>.

- [GP18] Mika Göös and Toniann Pitassi. *Communication Lower Bounds via Critical Block Sensitivity*. SIAM Journal on Computing, 47(5):1778–1806, 2018. Preliminary version in STOC 2014. <https://doi.org/10.1137/16M1082007>.
- [dRNV16] Susanna F. de Rezende, Jakob Nordström, and Marc Vinyals. *How Limited Interaction Hinders Real Communication (and What It Means for Proof and Circuit Complexity)*. In Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 295–304, 2016. <https://doi.org/10.1109/FOCS.2016.40>.
- [PR17] Toniann Pitassi and Robert Robere. *Strongly Exponential Lower Bounds for Monotone Computation*. In Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC), pages 1246–1255, 2017. <https://doi.org/10.1145/3055399.3055478>.
- [GGKS20] Ankit Garg, Mika Göös, Pritish Kamath, and Dmitry Sokolov. *Monotone Circuit Lower Bounds from Resolution*. Theory of Computing, 16(13):1–30, 2020. Preliminary version in STOC 2018. <https://doi.org/10.4086/toc.2020.v016a013>.
- [GKRS19] Mika Göös, Pritish Kamath, Robert Robere, and Dmitry Sokolov. *Adventures in Monotone Complexity and TFNP*. In 10th Innovations in Theoretical Computer Science Conference (ITCS), LIPIcs 124, pages 38:1–38:19, 2019. <https://doi.org/10.4230/LIPIcs.ITCS.2019.38>.
- [dRMNPRV20] Susanna F. de Rezende, Or Meir, Jakob Nordström, Toniann Pitassi, Robert Robere, and Marc Vinyals. *Lifting with Simple Gadgets and Applications to Circuit and Proof Complexity*. In Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS), pages 24–30, 2020. <https://doi.org/10.1109/FOCS46700.2020.00011>.
- [dRV25] Susanna F. de Rezende and Marc Vinyals. *Lifting with Colourful Sunflowers*. In 40th Computational Complexity Conference (CCC), LIPIcs 339, pages 36:1–36:19, 2025. <https://doi.org/10.4230/LIPIcs.CCC.2025.36>.
- [KR26] Gülce Kardeş and Benjamin Rossman. *On Top-Down and Local Lower Bounds for AC^0 Circuits*. arXiv:2609.01759, 2026. <https://arxiv.org/abs/2609.01759>.
- [Sip84] Michael Sipser. *A Topological View of Some Problems in Complexity Theory*. In Mathematical Foundations of Computer Science (MFCS), Lecture Notes in Computer Science 176, pages 567–572. Springer, 1984. <https://doi.org/10.1007/BFb0030341>.