Sharp Thresholds And Circuit Lower Bounds

COMS 6998: Unconditional Lower Bounds and Derandomization

Ryan Anselm, Akshat Yaparla, Patrick Yeh

May 9, 2024

1 Introduction

Our project is thematically focused on understanding the nature of the relationship between (circuit) complexity theory and 'sharp threshold' phenomena, a topic that naturally emerges in theoretical computer science, statistical physics, social choice theory, statistics, and other areas. One of the most compelling motivations for pursuing this line of work is to establish a *dictionary* between theoretical computer science and other fields such as statistical physics, where ideas and results from one field can be borrowed for application in the others. As a contemporary example of this, [DRT18] proved new results for the random-cluster and Potts models from statistical physics while making use of the OSSS inequality [ODo+05], which originated from the study of decision trees in theoretical computer science.

More concretely, we focus on the recent result of [GMZ23], which establishes lower bounds on the size/depth of boolean circuits that, in a certain approximate sense, compute functions which exhibit a sharp threshold. This result establishes that phase transition-like sharp threshold behavior, ubiquitous in physics and elsewhere, has rigorously provable computational hardness.

Our report is structured as follows: In section 2, we provide background information relevant to the study of sharp thresholds and key prior results that constitute the foundation for [GMZ23]. Next, in section 3 we give an informal statement of the main result and demonstrate how it can be used in several contexts. In section 4, we give the proof of the main result. Lastly, in section 5 we discuss future lines of work.

2 Background

2.1 Preliminaries

Let a Boolean function $f : \{0,1\}^n \to \{0,1\}$ be *monotone* if on all inputs, flipping any input bit from 0 to 1 can never cause the corresponding output to flip from 1 to 0. Formally, f is monotone if $\forall x, y \in \{0,1\}^n : (\forall i \in [n] : x_i \leq y_i) \implies f(x) \leq f(y)$. For a monotone Boolean function $f : \{0,1\}^n \to \{0,1\}$, consider the probability that an input $X \in \{0,1\}^n$ drawn from a p-biased product distribution $X \sim \mathbb{P}_p = \mathrm{Bern}(p)^{\otimes n}$ satisfies $f(X) = 1$. Denote $\mathbb{E}_p[f] \equiv \mathbb{E}_{X\sim\mathbb{P}_p}[f(X)] = \Pr_{X\sim\mathbb{P}_p}[f(X) = 1]$ to be the probability that f is satisfied when drawing an input from $\mathbb{P}_p$ and the

critical probability p_c to be the probability such that $\mathbb{E}_{p_c}[f] = 1/2$. Note that $g_f : [0,1] \to [0,1]$ where $g_f(p) = \mathbb{E}_p[f]$ is a monotonically increasing real function.

A sharp threshold is a phenomenon where within a small window frame of increasing the Bernoulli parameter p of the distribution $\mathbb{P}_p$ being sampled from, the probability $\mathbb{E}_p[f]$ of a random input to f having an output of 1 jumps rapidly from near 0 to near 1. A formal definition is given below. Concretely we assume that 'near 0' means less than 0.01 and 'near 1' means greater than 0.99, though these numbers are arbitrary and for any $\varepsilon \in (0, 1/2)$ we could use ε and $1 - \varepsilon$ as the respective cutoffs.

Definition 1 (Sharp Threshold)**.** *A monotone Boolean function $f : \{0,1\}^n \to \{0,1\}$ has a Δ-**sharp threshold** around p_c if there exists p_1, p_2 where $0 < p_1 < p_c < p_2 < 1$ such that $\mathbb{E}_{p_1}[f] < 0.01$ and $\mathbb{E}_{p_2}[f] > 0.99$ where $\frac{p_2 - p_1}{\min\{p_c, 1-p_c\}} < \frac{1}{\Delta}$.*

Define critical window width to be $\varepsilon_n = p_2 - p_1$. Note that the sharpness of the threshold is inversely proportional to the smallest possible ratio between the width of the **'critical window'** where $\mathbb{E}_p[f]$ goes from near 0 to near 1 and the closeness of the critical point p_c to 0 or 1. The Bollobás-Thomason Theorem [BT87] states that for every monotone Boolean function f, $\varepsilon_n = p_2 - p_1 = O(\min\{p_c, 1-p_c\})$, so we always have that $\frac{p_2 - p_1}{\min\{p_c, 1-p_c\}} = O(1)$. Colloquially, a threshold is considered to be 'sharp' when $\frac{p_2 - p_1}{\min\{p_c, 1-p_c\}} = o(1)$ and otherwise it is considered to be a coarse threshold.

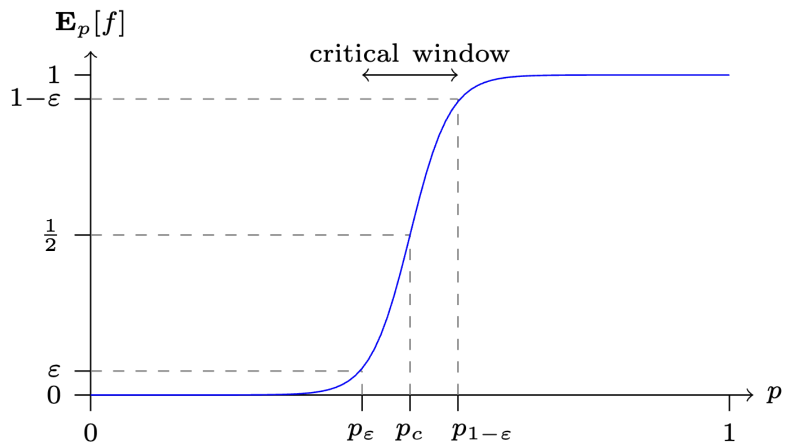

Figure 1: $\mathbb{E}_p[f]$ vs. p graph of some hypothetical function f exhibiting a threshold with the critical probability p_c, $p_\varepsilon \equiv p_1$, and $p_{1-\varepsilon} \equiv p_2$ labeled.

As an example of a function that exhibits a sharp threshold, we will briefly look at the majority function on n bits $\text{MAJ}_n : \{0,1\}^n \to \{0,1\}$ which outputs 1 if and only if a majority of the n input bits are 1. There is a critical point at $p_c = 0.5$ for this function. This can be verified because for any $p < 0.5$, as $n \to \infty$ the probability of at least half of the inputs being 1 goes to 0, so $\mathbb{E}_p[\text{MAJ}_n] \to 0$, and for any $p > 0.5$, as $n \to \infty$ the probability of at least half of the inputs being 1 goes to 1, so $\mathbb{E}_p[\text{MAJ}_n] \to 1$.

The sharpness of the threshold for MAJ_n can be derived from the central limit theorem, as $\text{MAJ}_n \equiv \mathbb{1}\left[\frac{1}{n}\sum_{i=1}^n x_i \geq 1/2\right]$, so the output is dependent on the average of n i.i.d. Bernoulli random variables, which will approximate a Gaussian random variable as n gets large. The width of the critical window depends on the value of p_1 such that $\mathbb{E}_{p_1}[\text{MAJ}_n] = 0.01$ and the value of p_2 such that $\mathbb{E}_{p_2}[\text{MAJ}_n] = 0.99$ (in particular, how far are they from $p_c = 0.5$). Observe that this is entirely dependent on the standard deviation of the distribution, which for $\frac{1}{n}\sum_{i=1}^{n} x_i$ scales as $\Theta(1/\sqrt{n})$. Therefore, we can infer that the critical window width is $\Theta(1/\sqrt{n})$ for MAJ_n, so MAJ_n has a $\Theta(\sqrt{n})$-sharp threshold.

Definition 2 (Computing a function on average around the critical window)**.** *Fix a family of functions $\{f_n\}_{n\in\mathbb{N}}$ where $f_n : \{0,1\}^n \to \{0,1\}$. A family of circuits $\{\mathcal{C}_n\}_{n\in\mathbb{N}}$ where $\mathcal{C}_n : \{0,1\}^n \to \{0,1\}$ computes f on average around the critical window if for some constant $\xi \in (0, 1/2)$, and $p_1, p_2 \in (0,1)$, with $\mathbb{E}_{p_1}[f] = \xi$ and $\mathbb{E}_{p_2}[f] = 1 - \xi$, it holds that for all $\varepsilon > 0$, there exists a sufficiently large n such that*

$$\max\{\mathop{\mathbb{E}}_{p_1}[|\mathcal{C}_n - f|], \mathop{\mathbb{E}}_{p_2}[|\mathcal{C}_n - f|]\} \leq \varepsilon$$

A function is *computed on average around the critical window* if a single circuit agrees with it with high probability for *at least two values* of p around the critical window. We are concerned with this specialized notion of average-case hardness because for several problems worst-case hardness is known but nothing can be said about the conventional notion of average-case hardness. For example, the k-clique problem asks to decide whether a given input graph with n nodes has a k-clique in it. However, if $k = \Theta(n)$, then the vast majority of input graphs with n nodes will not have a k-clique, so the problem is trivial to compute on average given a uniform distribution over inputs. However, computing k-clique in the worst-case instance is known to be NP-hard. Average-case hardness around the critical window strikes a balance in that it is a stronger requirement than worst-case hardness but still holds for many problems of interest where conventional average-case hardness trivially does not hold.

Lastly, we give the definition of AC_0 circuits for completeness because of their relevance to the circuit lower bounds that we will be able to prove.

Definition 3 (AC_0 circuit)**.** *A Boolean circuit using unbounded fan-in AND, OR, and NOT gates with n inputs is an AC_0 circuit if it is of poly(n) size and $O(1)$ depth.*

2.2 Friedgut's Theorem

Next, we turn to discuss some of the significant prior work on sharp thresholds on graph properties. In his survey on sharp thresholds [Fri05], Friedgut gives an overview of the work done concerning sharp thresholds up to that point, but primarily focusing on random graphs. He first provides examples of sharp thresholds of random graphs, elaborated on later, provides theorems from [FB99], and uses them to provide theorems on coarse thresholds and their relationship with random graphs.

Friedgut provides his main theorem, stated below.

Theorem 4 ([FB99] Theorem 1.1)**.** *There exists a function $k(\epsilon, c)$ such that for all $c > 0$, any n, and any monotone symmetric family of graphs A on n vertices such that $p \cdot \frac{d\mu_p(A)}{dp} \leq c$, for every $\epsilon > 0$, there exists a monotone symmetric family B such that $\|B\| \leq k(\epsilon, c)$ and $\mu_p(A \triangle B) \leq \epsilon$. Furthermore, the minimal graphs in B are all balanced.*

He uses an equivalent definition of sharpness, where $\mu_p(A) \equiv \mathbb{E}_p[f]$, but A is a product space. Summarized by Friedgut in [Fri05], this theorem says

"All monotone graph properties with a coarse thresholds (denoted by A in the theorem) may be approximated by a local property (denoted by B)."

In addition to the main theorem, he also conjectured the following, proven in the appendix by Bourgain.

Theorem 5 ([FB99] Theorem 2.4, due to Bourgain)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be the characteristic function of A, a monotone subset of $\{0,1\}^n$. Let $p = p_c(A)$. For $\tau > 0$, let*

$$\Omega_r = \{S : \hat{f}(S)^2 < \tau p^{|S|}\}.$$

Let $p_c \to 0$, $r \to 0$, and $n \to \infty$. Then, if $p_c \cdot d\mu(A)/dp|_{p=p_c} < c$,

$$\sum_{S \in \Omega_r} \hat{f}^2(S) = o(1).$$

This conjecture is summarized as

Properties with sharp thresholds cannot be mostly supported on low-degree Fourier terms, provided the critical point is asymptotically close to 0.

A coarse threshold is the opposite of a sharp threshold, that is, one where the probability is bounded away from 0 and 1 for a non-negligible ($\Theta(1)$ width) range of p values.

Theorem 6 ([Fri05] Theorem 2.4)**.** *Let $\mathcal{A}$ be a graph property with a coarse threshold. Then there exists $p = p(n)$, $\tau > 0$, a fixed graph M with $\Pr[M \subseteq G(n,p)] > \tau, \alpha > 0$ with*

$$\alpha < \Pr[G(n,p) \in \mathcal{A}] < 1 - 3\alpha,$$

and a constant $\epsilon > 0$ such that for every graph property $\mathcal{G}$ such that $G(n,p) \in \mathcal{G}$ holds a.a.s there exists an infinte series of n's, and for each n, a graph $G \in \mathcal{G}$ on n vertices such that the following holds.

$$\Pr[(G \cup M^*) \in \mathcal{A}] > 1 - \alpha,$$
$$\Pr[(G \cup G(n, \epsilon p) \in \mathcal{A}] < 1 - 2\alpha,$$

where the random graph $G(n, \epsilon p)$ is taken on the same vertex set as G. That is, adding a random copy of M boosts the probability of $\mathcal{A}$ more than adding $\epsilon p\binom{n}{2}$ random edges.

He uses the above theorem to then show sharpness in graph colorability and matching in the following two lemmas.

Lemma 7 ([Fri05] Claim 3.2)**.** *Let $k \geq 2$ be fixed, and let $H(n,p)$ be the random $k-$uniform hypergraph on n vertices. Then the threshold for appearance of a perfect matching in $H(n,p)$ is sharp.*

Lemma 8 ([Fri05] Claim 3.3)**.** *Let $k > 2$ be fixed, and let $H(n,p)$ be the random k-uniform hypergraph on n vertices. Then the threshold for $H(n,p)$ being non-2-colorable is sharp.*

The proofs of both are based on the hypergraph version of Theorem 6. They use a proof by contradiction, assuming that these problems have a coarse threshold, and show a contradiction with one of the conditions laid out in Theorem 6. The former problem is of particular interest as finding a maximal matching in a graph is a famous $\mathsf{NP}-$complete problem. One of the problems we want to solve (discussed later) is finding circuit lower bounds for the Ising model in statistical physics. This problem is known to be reducible to the MaxCut graph problem, another $\mathsf{NP}-$complete problem. This gives evidence that techniques cited in this survey may be helpful in proving sharp thresholds for a maximal cut of a graph, and thus finding lower bounds for the Ising model.

2.3 LMN Theorem

In their celebrated work, Linial, Mansour, and Nisan [LMN93] discovered the relationship between functions with low-degree Fourier expansions and circuit complexity. Henceforth, the following theorem will be known as the LMN theorem.

Theorem 9 ([LMN93] Main Lemma)**.** *Let f be a Boolean function on n variables computable by a boolean circuit of depth d and size M, and let t be any integer. Then,*

$$\sum_{S \subset [n], |S| > t} \hat{f}(S)^2 \leq 2M 2^{-t^{1/d}/20},$$

where $\hat{f}(S)$ denotes the Fourier transform of f at S.

The result we will prove uses an extension of the LMN theorem to arbitrary bias p, rather than the original LMN theorem. A key condition of the extension is that the Fourier weight bound depends on a term $\frac{1}{p(1-p)}$, so p which are extremely close to 0 or 1 end up giving a very poor bound. Informally, we can summarize this extended theorem as

Constant-depth, polynomial-size circuits with random p-biased inputs have a low-degree Fourier expansion, provided that the bias p is bounded away from 0 or 1.

Taken together, Friedgut's Theorem and the LMN Theorem almost provide a working proof to the fact that sharp thresholds imply circuit lower bounds. The reasoning roughly works as follows:

"Sharp threshold $\implies$ High-degree Fourier expansion $\implies$ Hard for small boolean circuits"

where the first implication comes from Friedgut's Theorem and the second implication comes from the LMN theorem. However, technical conditions requiring the bias to be bounded away from 0 or 1 for the extended LMN theorem to hold prevent these theorems from fitting together, though the high level idea still turns out to work. As will be shown, [GMZ23] mends the gap between these two to complete the chain.

3 Example Applications

Let us give a "boiled-down" statement of the main result of [GMZ23].

Theorem 10 ([GMZ23], Informal)**.** *Any Boolean circuit $\mathcal{C} : \{0,1\}^n \to \{0,1\}$ of depth $d = O(\log n / \log \log n)$ which exhibits a Δ-sharp threshold must have size at least* $\exp(\Delta^{1/d})$.

This statement is equivalent to saying that for a Boolean circuit with unbounded fan-in AND, OR, and NOT gates to compute a function with a sufficiently sharp threshold, it must either have sufficiently large depth $d = \Omega(\log n / \log \log n)$ or sufficiently large size $s = \Omega(\exp(\Delta^{1/d})$. A corollary of this theorem for AC_0 circuits follows from the requirement that any constant-depth circuit must have exponential size in $\Delta^{\Theta(1)}$.

Corollary 11 (AC_0 hardness of sharp thresholds)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function which has a sharp threshold with arbitrary critical point $p_c(n) \in (0,1)$ and sharpness $\Delta = n^{\Omega(1)}$. Then f cannot be computed on average around the critical window by any AC_0 circuit.*

We will now give a few examples of newly established results that come as simple applications of Corollary 11 to a variety of settings.

3.1 k-clique in random graphs

Definition 12. (The k-clique problem). Let $n, k \in \mathbb{N}$ with $k \leq n$. Let $N = \binom{n}{2}$. The k-clique Boolean function $f : \{0,1\}^N \to \{0,1\}$ equals to 1 if and only if the n-vertex graph $G \in \{0,1\}^N$ contains a k-clique. Each possible edge in G is parameterized by one of $\binom{n}{2}$ bits and is present if the value is 1 and is missing otherwise.

We use the following properties of the k-clique threshold, presented without proof.

Fact 13. A circuit $\mathcal{C}$ that computes the k-clique Boolean function on average for $k = \Theta(n)$ will exhibit a sharp threshold at $p_c = 1 - \Theta(1/n)$ with a critical window size $\varepsilon_n = o(n^{-3/2+\gamma})$ for a small $\gamma > 0$.

Note $\min\{p_c, 1-p_c\} = \Theta(1/n)$. Then, $\frac{\varepsilon_N}{\min\{p_c, 1-p_c\}} < n^{-1/2+\gamma}$ so the sharpness of the circuit is then $\Delta = n^{1/2-o(1)}$. Applying corollary 11 to this sharp threshold, we can see that AC_0 cannot compute the k-clique problem on average for $k = \Theta(n)$ around its critical threshold.

The existence of a k-clique is an example of a monotone graph property. Briefly, a monotone graph property is a property such that if it holds for any subgraph H of G, it also holds for G. As was first shown in [FK96], it is known that every monotone graph property has a sharp threshold. Therefore, we can expect to draw similar conclusions about AC_0 hardness for other monotone graph properties as well.

3.2 Random 2-SAT

Definition 14. (Random 2-SAT). A 2-SAT formula is a conjunction of m distinct clauses which are each the disjunction of two distinct literals in $\{x_1, \ldots, x_n, \overline{x_1}, \ldots, \overline{x_n}\}$. Let $C_1, \ldots, C_{2n(n-1)}$ be an enumeration of all possible 2-SAT clauses. Let $N = 2n(n-1)$. Let $X \in \{0,1\}^N$ be an encoding of a 2-SAT formula where $X_i = 1$ iff clause C_i is used in the formula. The 2-SAT Boolean function $f : \{0,1\}^N \to \{0,1\}$ is defined such that $f(X) = 1$ if and only if the corresponding 2-SAT formula of X is satisfiable.

We use the following properties of the random 2-SAT threshold which were first shown by [Bol+01], presented here without proof.

Fact 15. A circuit $\mathcal{C}$ that computes the 2-SAT function f on average will exhibit a sharp threshold at $p_c = \Theta(1/n)$ with a critical window size $\varepsilon_N = \Theta(n^{-4/3})$.

Note $\frac{\varepsilon_N}{\min\{p_c, 1-p_c\}} = \Theta(n^{-1/3})$ so the sharpness of the circuit is $\Delta = n^{1/3}$. Applying corollary 11 to this sharp threshold, we can see that AC_0 cannot compute the 2-SAT Boolean function on average around its critical threshold.

3.3 Statistical estimation of planted k-clique

This section gives an example of a circuit lower bound against a statistical estimation problem, showing that the result can be applied in a variety of settings.

Definition 16 (Hidden Subset Problem)**.** *Fix N and an arbitrary known prior distribution $\mathcal{P}$ over $\{0,1\}^N$. The Hidden Subset Problem is a task parameterized by prior $\mathcal{P}$. For some noise level $p \in (0,1)$ and a sample $S \sim \mathcal{P}$, you observe $S \vee X$ where $X \sim \mathbb{P}_p$. The goal is for you to construct* *an estimator function $A = A(N, \mathcal{P}, p) : \{0,1\}^N \to \{0,1\}^N$ which achieves exact recovery with a good probability of success, i.e. for sufficiently large N it holds that*

$$\Pr_{S\sim\mathcal{P}, X\sim\mathbb{P}_p}[A(S \vee X) = S] \geq 0.9$$

The larger the noise level p, the harder it will be to recover the hidden subset S. There is an information-theoretic threshold p_{IT} where for any noise level below the threshold there exists an estimator which achieves exact recovery, while for any noise level above the threshold no estimator can achieve exact recovery. The *All-or-Nothing* (AoN) phenomenon occurs when there is a sharp threshold present in the maximum value over all boolean functions A of $\Pr_{S\sim\mathcal{P},X\sim\mathbb{P}_p}[A(S \vee X) = S]$ as a function of the noise level p.

Definition 17 (Planted k-clique problem)**.** *Fix n vertices and some k. Let $N = \binom{n}{2}$ be used to index all possible undirected edges between the n vertices. The k-clique problem is the Hidden Subset Problem where $S \sim \mathcal{P}$ is the prior on $\{0,1\}^N$ which chooses k vertices at random and adds the edges between them to form a clique. Let $S \in \{0,1\}^{\binom{n}{2}}$ be the indicator of edges in this k-clique. Then the estimator is given a noisy sample $S \vee X$ where $X \sim \mathbb{P}_p$.*

This corresponds to a graph where S is a planted k-clique and every other edge appears independently with probability p, and the goal is to identify S exactly. We then use the following facts, presented without proof.

Fact 18. *For k where $(\log n)^{\omega(1)} = k \leq n^{1/3-\Omega(1)}$, the k-planted clique problem exhibits the All-or-Nothing phenomenon for exact recovery at some $p_{IT} = 1 - \Theta(\log n/k)$ and with critical window size $\varepsilon = \Theta(1/k^2)$.*

We obtain that the sharpness of the All-or-Nothing threshold is $\Delta = \Omega(k)$. For k where $(\log n)^{\omega(1)} = k \leq n^{1/3-\Omega(1)}$, this implies that no circuit in AC_0 can solve the planted k-clique statistical estimation problem.

4 Main Proof of [GMZ23]

4.1 High-level idea & sketch

The key motivation for this paper is to formalize a connection between the presence of a sharp threshold in a Boolean function and its corresponding circuit complexity. The authors provide some background for the result. First off, we are given some definitions for what exactly a sharp threshold is, in terms of a window size ε and a critical point p_c. In particular, Friedgut's paper [FB99] established that properties with sharp thresholds cannot be supported on low degree Fourier terms. However, the issue with this argument is that Friedgut's theorem primarily applies for cases when the bias p_n is close to zero, whereas bounds from the LMN theorem only work for $p_n = O(1)$ and degrade rapidly as the bias converges to zero.

This is addressed in [GMZ23] by working on a "debiasing" technique first introduced in [GJW22]. At a high level, they construct for any threshold p_c a "biasing" layer and apply this to a circuit. Whenever the original circuit exhibits a sharp threshold, the transformed circuit also exhibits a sharp threshold, but this time at a constant critical probability bounded away from 0 or 1. With this new critical probability, we can then apply other known results on this modified circuit. The application of this debiasing technique enables average-case around the critical point circuit lower bounds analysis for problems with critical thresholds p_c arbitrarily close to 0 or 1, for example the satisfiability of a random 2-SAT formula or the existence of a k-clique in a random graph.

As a result, the general approach for utilizing this result is as follows: we take a Boolean function f with a sharp threshold at p_c with a critical window size of ε_n. If some depth-d (where $d = O(\log n/\log\log n)$) circuit $\mathcal{C}$ agrees with f with high probability around the critical threshold p_c, then we can apply the main result to conclude a lower bound on the size s of $\mathcal{C}$ to be at least $\exp(\Delta^{1/d})$ where $\Delta = \frac{\min\{p_c, 1-p_c\}}{\varepsilon_N}$. When $\frac{\varepsilon_N}{\min\{p_c, 1-p_c\}} = O(N^{-c})$ for $c > 0$ this implies $f \notin \mathsf{AC}_0$.

The main proof will proceed as follows.

1. We will first reduce the argument to the $p_c \leq 1/2$ case since we can argue that having a Bernoulli parameter of $p_c > 1/2$ for the positive affirmation of an instance is equivalent to having another Bernoulli parameter $p_c' \leq 1/2$ for the negative instance.
2. We then introduce the concept of a "debiasing" layer Φ, which takes in $N \log 1/p_0$ input bits with a large constant bias $p_1 = \Theta(1)$ and maps them to N output bits with small bias $p_0 = o(1)$.
3. This is then appended to the bottom of a target circuit $\mathcal{C}$ and helps us provide input bits to $\mathcal{C}$ such that they "typecheck" while also enabling us to use a more amenable Bernoulli parameter p_1 that is conducive to analysis via theorems such as LMN.

Therefore, given this purpose, the onus is on us now to provide a valid construction and justification of Φ. We then will demonstrate a correspondence between the sharp threshold for circuit $\mathcal{C}$ and the sharp threshold for the circuit with the debiasing layer concatenated in front $\mathcal{C} \circ \Phi$. This then validates our debiasing layer construction with respect to our higher-level needs and purpose. Once we have a friendlier circuit $\mathcal{C} \circ \Phi$ to deal with, we can then apply well-established results from the analysis of boolean functions on $\mathcal{C} \circ \Phi$ to arrive at our desired size and depth trade-offs for the original $\mathcal{C}$.

4.2 Proof

Recall that for notation purposes, $\mathbb{E}_p f = \mathbb{E}_{\mathbf{X} \sim \mathbb{P}_p} f(\mathbf{X})$. Let us use a slightly different but equivalent definition of a sharp threshold to what we saw earlier:

Definition 19. *A function f exhibits a sharp threshold with window size $\varepsilon = \varepsilon_N \in (0,1)$ if for some critical threshold $0 < p_c = (p_c)_N \in (0,1)$ and jump size $\delta = \delta_N \in (0,1)$, for sufficiently large N, $|\mathbb{E}_{(1+\varepsilon)p_c} f - \mathbb{E}_{(1-\varepsilon)p_c} f| \geq \delta$.*

The main statement we will prove is as follows:

Theorem 20. *For some universal constants $c_1, c_2 > 0$ the following holds. Let $C : \{0,1\}^N \to \{0,1\}$ be a circuit of size $s = s_N$ and depth $d = d_N$. Suppose C exhibits a sharp threshold with window size $\varepsilon = \varepsilon_N$, jump size $\delta = \delta_N$ at critical threshold $p_c = (p_c)_N$.*

Let $\beta := \min\{p_c, 1-p_c\}$ and assume $\beta = N^{-\Omega(1)}$ and $\varepsilon = o((1-p_c)/\log 1/\beta)$. Then for some universal constants $c_1, c_2, c_3 > 0$, the following holds for sufficiently large N:

(1) Either,

$$d \geq \frac{1}{2\log\log N} \log\left[\frac{\delta(1-p_c)}{\varepsilon \log 1/\beta}\right] - 3,$$

i.e., the "depth is large"

(2) Or,

$$s \geq c_1 \exp\left(c_2 \left(\frac{\delta(1-p_c)}{\varepsilon \log 1/\beta}\right)^{1/(d+3)}\right),$$

i.e., the "size is large."

First, observe that for any circuit $\mathcal{C}$, we can define the circuit $\mathcal{C}'(X) := \mathcal{C}(\neg X)$, where $\mathcal{C}'$ has almost the same size and depth as $\mathcal{C}$ and satisfies

$$\mathop{\mathbb{E}}_{p}[\mathcal{C}'] = \mathop{\mathbb{E}}_{1-p}[\mathcal{C}]$$

We simply negate the inputs for circuit $\mathcal{C}$. Thus, if each input of $\mathcal{C}$ is assigned 1 independently with probability $p > 1/2$, then each input of $\mathcal{C}'$ will be assigned 1 with probability $p' \leq 1/2$. Since the size and depth are barely affected (we simply need to negate inputs), this transformation can be done without affecting our conclusions. Moreover, $\mathcal{C}$ has a sharp threshold at p_c with window size ε_N if and only if $\mathcal{C}'$ has a sharp threshold at $1-p_c$ with window size $\varepsilon'_N = \varepsilon_N p_c/(1-p_c)$. Lower bounds are thus shared between $\mathcal{C}$ and $\mathcal{C}'$ since they are equivalent up to negation. As a result, we only need to worry about when $p_c \leq 1/2$: if $\mathcal{C}$ has a critical point $p_c \geq 1/2$, then just consider $\mathcal{C}'$ which will have critical point $1-p_c \leq 1/2$. This allows us to generalize our proof to the $p_c \leq 1/2$ case.

Next, we introduce a tool that lets us move a critical point from close to 0 to a constant region by appending a depth-1 "debiasing layer" of bounded size to the bottom of our circuit:

Lemma 21. *Let $N \in \mathbb{N}$ and arbitrary $p_0 = (p_0)_N \leq 1/2$. There exists a depth-1 and size $O(N \log 1/p_0)$ circuit $\Phi : \{0,1\}^{N\lceil \log 1/p_0 \rceil} \to \{0,1\}^N$ and some $p_1 = (p_1)_N \in [1/2, 1/\sqrt{2})$ such that*

- $\Phi\left(Bern(p_1)^{\otimes N\lceil \log 1/p_0 \rceil}\right) \stackrel{d}{=} Bern(p_0)^{\otimes N}$
- *For any $0 < \gamma = \gamma_N = o\left(\frac{1}{\log 1/p_0}\right)$, there exists $0 < r_N = \Theta(\gamma \log \frac{1}{p_0}) = o(1)$ with*

$$\Phi\left(Bern(p_1 - r_n)^{\otimes N\lceil \log 1/p_0 \rceil}\right) \stackrel{d}{=} Bern(p_0(1-\gamma))^{\otimes N}$$

A proof of this lemma will follow in the next section, but using this lemma for our main proof, we note that condition 1 of the lemma essentially let's us shift a Bernoulli parameter p_0 to a "nicer" one p_1. Assuming $p_c \in (0, 1/2)$, we then choose $p_0 := (1+\varepsilon_N)p_c$. Then, for some $p_1 \in [1/2, 1/\sqrt{2})$, we get the following:

$$\mathop{\mathbb{E}}_{p_1}[\mathcal{C} \circ \Phi] = \mathop{\mathbb{E}}_{(1+\varepsilon_N)p_c}[\mathcal{C}].$$

In other words, we have successfully shifted away p_0 to a "nice" p_1 that enables us to use our tools for analysis.

Likewise, condition 2 of the lemma is targeted towards the preservation of sharpness in our circuit. Letting $\gamma = 1 - \frac{1-\varepsilon_N}{1+\varepsilon_N}$, we then obtain some $r_N = \Theta(\varepsilon_N \log 1/p_c) = o(1)$ such that

$$\mathop{\mathbb{E}}_{p_1 - r_N}[\mathcal{C} \circ \Phi] = \mathop{\mathbb{E}}_{(1-\varepsilon_N)p_c}[\mathcal{C}].$$

In other words, the region around our "nice" p_1 also acts as a sharp threshold, thus allowing us to connect these sharp thresholds with our desired size and depth trade off.

The interval $(1-\varepsilon_N)p_c$ to $(1+\varepsilon_N)p_c$ is the critical window of $\mathcal{C}$. Using the debiasing layer, the critical window is mapped to the interval $p_1 - r_N$ to p_1 for $\mathcal{C} \circ \Phi$. Over an interval of width $r_N = \Theta(\varepsilon_N \log 1/p_c)$, the value of $\mathbb{E}_p[\mathcal{C} \circ \Phi]$ will change from near 0 to near 1. We assume for simplicity that the change in $\mathbb{E}_p[\mathcal{C} \circ \Phi]$ in this range is exactly equal to 1. Since $p_1 \in [1/2, 1/\sqrt{2})$ and $r_N = \Theta(\varepsilon_N \log 1/p_c) = o(1)$, the range $(p_1 - r_N, p_1) \subseteq [1/3, 2/3]$. By the mean value theorem,

$$\max_{p \in [1/3,2/3]} \left| \frac{d}{dp} \mathop{\mathbb{E}}_p[\mathcal{C} \circ \Phi] \right| \geq 1/r_N = \frac{1}{\Theta\left(\varepsilon_N \log 1/p_c\right)}$$

We can then invoke two well-established results that are used here without further proof.

Lemma 22. *(One-sided Russo-Margulis lemma.) For any Boolean function $f : \{0,1\}^N \to \{0,1\}$, we have*

$$\left| \frac{d}{dp} \mathbb{E}_p[f] \right| \leq (p(1-p))^{-1} I_p(f)$$

where $I_p(f) = \sum_{k=1}^{n} I_p^k(f)$ is the *total influence* of f.

Lemma 23. *(Extension of LMN Theorem to arbitrary p.) For some constant $c_0 > 0$, if $\mathcal{C} : \{0,1\}^N \to \{0,1\}$ is a Boolean circuit of depth D and size S, then*

$$I_p(\mathcal{C}) \leq c_0 \left(\frac{10 \log S}{p(1-p)} \right)^{D+2}$$

These lemmas allow us to control the slope, $|\frac{d}{dp} \mathbb{E}_p[f]|$ when p is bounded away from 0 or 1, as a function of size S and depth D.

To finish the main proof, we let $\mathcal{C}$ be of size s and depth d. Circuit $\mathcal{C} \circ \Phi$ is thus of size $S = s + O(N \log 1/p_c)$ and depth $D = d+1$. Recall we showed earlier that

$$\max_{p \in [1/3,2/3]} |\frac{d}{dp} \mathop{\mathbb{E}}_p[\mathcal{C} \circ \Phi]| \geq \frac{1}{\Theta(\varepsilon_N \log 1/p_c)}$$

By chaining this together with the inequalities from the lemma we obtain that

$$\exp\left(O\left(\left(\frac{1}{\varepsilon_N \log 1/p_c} \right) \right)^{1/(d+3)} \right) \leq s + N \log 1/p_c$$

This inequality implies that at least one of the following must be true:

$$\exp\left(O\left(\left(\frac{1}{\varepsilon_N \log 1/p_c} \right) \right)^{1/(d+3)} \right) \leq s$$

or

$$\exp\left(O\left(\left(\frac{1}{\varepsilon_N \log 1/p_c} \right) \right)^{1/(d+3)} \right) \leq N \log 1/p_c$$

The depth-size trade-off of the original circuit $\mathcal{C}$ can at last be obtained via algebraic manipulation, yielding the desired circuit lower bounds. ■

4.3 Proof of Lemma 21

To finish the main proof, we will prove the key lemma surrounding the debiasing layer. To reiterate, the debiasing layer is a construction we'll use to make a circuit amenable to analysis. Its purpose is to shift the critical point away from 0 (which is a problem when applying the Friedgut and LMN results). Using the debiasing layer, we create a new circuit $\mathcal{C}^* = \mathcal{C} \circ \Phi$ such that we can have $\mathcal{C}^*$ use a "nice" critical point but still behave like our original $\mathcal{C}$. Furthermore, bounds on the depth and size of the debiasing layer means that any size-depth tradeoff from $\mathcal{C}^*$ is also inherent to $\mathcal{C}$.

The construction of the debiasing layer is as follows. We split input $X \in \{0,1\}^{N\lceil \log 1/p_0 \rceil}$ into N disjoint blocks of $\lceil \log 1/p_0 \rceil$ consecutive bits $X_1, ..., X_N \in \{0,1\}^{\lceil \log 1/p_0 \rceil}$. Then, for each block X_i, set $\Phi(X)_i = \bigwedge_{j=1}^{\lceil \log 1/p_0 \rceil} (X_i)_j$. In other words, we take the AND of every bit in block X_i. We can see that Φ can be modeled as a depth-1, size-$O(N \log 1/p_0)$ circuit.

Observe that the AND gate for each block X_i lights up as 1 if and only if all $\lceil \log 1/p_0 \rceil$ bits inside the block are 1. For any $p \in (0,1)$, each individual bit is 1 with probability p. Therefore, the probability the whole block X_i is 1 is $p^{\lceil \log 1/p_0 \rceil}$. In other words, the circuit Φ outputs N Bernoulli trials, each with success probability $p^{\lceil \log 1/p_0 \rceil}$. This is the key idea that lets us shift our Bernoulli parameter towards favorable terms.

We will now show the two desired statements:

1) $\Phi\left(\text{Bern}(p_1)^{\otimes N\lceil \log 1/p_0 \rceil}\right) \overset{d}{=} \text{Bern}(p_0)^{\otimes N}$

2) For any $0 < \gamma = \gamma_N = o\left(\frac{1}{\log 1/p_0}\right)$, there exists $0 < r_N = \Theta(\gamma \log \frac{1}{p_0}) = o(1)$ with

$$\Phi\left(\text{Bern}(p_1 - r_n)^{\otimes N\lceil \log 1/p_0 \rceil}\right) \overset{d}{=} \text{Bern}(p_0(1-\gamma))^{\otimes N}$$

Condition 1 is fairly straightforward. Let $p_1 = p_0^{1/\lceil \log 1/p_0 \rceil}$. We know by assumption $p_0 \leq 1/2$, so $p_1 \in [1/2, 1/\sqrt{2})$ as desired:

$$p_1 = 2^{-(\log \frac{1}{p_0})/\lceil \log \frac{1}{p_0} \rceil} \in [1/2, 1/\sqrt{2}) = \Theta(1)$$

We also know $p_1^{\lceil \log 1/p_0 \rceil} = p_0$. Therefore, we can now use the "nice" Bernoulli parameter p_1. Our blocks X_i then simulate inputs over the original parameter p_0 for the original circuit $\mathcal{C}$.

To show Condition 2, a technical statement saying that the sharpness is preserved (by relating window sizes r_n and γ_n of the modified and original circuit respectively), we set $p = p_1 - r_n$. Remember that the debiasing layer essentially took in $N\lceil \log 1/p_0 \rceil$ Bernoulli trials under parameter p and turned them into N Bernoulli trials under parameter $p^{\lceil \log 1/p_0 \rceil}$:

$$\Phi\left(\text{Bern}(p)^{\otimes N\lceil \log 1/p_0 \rceil}\right) \overset{d}{=} \text{Bern}(p^{\lceil \log 1/p_0 \rceil})^{\otimes N}$$

We find that $p = p_1 - r_n$ satisfies the solution $r_n > 0$ of

$$(p_1 - r_n)^{\lceil \log 1/p_0 \rceil} = p_0(1-\gamma)$$

where $r_n = \Theta(\gamma \log 1/p_0)$. This equation thus implies:

$$p_1 - r_n = p_1(1-\gamma)^{1/\lceil \log 1/p_0 \rceil}$$

Finally, because we know $\gamma \log 1/p_0 = o(1)$ and $p_1 = \Theta(1)$:

$$r_n = p_1 - p_1(1 - \Theta(\gamma \log 1/p_0)) = \Theta(\gamma \log 1/p_0)$$

We are thus able to preserve the sharp threshold by showing this relationship between window sizes for both $\mathcal{C}$ and $\mathcal{C}^* = \mathcal{C} \circ \Phi$. This finishes our proof of the lemma. ■

5 Problems & Future Directions

5.1 Application to the Ising Model

As previously mentioned, sharp thresholds have wide-ranging applications in statistical mechanics. A prime candidate for sharp thresholds is the Ising model. The Ising model consists of an ensemble of particles which can either take on a spin value of $+1$ or -1. Each particle has a magnetic interaction with its neighbor as well as with an external magnetic field. The Hamilton, or energy of such a system consisting of particles $S = \{s_i\}$ can be described as

$$H(S) = -\sum_{\langle i,j \rangle} J_{ij} s_i s_j - h \sum_i s_i,$$

where $\langle i, j \rangle$ denotes the particles s_j that are adjacent neighbors of s_i, J_{ij} is the strength of the local interaction between s_i and s_j, and h is the strength of the local magnetic field. If $J_{ij} > 0$, then neighboring particles tend to have *symmetric* spin, that is, both have spin of either $+1$ or -1. Otherwise, if $J_{ij} < 0$, neighboring particles tend to have antisymmetric spin. At low temperatures, the particles in the system behave either symmetrically or antisymmetrically depending on J. However, at high temperatures, the particles no longer follow such a regime; due to high entropy, the particles are in disarray and have spin values that are independent of each other. This suggests the existence of a *critical* temperature T_c at which there exists a sharp threshold of a phase transition.

There is an interesting reduction from the Ising problem in 2 dimensions without an external field to the MaxCut problem. Imagine the spin values of the particles $s_i \in S$ to be vertices of a graph G. Then, redefine the Ising model on S to be over the graph G:

$$H(S) = -\sum_{(i,j) \in E(G)} J_{ij} s_i s_j.$$

Recall that the goal is to find the configuration of S such that the Hamiltonian is minimized, otherwise known as a ground state. The idea is to now construct 2 subgraphs of G with different vertices. Denote the set of vertices V_+ such that $s_i \in V_+$ if and only $s_i = +1$, and likewise, V_- such that $s_i \in V_-$ if and only $s_i = -1$. Then, denote $\delta(V_+)$ to be the set of edges that connects the disjoint subsets V_+ and V_-. In other words, this is the cut of G. Then, define the weight of each edge (i, j) to be $W_{ij} = -J_{ij}$. The size of the cut is then given by

$$|\delta(V_+)| = \frac{1}{2} \sum_{(i,j) \in \delta(V_+)} W_{ij},$$

and the scaling factor of $1/2$ is used to compensate for the double counting of weights $W_{ij} = W_{ji}$

in an undirected graph. Then,

$$\begin{aligned} H(S) &= -\sum_{(i,j)\in E(V_+)} J_{ij} - \sum_{(i,j)\in E(V_-)} J_{ij} + \sum_{(i,j)\in \delta(V_+)} J_{ij} \\ &= -\sum_{(i,j)\in E(G)} J_{ij} + 2\sum_{(i,j)\in \delta(V_+)} J_{ij}. \end{aligned}$$

As the first term of the second equation does not depend on S as it encompasses the entire graph, then minimizing

$$\sum_{(i,j)\in \delta(V_+)} J_{ij}$$

minimizes $H(S)$. Since we defined edge weights W_{ij} to be the negation of J_{ij}, the maximum cut, that is, the maximum value of $|\delta(V_+)|$, also minimizes H, showing a reduction from the Ising problem to MaxCut.

Although the 2D Ising model has an exact solution for $h = 0$ due to Onsager in 1944, the above reduction shows that approximating the sharpness of the phase transition in an Ising model of 3 or more dimensions or in the presence of an external magnetic field through similar methods is a worthwhile problem to explore.

5.2 Smallest Possible Width of the Critical Window

Another open question we considered was finding further problems with sharp thresholds. Areas such as graph theory, for example, seemed promising. In particular, under the $G(n, p)$ Erdos-Renyi model, which is rich with sharp thresholds, we wanted to analyze other graph properties. For example, the detection of other subgraph structures (not just limited to cliques) was one interesting avenue. Ultimately, these problems involved narrowing our critical windows so that they were sharp enough to apply non-trivial bounds. If we can provide better sharp thresholds, then we can utilize the relationship between the window size ε_N and the circuit depth and size lower bound $S \geq \exp 1/\varepsilon_N^{1/d}$ to conclude meaningful circuit lower bounds for detecting these properties. Friedgut's line of work in sharp thresholds for random graph properties thresholds further promise to this idea in terms of the toolkit needed for such analysis.

5.3 Building on the Debiasing Technique

Another direction we thought about was generalizations of the "debiasing" layer technique that exactly transforms a p_1-biased Bernoulli distribution to a p_2-biased Bernoulli distribution. In particular, imagine that we establish that computing a function g is hard on average under some distribution $\mathcal{D}_h$ over $\{0,1\}^m$ and computing another function f under another distribution $\mathcal{D}_e$ over $\{0,1\}^n$ is easy on average. Suppose that you have access to a 'debiasing filter' Boolean circuit $\Phi : \{0,1\}^m \to \{0,1\}^n$ such that when it takes in samples from $\mathcal{D}_h$, it outputs samples distributed approximately as $\mathcal{D}_e$. Suppose you have a small circuit $\mathcal{C}$ which is able to compute f on average when you give it the easy distribution $\mathcal{D}_e$. If you feed $\mathcal{D}_h$ into $\mathcal{C} \circ \Phi$, it will also be able to compute the function defined by $k(x) := \mathcal{C} \circ \Phi(x)$ on average under $\mathcal{D}_h$. If we can show that function k is hard to compute under $\mathcal{D}_h$ (or somehow choose f and g in a way that lets g have high correlation with $k \approx f \circ \Phi$), then that implies that the probability mapping Φ must require a large circuit to compute in order to compensate for the small circuit used to compute f.

5.4 Other families of probability distributions

Additionally, we thought about how to obtain sharp thresholds for other continuously parameterized families of probability distributions besides the product Bernoulli distributions $\mathrm{Bern}(p)^{\otimes n}$. One way of obtaining a parameterized family of distributions over $\{0,1\}^n$ is to sample points with probabilities weighted according to an n-dimensional Gaussian distribution centered at one of the corners of the cube (or an arbitrary coordinate elsewhere). The continuously varying parameter $p \in [0,1]$ would be inversely proportional to the variance of the Gaussian distribution, so at $p = 0$ it would sample from a Gaussian with variance ∞ giving a uniform distribution, and then at $p = 1$ it would sample over a Gaussian with small variance, only giving points close to the center. Depending on what the center is, this could be another class of distribution families where sharp thresholds exist, especially if the Gaussian is centered on a hard instance of the problem, while the problem is easy under the uniform distribution. Note that a distribution that only samples from one (or a few) instances of the problem should be easy to compute because a circuit could just 'hard code' the answer into it. However, when the variance is increased slightly more it might become non-trivial to compute, and then when the variance is increased yet more it approaches close enough to a uniform distribution that it becomes easy to compute again, suggesting that hardness would lay only around the sharp threshold.

References

[Bol+01] Béla Bollobás et al. "The scaling window of the 2-SAT transition". In: *Random Structures amp; Algorithms* 18.3 (Apr. 2001), pp. 201–256. ISSN: 1098-2418. DOI: 10.1002/rsa.1006. URL: http://dx.doi.org/10.1002/rsa.1006.

[BT87] Béla Bollobás and Andrew G. Thomason. "Threshold functions". In: *Combinatorica* 7 (1987), pp. 35–38.

[DRT18] Hugo Duminil-Copin, Aran Raoufi, and Vincent Tassion. *Sharp phase transition for the random-cluster and Potts models via decision trees*. 2018. arXiv: 1705.03104 [math.PR].

[FB99] Ehud Friedgut and Jean Bourgain. "Sharp Thresholds of Graph Properties, and the k-Sat Problem". In: *Journal of the American Mathematical Society* 12.4 (1999), pp. 1017–1054. ISSN: 08940347, 10886834. URL: http://www.jstor.org/stable/2646096 (visited on 04/04/2024).

[FK96] Ehud Friedgut and Gil Kalai. "Every monotone graph property has a sharp threshold". In: 1996.

[Fri05] Ehud Friedgut. "Hunting for sharp thresholds". In: vol. 26. Jan. 2005, pp. 37–51. DOI: 10.1002/rsa.20042.

[GJW22] David Gamarnik, Aukosh Jagannath, and Alexander S. Wein. *Hardness of Random Optimization Problems for Boolean Circuits, Low-Degree Polynomials, and Langevin Dynamics*. 2022. arXiv: 2004.12063 [cs.CC].

[GMZ23] David Gamarnik, Elchanan Mossel, and Ilias Zadik. *Sharp Thresholds Imply Circuit Lower Bounds: from random 2-SAT to Planted Clique*. 2023. arXiv: 2311.04204 [cs.CC].

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. "Constant Depth Circuits, Fourier Transform, and Learnability". In: (1993), pp. 607–620. URL: https://dl.acm.org/doi/10.1145/174130.174138.

[ODo+05] R. O'Donnell et al. "Every decision tree has an influential variable". In: *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*. 2005, pp. 31–39. DOI: 10.1109/SFCS.2005.34.