Correlation Bounds Against Monotone NC^1

Benjamin Rossman*

National Institute of Informatics, Tokyo, Japan
Simons Institute for the Theory of Computation, Berkeley, USA
rossman@nii.ac.jp

Abstract

This paper gives the first correlation bounds under product distributions, including the uniform distribution, against the class mNC^1 of polynomial-size $O(\log n)$-depth monotone circuits. Our main theorem, proved using the pathset complexity framework introduced in [56], shows that the average-case k-CYCLE problem (on Erdős-Rényi random graphs with an appropriate edge density) is $\frac{1}{2} + \frac{1}{\mathrm{poly}(n)}$ hard for mNC^1. Combining this result with O'Donnell's hardness amplification theorem [43], we obtain an explicit monotone function of n variables (in the class mSAC^1) which is $\frac{1}{2}+n^{-1/2+\varepsilon}$ hard for mNC^1 under the uniform distribution for any desired constant $\varepsilon > 0$. This bound is nearly best possible, since every monotone function has agreement $\frac{1}{2} + \Omega(\frac{\log n}{\sqrt{n}})$ with some function in mNC^1 [44].

Our correlation bounds against mNC^1 extend smoothly to non-monotone NC^1 circuits with a bounded number of negation gates. Using Holley's monotone coupling theorem [30], we prove the following lemma: with respect to any product distribution, if a balanced monotone function f is $\frac{1}{2} + \delta$ hard for monotone circuits of a given size and depth, then f is $\frac{1}{2} + (2^{t+1} - 1)\delta$ hard for (non-monotone) circuits of the same size and depth with at most t negation gates. We thus achieve a lower bound against NC^1 circuits with $(\frac{1}{2} - \varepsilon)\log n$ negation gates, improving the previous record of $\frac{1}{6}\log\log n$ [7]. Our bound on negations is "half" optimal, since $\lceil \log(n+1) \rceil$ negation gates are known to be fully powerful for NC^1 [3, 21].

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases circuit complexity, average-case complexity

Digital Object Identifier 10.4230/LIPIcs.CCC.2015.392

1 Introduction

The majority of research in boolean circuit complexity is focused on restricted classes of circuits. Super-polynomial lower bounds are known in two basic settings: *bounded-depth circuits* (i.e. AC^0) [1, 24] and *monotone circuits* [51]. For another natural class, *deMorgan formulas* (tree-like circuits with fan-out 1), nearly cubic $n^{3-o(1)}$ lower bounds are known [28]. For bounded-depth circuits as well as deMorgan formulas, the state-of-the-art worst-case lower bounds (obtained in the 1980's and 90's) have recently been matched by tight *average-case lower bounds*, also known as *correlation bounds*, under the uniform distribution. It is now known that

- PARITY is $\frac{1}{2} + 2^{-\Omega(n/(\log S)^{d-1})}$ hard for depth-d (unbounded fan-in) circuits of size S [29],
- there is an explicit function in P which is $\frac{1}{2} + 2^{-\Omega(r)}$ hard for deMorgan formulas of size $n^{3-o(1)}/r^2$ [40].

* Supported by the JST ERATO Kawarabayashi Large Graph Project and the Simons Institute Research Fellowship.

(A boolean function f is said to be γ-*hard* for a class of circuits $\mathcal{C}$ under a distribution μ if $\mathbb{P}_{x\sim\mu}[\, f(x) = \mathfrak{C}(x)\,] \leq \gamma$ for every circuit $\mathfrak{C} \in \mathcal{C}$. By default μ is the uniform distribution and γ is typically expressed as $\frac{1}{2} + \delta$ or $1 - \delta$ where $\delta(n) \to 0$.)

In the monotone setting, there is an excellent knowledge of worst-case lower bounds: a long line of works [4, 8, 27, 37, 45, 48, 46, 50, 51] (among many others) have achieved separations between the monotone versions of nearly all the important complexity classes, as defined by Grigni and Sipser [26]. However, when it comes to average-case lower bounds under the uniform distribution or any product distribution, essentially nothing has been known; it is still open, for instance, whether any monotone function in NP is $1 - \frac{1}{\mathrm{poly}(n)}$ hard for polynomial-size monotone circuits. This represents a major gap in the basic understanding of monotone computation, given the importance of product distributions in the monotone setting. Product distributions are believed to be a natural source of hard instances for many monotone problem: k-SAT and k-CLIQUE are famously conjectured to be hard-on-average at an appropriate threshold density [38]. Product distributions are also significant in the real analysis of monotone functions (see the FKG inequality [22], the Bollobás-Thomason theorem [18], Friedgut's threshold theorem [23], etc.)

1.1 Previous Work

Many of the aforementioned *worst-case* lower bounds in the monotone setting can be viewed as average-case lower bounds under specific *non-product* distributions. To take one example, consider Razborov's celebrated lower bound for the k-CLIQUE function [51]. Let μ denote the distribution on n-vertex graphs which, half of the time, is a uniform random k-clique, and the other half is a uniform random $(k-1)$-coclique (i.e. complete $(k-1)$-partite graph). The result of [51] (together with the quantitative improvement [4]) shows that, for all $3 \leq k \leq n^{1/4}$, k-CLIQUE is $\frac{1}{2} + n^{-\Omega(\sqrt{k})}$ hard under μ for the class mP of polynomial-size monotone circuits (for $k \leq \log n$, this bound improves to $\frac{1}{2} + n^{-\Omega(k)}$). (Correlation bounds under similar (non-product) distributions were recently obtained for monotone classes within mP [20, 25], strengthening previous worst-case separations among these classes.)

Toward the goal of worst-case lower bounds, this distribution μ has a very sensible property: it is supported entirely on *minterms* (minimal 1-instances, i.e., k-cliques) and *maxterms* (maximal 0-instances, of which $(k-1)$-cocliques are a subset). Thus μ exploits monotonicity in the strongest possible way. However, there is something backwards about μ: every 1-instance has Hamming weight $\binom{k}{2}$ $(\leq \sqrt{n})$, which is less the minimum Hamming weight $\binom{k-1}{2}\big(\frac{n}{k-1}\big)^2$ $(\geq n^2/2)$ of any 0-instance. It follows that k-CLIQUE is equivalent under μ to the *anti-monotone* threshold function $\mathrm{THR}_{<n^2/2}$ (which is 1 on graphs with fewer than $n^2/2$ edges). Therefore, even though k-CLIQUE is hard under μ for monotone circuits, it is easy under μ for polynomial-size circuits with a single negation gate (in fact, $\mathrm{THR}_{<n^2/2}$ is computable by polynomial-size formulas with a single negation [3]).

This discomfort was addressed in work of Amano and Maruoka [7], who extended the k-CLIQUE lower bound of [4, 51] to polynomial-size circuits with $\frac{1}{6}\log\log n$ negation gates by considering a modified distribution μ' (a certain convex combination, over various $\ell \in \{k, \ldots, n\}$, of ℓ-cliques and $(k-1)$-cocliques supported on sets of size ℓ). While the core of the proof in [7] is still a monotone circuit lower bound for cliques vs. cocliques, this result contributed an insight that sufficiently strong lower bounds against monotone circuits imply lower bounds against negation-limited boolean circuits (we capitalize on this insight in Lemma 1.3).

A more natural distribution for the average-case analysis of k-CLIQUE is given by the Erdős-Rényi random graph $G(n,p)$. Here we take p $(= p(n,k))$ to be the unique "$\frac{1}{2}$-threshold"

such that $\mathbb{P}[\, G(n,p)$ contains a k-clique$\,] = \frac{1}{2}$. (Note that $G(n,p)$ is a product distribution on $\{0,1\}^{\binom{n}{2}}$.) Karp [38] famously conjectured that k-CLIQUE is hard-on-average under $G(n,p)$ (in the regime $k \approx 2\log n$). Previous work of the author [54] confirmed this conjecture in the (non-monotone) AC^0 setting by showing that k-CLIQUE is $\frac{1}{2} + n^{-\Omega(k)}$ hard under $G(n,p)$ for AC^0 (polynomial-size constant-depth circuits) for all $k \leq \log^{1/2} n$. Follow-up work of the author [55] combined the technique of [54] with the "approximation method" framework of Razborov [51] to prove a correlation bound against monotone circuits under a different distribution ν on n-vertex graphs: half of the time, ν is $G(n,p)$ plus a uniform random planted k-clique, and the other half ν is $G(n,2p)$ conditioned on being k-clique-free. The result of [55] shows that k-CLIQUE is $\frac{1}{2} + n^{-\Omega(k)}$ hard under ν for mP for all $k \leq \log^{1/2} n$. While ν is (morally speaking) similar to $G(n,p)$, it is unfortunately not a product distribution and suffers the same shortcoming as μ: the 0-instances and 1-instances under ν, although no longer minterms and maxterms, are nevertheless separable with high probability by an anti-monotone threshold function (in this case $\mathrm{THR}_{<\frac{3}{2}\binom{n}{2}p}$). It was left as an open problem in [55] to prove a correlation bound against mP for k-CLIQUE under $G(n,p)$. In the present paper, we do not succeed in this task; however, we prove a correlation bound against a significant subclass of mP (mNC^1) for a different monotone graph property (k-CYCLE) under an appropriate product distribution.

1.2 Our Results

Our main theorem is a correlation bound for the average-case k-CYCLE problem against the class mNC^1 of polynomial-size $O(\log n)$-depth monotone circuits (equivalently, polynomial-size monotone formulas). Rather than the standard Erdős-Rényi random graph, we find it convenient to restrict attention to "C_k-partite" input graphs with kn vertices and kn^2 potential edges (Def. 4.2). For the average-case analysis of k-CYCLE, we consider a random C_k-partite graph, denoted Γ, in which each potential edge is independently included with probability p where p is the unique "$\frac{1}{2}$-threshold" such that $\mathbb{P}[\, \Gamma$ contains a k-cycle$\,] = \frac{1}{2}$. (Note: $p \sim c_k/n$ for a constant c_k depending on k.) A monotone function f on kn^2 variables is said to compute k-CYCLE on Γ with *advantage* δ if $\mathbb{P}[\, f(\Gamma) = k\text{-CYCLE}(\Gamma)\,] \geq \frac{1}{2} + \delta$.

▶ **Theorem 1.1** (Main Theorem). *For all $k \leq \log\log n$, if a monotone formula computes k-CYCLE on Γ with advantage $n^{-1/2+c}$, then it has size $n^{\Omega(c \log k)}$. In particular, $\log\log n$-CYCLE is $\frac{1}{2} + n^{-1/2+o(1)}$ hard under Γ for monotone formulas of size $n^{o(\log\log\log n)}$ (and hence for mNC^1).*

The lower bound in Theorem 1.1 is essentially tight, since k-CYCLE is computable (in the worst-case) by monotone formulas of size $n^{O(\log k)}$, as well as by polynomial-size $O(\log k)$-depth monotone circuits with semi-unbounded fan-in (i.e. binary AND gates and unbounded OR gates). This places k-CYCLE in the class mSAC^1. (In terms of space complexity, k-CYCLE is computable in NL as well as Ave-L, as defined in [12].) Theorem 1.1 thus gives a very sharp average-case separation of mNC^1 from higher complexity classes.

As a corollary of Theorem 1.1, we obtain nearly optimal correlation bounds against mNC^1 under the *uniform distribution*. Note that the random graph Γ, while a product distribution, is not the uniform distribution on $\{0,1\}^{kn^2}$; moreover, the correlation bound in Theorem 1.1 is only $\frac{1}{2} + (kn^2)^{-1/4+o(1)}$ in terms of the input size kn^2. Nevertheless, using O'Donnell's hardness amplification theorem [43], we have the following result:

▶ **Corollary 1.2.** *For every $\varepsilon > 0$, there is an explicit monotone function of N variables (in the class mSAC^1) which is $\frac{1}{2} + N^{-1/2+\varepsilon}$ hard for mNC^1 under the uniform distribution.*

This function is the composite function[1] TRIBES $\otimes \log\log n$-CYCLE $\otimes$ p-BIAS on $N = \text{poly}(n)$ variables where

- p-BIAS : $\{0,1\}^n \to \{0,1\}$ is any n-term monotone DNF with exactly $\lceil p2^n \rceil$ satisfying assignments,[2]
- TRIBES : $\{0,1\}^{n^c} \to \{0,1\}$ is the "tribes" function of Ben-Or and Linial [13] on n^c variables, where c $(= \Omega(1/\varepsilon))$ is a sufficiently large constant.

See O'Donnell's paper [43] for background on the hardness amplification theorem which yields Corollary 1.2 from Theorem 1.1. We only remark that all results in [43], while stated in terms of the class NP, apply equally to mNC^1. This observation relies on the fact that MAJ $\in$ mNC^1 [3, 63] (where MAJ is the majority function); this is essential for the application of Implagliazzo's "hard-core set" theorem [31, 39], which is a main tool in [43].

The correlation bound in Corollary 1.2 is nearly best possible under the uniform distribution, as O'Donnell and Wimmer [44] have shown that every monotone function $\{0,1\}^n \to \{0,1\}$ has agreement $\frac{1}{2} + \Omega(\frac{\log n}{\sqrt{n}})$ with one of functions $0, 1, x_1, \ldots, x_n, \text{MAJ}$. Since these functions are all in mNC^1, it follows that no monotone function is $\frac{1}{2} + o(\frac{\log n}{\sqrt{n}})$ hard for mNC^1. Corollary 1.2 shows that this correlation bound is nearly achieved by an explicit monotone function. (By counting arguments, there exist (non-explicit) monotone functions achieving similar correlation bounds [9, 36].)

Finally, we extend these results to non-monotone circuits with a bounded number of negation gates, by means of a general lemma on correlation bounds under product distribution. In fact, our observation applies to the broader class of distributions μ on $\{0,1\}^n$ which satisfy the *FKG lattice condition* [22] if

$$\mu(x)\mu(y) \leq \mu(x \wedge y)\mu(x \vee y) \qquad \text{for all } x, y \in \{0,1\}^n. \tag{1}$$

Note that every product distribution satisfies (1) with equality.

▶ **Lemma 1.3.** *Let μ be a distribution which satisfies the FKG lattice condition* (1) *and let f be a monotone function which is balanced under μ (i.e. $\mathbb{E}_\mu(f) = \frac{1}{2}$). If f is $\frac{1}{2} + \delta$ hard under μ for monotone circuits of a given size and depth, then f is $\frac{1}{2} + (2^{t+1} - 1)\delta$ hard under μ, up to the same size and depth, for (non-monotone) circuits with t negation gates.*

Via Lemma 1.3, the correlation bound of Corollary 1.2 extends to NC^1 circuits with $(\frac{1}{2} - \varepsilon)\log n$ negation gates.

▶ **Corollary 1.4.** *For every $\varepsilon > 0$, there is an explicit function in* mSAC^1 *which is $\frac{1}{2} + o(1)$ hard for* NC^1 *circuits with $(\frac{1}{2} - \varepsilon)\log n$ negations under the uniform distribution.*

Corollary 1.4 is half optimal, in the sense that NC^1 circuits with $\lceil \log(n+1) \rceil$ negations are known to be equivalent to full NC^1 by well-known results of Markov [42] and Fischer [21] (again using the fact that MAJ $\in$ NC^1). This improves a previous $\frac{1}{6}\log\log n$ lower bound of Amano and Maruoka [7] on the negation-limited complexity of an explicit monotone function $\{0,1\}^n \to \{0,1\}$ (however, unlike Corollary 1.4, the result of [7] applies to polynomial-size circuits of unbounded depth). For multi-output monotone functions $\{0,1\}^n \to \{0,1\}^n$, Jukna

[1] For boolean functions $h : \{0,1\}^l \to \{0,1\}$ and $g : \{0,1\}^m \to \{0,1\}$, the composite function $g \otimes h : (\{0,1\}^l)^m \to \{0,1\}$ is defined by $(g \otimes h)(y_1, \ldots, y_m) = g(h(y_1), \ldots, h(y_m))$.

[2] It is an easy exercise to show that there is an n-term monotone DNF with exactly m satisfying assignments for every $m \in \{0, \ldots, 2^n\}$. Note that p-BIAS generates a single p'-biased bit from the uniform distribution on $\{0,1\}^n$ where $p \leq p' < p + 2^{-n}$.

[34] proved a worst-case lower bound of $\log n - O(\log \log n)$. (There is an extensive literature on negation-limited complexity; see Chapter 10 of [35] and papers [11, 14, 16, 36, 64, 67] besides those already mentioned.)

1.3 Overview

We present an outline of the paper, highlighting the main ideas in the proof of Theorem 1.1.

Persistent Minterms

In Section 3 we introduce the key notion of *persistent minterms* of a monotone function f under an increasing sequence of monotone restrictions. Formally, we consider the sequence of monotone functions $f^{\vee\rho_0} \leq f^{\vee\rho_1} \leq \cdots \leq f^{\vee\rho_m}$ where $\rho_0 \leq \rho_1 \leq \cdots \leq \rho_m$ are elements in $\{0,1\}^n$ and $f^{\vee\rho_i}(x) := f(x \vee \rho_i)$. An element $x \in \{0,1\}^n$ of Hamming weight $|x| = k$ is called a *d-persistent minterm* of f under $\vec{\rho}$ if it is a common minterm of $\binom{d+k-1}{k-1}$ many functions $f^{\vee\rho_i}$.

Persistent minterms behave like ordinary minterms under operations $\vee$ and $\wedge$ (Lemma 3.4). In additional, persistent minterms have the desirable property of being "noise-insensitive" in a certain sense. Suppose $\xi^{(1)}, \ldots, \xi^{(m)}$ are independent samples from a distribution of random "noise" over $\{0,1\}^n$. If we now define ρ_i by $\xi^{(1)} \cup \cdots \cup \xi^{(i)}$, then every persistent minterm is noise-insensitive, insofar as it has survived at least one hit of random noise. This translates into a lemma to the effect that every monotone function whatsoever has "few" persistent minterms with high probability (Lemma 5.13).

Average-Case k-CYCLE

In Section 4 we consider the average-case k-CYCLE problem on the random graph Γ (i.e. the p-biased product distribution on $\{0,1\}^{kn^2}$ for appropriate $p \sim 1/n$). We introduce an auxiliary random graph Ξ_ℓ consisting of ℓ ($\ll \sqrt{n}$) independent paths of length $k-1$. Crucially, Ξ_ℓ lives "within the variance" of the random graph Γ, in the sense that Γ and $\Gamma \cup \Xi_\ell$ have small total variation distance. Roughly speaking, we are able to show: if a monotone function f has correlation $\gg \ell k^2/\sqrt{n}$ with k-CYCLE under Γ, then a non-negligible fraction (at least $1/\sqrt{n}$) of k-cycles are persistent minterms of f with respect to random noise Ξ_ℓ (Lemma 4.5).

Pathset Complexity

In Section 5 we present the pathset complexity framework and state a lower bound proved in [56]. Very roughly speaking, for a subgraph $A = (V_A, E_A)$ of the k-cycle, a *pathset over* A is a set of isomorphic copies of A embedded (as "sections") in $V_A \times [n]$. *Pathset complexity* is a (formula-like) complexity measure on pathsets with respect to operations $\cup$ and $\bowtie$ (union and join). Crucially, pathsets are subject to a collection of density constraints called *smallness*; this bottleneck accounts for the high complexity of constructing pathsets beyond a certain size.

The pathset complexity framework was introduced in [56] for the purpose of separating formula-size and circuit-size within AC^0. The technique is tailored to the formula complexity of the (virtually equivalent) average-case k-STCONN / k-CYCLE problems. The paper [56] proves a lower bound of $n^{\Omega(\log k)}$ on the pathset complexity of any sufficiently dense pathset over the k-path / k-cycle (Theorem 5.8).

Our correlation bound against mNC^1 (Theorem 1.1) is proved by reduction to this pathset complexity lower bound. Given a monotone formula which correlates well with k-CYCLE under Γ, we define (random) pathsets at all gates of the formula in terms of persistent minterms. We show (Lemma 5.13) that all of these pathsets satisfy the smallness constraint with high probability. In this way, we are able to obtain a formula-size lower bound from pathset complexity. The proof of Theorem 1.1 is given in Section 6; proofs of various lemmas are included in appendices (due to space limitation, two appendices which appear in the full version of this paper have been removed from this version).

2 Preliminaries

Let $\mathbb{N} = \{0, 1, 2, \dots\}$. For $n \in \mathbb{N}$, let $[n] = \{1, \dots, n\}$. We write $\ln(\cdot)$ for the natural logarithm and $\log(\cdot)$ for the base-2 logarithm. For random variables X and Y, notation $X \overset{\mathrm{d}}{=} Y$ expresses "X and Y are identically distributed".

▶ **Definition 2.1** (Monotone Functions, Minterms, Monotone Restrictions). $\mathbb{B}_n^+$ denotes the lattice of monotone (non-decreasing) boolean functions $\{0,1\}^n \to \{0,1\}$. f, g represent functions in $\mathbb{B}_n^+$. $f \leq g$ denotes $f(x) \leq g(x)$ for all $x \in \{0,1\}^n$. For $f \in \mathbb{B}_n^+$ and $x \in \{0,1\}^n$, we say that x is a *minterm* of f if $f(x) = 1$ and $f(x') = 0$ for all $x' < x$. The set of minterms of f is denoted by $\mathcal{M}(f)$. (Note that $\mathcal{M}(\cdot)$ gives a bijection from $\mathbb{B}_n^+$ to anti-chains in $\{0,1\}^n$.) For $f \in \mathbb{B}_n^+$ and $\rho \in \{0,1\}^n$, we denote by $f^{\vee\rho}$ be the monotone function $f^{\vee\rho}(x) := f(x \vee \rho)$. (Note that $f \leq f^{\vee\rho}$.) In this context, we view $\rho \in \{0,1\}^n$ as a "monotone restriction" which sets some variables to 1 (namely, $i \in [n]$ such that $\rho_i = 1$) and leaves the remaining variables unset.

▶ **Lemma 2.2** (Minterm Lemma). *For all $f, g \in \mathbb{B}_n^+$,*

$$\mathcal{M}(f \vee g) \subseteq \mathcal{M}(f) \cup \mathcal{M}(g), \qquad \mathcal{M}(f \wedge g) \subseteq \{x \vee y : x \in \mathcal{M}(f),\, y \in \mathcal{M}(g)\}. \tag{2}$$

In other words, every minterm of $f \vee g$ is a minterm of f or a minterm of of g, and every minterm of $f \wedge g$ is the disjunction of a minterm of f and a minterm of g. (This is easy to see, for instance, by thinking of the DNF representations of f and g.)

▶ **Definition 2.3** (Monotone Formulas). A *monotone formula* on n variables is a finite rooted binary tree whose leaves (inputs) are labeled by elements of $[n] \cup \{\underline{0}, \underline{1}\}$ and whose non-leaves (gates) are labeled $\wedge$ or $\vee$. (In this paper all AND and OR gates have fan-in 2.) Every monotone formula Φ on n variables computes a monotone function in $\mathbb{B}_n^+$ (in the usual way). For $x \in \{0,1\}^n$, we write $\Phi(x)$ for the value of the monotone function computed by Φ on input x. $\mathrm{Sub}(\Phi)$ denotes the set of (syntactic) sub-formulas of Φ. For example, if Φ is the formula $\Psi \wedge \Psi$, then $\mathrm{Sub}(\Phi)$ contains both (left and right) copies of Ψ. $\mathrm{Leaves}(\Phi)$ ($\subseteq \mathrm{Sub}(\Phi)$) denotes the set of leaves in Φ. The *(leaf) size* of Φ is defined as $\mathrm{size}(\Phi) := |\mathrm{Leaves}(\Phi)|$ ($= \frac{1}{2}(|\mathrm{Sub}(\Phi)| + 1)$). The *depth* of Φ is its height as a tree (where a single leaf has depth 0).

3 Persistent Minterms

▶ **Notation 3.1.** For a partially ordered set L and $m \in \mathbb{N}$, we denote by $\mathrm{Seq}_{\leq}^m(L)$ the set of non-decreasing chains $\vec{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_m)$ such that $\lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_m$. (We will consistently index coordinates of $\vec{\lambda}$ by λ_s, λ_t where $0 \leq s \leq t \leq m$.)

▶ **Notation 3.2.** For $d, k \in \mathbb{N}$, let $\left\langle {d \atop k} \right\rangle := \binom{d+k-1}{k-1}$. Note the identity $\left\langle {d \atop k} \right\rangle = \left\langle {d-1 \atop k} \right\rangle + \left\langle {d \atop k-1} \right\rangle$.

▶ **Lemma 3.1.** *For all $d, k \geq 1$ and $\vec{a} \in \mathrm{Seq}^k_{\leq}(\mathbb{R})$, if $a_k - a_0 > \left\langle {d \atop k} \right\rangle$, then $a_j - a_{j-1} > \left\langle {d-1 \atop j} \right\rangle$ for some $j \in \{1, \ldots, k\}$.*

Proof. By induction on k: assuming $a_k - a_0 > \left\langle {d \atop k} \right\rangle$, either $a_k - a_{k-1} > \left\langle {d-1 \atop k} \right\rangle$, in which case the lemma is satisfied with $j = k$, or else $a_{k-1} - a_0 = (a_k - a_0) - (a_k - a_{k-1}) > \left\langle {d \atop k} \right\rangle - \left\langle {d-1 \atop k} \right\rangle = \left\langle {d \atop k-1} \right\rangle$, in which case we use the induction hypothesis for $(a_0, \ldots, a_{k-1}) \in \mathrm{Seq}^{k-1}_{\leq}(\mathbb{R})$. ◀

By the same basic induction, we have:

▶ **Lemma 3.2.** *For all $d, m \geq 1$ and $\vec{x} \in \mathrm{Seq}^m_{\leq}(\{0,1\}^n)$, if $m \geq \left\langle {d \atop |x_m|} \right\rangle$, then $x_s = x_t$ for some $0 \leq s \leq t \leq m$ with $t - s \geq \left\langle {d-1 \atop |x_s|} \right\rangle$.*

Proof. Suppose $m \geq \left\langle {d \atop |x_m|} \right\rangle$ and let $\ell := \min\{s \geq 0 : |x_s| = |x_m|\}$. If $m - \ell \geq \left\langle {d-1 \atop |x_m|} \right\rangle$, then we are done. Otherwise, $\ell - 1 = (m-1) - (m-\ell) \geq (\left\langle {d \atop |x_m|} \right\rangle - 1) - (\left\langle {d-1 \atop |x_m|} \right\rangle - 1) \geq \left\langle {d \atop |x_m|-1} \right\rangle \geq \left\langle {d \atop |x_{\ell-1}|} \right\rangle$ and we use the induction hypothesis for the truncated sequence $(x_0, \ldots, x_{\ell-1}) \in \mathrm{Seq}^{\ell-1}_{\leq}(\{0,1\}^n)$. ◀

▶ **Definition 3.3** (Persistent Minterms)**.** For $\vec{f} \in \mathrm{Seq}^m_{\leq}(\mathbb{B}^+_n)$ and $x \in \{0,1\}^n$, we say that x is a *d-persistent minterm* of $\vec{f}$ if it is a common minterm of f_s and f_t (i.e. $x \in \mathcal{M}(f_s) \cap \mathcal{M}(f_t)$) for some $0 \leq s \leq t \leq m$ such that $t - s \geq \left\langle {d \atop |x|} \right\rangle$. The set of d-persistent minterms of $\vec{f}$ is denoted by $\mathcal{M}_d(\vec{f})$.

We have defined persistent minterms in a general way for sequences $f_0 \leq f_1 \leq \cdots \leq f_m$ of monotone functions. However, we will be interested in the persistent minterms of an *individual* monotone function f under a sequence $\rho_0 \leq \rho_1 \leq \cdots \leq \rho_d$ of monotone restrictions. (Eventually, we will utilize this notion by choosing *random* restrictions $\vec{\rho}$.)

▶ **Notation 3.3.** For $f \in \mathbb{B}^+_n$ and $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\{0,1\}^n)$, let $\mathcal{M}^{\vec{\rho}}_d(f) := \mathcal{M}_d(f^{\vee\rho_0} \leq f^{\vee\rho_1} \leq \cdots \leq f^{\vee\rho_m})$.

▶ **Lemma 3.4** (Persistent Minterm Lemma)**.** *For all $f, g \in \mathbb{B}^+_n$ and $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\{0,1\}^n)$ and $d, m \geq 1$,*

$$\mathcal{M}^{\vec{\rho}}_d(f \vee g) \subseteq \mathcal{M}^{\vec{\rho}}_{d-1}(f) \cup \mathcal{M}^{\vec{\rho}}_{d-1}(g), \tag{3}$$

$$\mathcal{M}^{\vec{\rho}}_d(f \wedge g) \subseteq \left\{x \vee y : x \in \mathcal{M}^{\vec{\rho}}_{d-1}(f),\, y \in \mathcal{M}^{\vec{\rho}}_{d-1}(g)\right\}. \tag{4}$$

The proof, which we include in Appendix A, is straightforward (in particular, we show (4) using Lemma 3.2). We will return to persistent minterms in Section 5.2.

4 Average-Case k-CYCLE

We depart from the setting of monotone functions $\{0,1\}^n \to \{0,1\}$ (on n variables) and instead consider a domain $\mathcal{G} \cong \{0,1\}^{k^2 n}$ of graphs (with kn^2 possible edges). Before defining $\mathcal{G}$, let us first clarify the role of k:

▶ **Definition 4.1.** Throughout the rest of this paper, let $k = k(n) \in \mathbb{N}$ be an arbitrary parameter (i.e. function of n) subject to $k \leq \log\log n$.

The constraint $k \leq \log\log n$ is due to the factor of $(1/2)^{O(2^k)}$ in Theorem 5.8. Outside of this theorem, all other lemmas in this paper hold for a larger range of k.

▶ **Definition 4.2** (K-Partite Graphs)**.** All graphs in this paper are finite directed graphs without isolated vertices. Formally, a *graph* is a pair $G = (V_G, E_G)$ where V_G is a finite set and $E_G \subseteq V_G \times V_G$ and $V_G = \bigcup_{vw \in E_G} \{v, w\}$. As a special case, $\varnothing$ denotes the *empty graph* with $V_\varnothing = E_\varnothing = \emptyset$ (the empty set). K denotes the *k-cycle graph* with $V_K = \{v_0, v_1, \ldots, v_{k-1}\}$ and $E_K = \{v_0v_1, v_1v_2, \ldots, v_{k-1}v_0\}$. (We never write these indices explicitly, instead always writing $v \in V_K$, $vw \in E_K$ or $e \in E_K$.) We denote by $\mathcal{G}$ $(= \mathcal{G}(k,n))$ the set of *K-partite graphs* G satisfying

$$V_G \subseteq \{v^{(i)} : v \in V_K,\, i \in [n]\}, \qquad E_G \subseteq \{v^{(i)}w^{(j)} : vw \in E_K,\, i,j \in [n]\}.$$

Here $v^{(i)}$ and $v^{(i)}w^{(j)}$ are just a friendly notation for ordered pairs (v,i) and $((v,i),(w,j))$. In the context of functions $\mathcal{G} \to \{0,1\}$, we identify $\mathcal{G}$ with the hypercube $\{0,1\}^{kn^2}$.

▶ **Definition 4.3** (k-CYCLE)**.** For $G \in \mathcal{G}$, we say that G is a *k-cycle* if G is isomorphic to K. Note that G is a k-cycle if and only if there exists a function $\iota : V_K \to [n]$ such that $E_G = \{v^{(\iota(v))}w^{(\iota(w))} : vw \in E_K\}$. We say that G has a k-cycle if it contains a k-cycle as a subgraph. k-CYCLE denotes the monotone function $\mathcal{G} \to \{0,1\}$ which takes value 1 on G if, and only if, G has a k-cycle.

We are interested in the average-case analysis of k-CYCLE. For this purpose, we define three random graphs needed to state our main lemma (on the noise-invariance of minterms of k-CYCLE).

▶ **Definition 4.4** (Random Graphs Γ, $\circlearrowright$ and Ξ_ℓ)**.** Let Γ, $\circlearrowright$ and Ξ_ℓ be the following (independent) random graphs in $\mathcal{G}$:

- Let Γ be the (K-partite, Erdős-Rényi) random graph in $\mathcal{G}$ which includes each potential edge independently with probability p (i.e. $\mathbb{P}[\,\Gamma = G\,] = p^{|E_G|}(1-p)^{kn^2-|E_G|}$) where $p = p(k,n)$ $(\sim (\ln 2)^{1/k}/n)$ is the unique "$\frac{1}{2}$-threshold" such that $\mathbb{P}[\,k\text{-CYCLE}(\Gamma) = 1\,] = \frac{1}{2}$.
- Let $\circlearrowright$ be a uniform random k-cycle in $\mathcal{G}$. For $e \in E_K$, we write $\circlearrowright^{-e}$ for the graph obtained from $\circlearrowright$ by deleting the edge in $\circlearrowright$ corresponding to e. Note that $\circlearrowright^{-e}$ is a path of length $k-1$.
- For $\ell \in \mathbb{N}$, let Ξ_ℓ be the random graph $\circlearrowright_1^{-e_1} \cup \cdots \cup \circlearrowright_\ell^{-e_\ell}$ where $\circlearrowright_1, \ldots, \circlearrowright_\ell$ are uniform random k-cycles and $e_1, \ldots, e_\ell$ are uniform random edges in E_K. Equivalently, Ξ_ℓ is the union of ℓ uniform random paths of length $k-1$.

We will only consider values of ℓ much less than $\sqrt{n}$, where random paths $\circlearrowright_1^{-e_1} \cup \cdots \cup \circlearrowright_\ell^{-e_\ell}$ are likely to be vertex-disjoint. The letter Ξ is mnemonic for this situation.

We state the key lemma of this section, whose proof is included in the full paper.

▶ **Lemma 4.5.** *For every monotone function $f : \mathcal{G} \to \{0,1\}$ and $\ell \in \mathbb{N}$, if*

$$\mathop{\mathbb{P}}_{\Gamma}\big[\, f(\Gamma) = k\text{-CYCLE}(\Gamma)\,\big] \geq \frac{1}{2} + \frac{C(\ell+1)k^2}{\sqrt{n}} \tag{5}$$

where $C > 0$ is a universal constant, then there exists $G \in \mathcal{G}$ such that

$$\mathop{\mathbb{P}}_{\Xi_\ell}\Big[\, \mathop{\mathbb{P}}_{\circlearrowright}\big[\,\circlearrowright \in \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G \cup \Xi_\ell})\,\big] \geq n^{-1/2}\Big] \geq n^{-1/2}. \tag{6}$$

Lemma 4.5 says the following: (in the case $\ell = 0$) if a monotone function f has correlation $\gg k^2/\sqrt{n}$ with k-CYCLE on Γ, then there exists a graph G such that a non-negligible fraction of k-cycles are minterms of $f^{\cup G}$. Moreover, (for $\ell \geq 1$) if this correlation is $\gg \ell k^2/\sqrt{n}$, then these minterms are "Ξ_ℓ-noise-invariant" in the following sense: with probability $\geq n^{-1/2}$ over Ξ_ℓ, at least $1/\sqrt{n}$ fraction of k-cycles are common minterms of $f^{\cup G}$ and $f^{\cup G \cup \Xi_\ell}$.

The tie-in to persistent minterms is clear. Let $d \in \mathbb{N}$ and suppose ℓ is a multiple of $m := \binom{d}{k}$. We may generate Ξ_ℓ as a union of independent $\Xi^{(1)}_{\ell/m}, \ldots, \Xi^{(m)}_{\ell/m}$. Writing ρ_s for the partial union $\Xi^{(1)}_{\ell/m} \cup \cdots \cup \Xi^{(s)}_{\ell/m}$, we have a non-decreasing sequence $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\mathcal{G})$. Notice that every k-cycle which is a common minterm in $\mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G \cup \Xi_\ell})$ is a d-persistent minterm in $\mathcal{M}^{\vec{\rho}}_d(f^{\cup G})$. (This observation shows up in the proof of Theorem 1.1 in Section 6.)

5 Pathset Complexity

5.1 The Basic Framework

We present the definitions required to state the pathset complexity lower bound (Theorem 5.8), which we use in our main theorem (Theorem 1.1). For background on these definitions (key examples, upper bounds, etc.), the reader is referred to the paper [56].

▶ **Definition 5.1** (Pattern Graphs). Subgraphs of K are called *pattern graphs* and designated by letters A, B, C. Recall that graphs (by definition in this paper) have no isolated vertices. Therefore, pattern graphs $A \subseteq K$ are in one-to-one correspondence with subsets $E_A \subseteq E_K$.

An important parameter of pattern graphs $A \subseteq K$ is the number $|V_A| - |E_A|$. Note that every pattern graph, other than K itself, is a disjoint union of paths. Therefore,

$$A \neq K \;\Rightarrow\; |V_A| - |E_A| = |\{\text{connected components of } A\}|. \tag{7}$$

Also note that $0 \leq |V_A| - |E_A| \leq k/2$ and $|V_A| - |E_A| = 0 \;\Leftrightarrow\; A \in \{\varnothing, K\}$.

▶ **Definition 5.2** (Sections). For $A \subseteq K$, an *A-section* is a graph $A' \in \mathcal{G}$ such that $E_{A'} = \{v^{(\iota(v))}w^{(\iota(w))} : vw \in E_A\}$ for some function $\iota : V_A \to [n]$. (As a special case, the empty graph $\varnothing$ is the unique $\varnothing$-section.) The set of all A-sections is denoted by $\mathcal{G}_A$. As a matter of notation, we consistently write A-sections using primes (A', A'', etc.) Every $A' \in \mathcal{G}_A$ is isomorphic to A via the projection $v^{(i)} \mapsto v$.

We have already encountered K-sections and $K \setminus \{e\}$-sections in the guise of random graphs $\circlearrowleft$ and $\circlearrowleft^{-e}$. (Note that K-sections are the same as k-cycles in $\mathcal{G}$ (Def. 4.2).)

▶ **Definition 5.3** (Pathsets). For $A \subseteq K$, subsets of $\mathcal{G}_A$ (i.e. sets of A-sections) are called *pathsets over A*. As a special case, note that there are two distinct pathsets over $\varnothing$: the empty set $\emptyset$ and the "identity" pathset $\{\varnothing\}$. Every non-empty pathset $\mathcal{A}$ is a pathset over a unique $A \subseteq K$, which we call its *underlying pattern graph*. Pathsets over A, B, C, K are consistently designated by the respective calligraphic letters $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{K}$. The *density* of a pathset $\mathcal{A}$ is defined by

$$\mathsf{density}(\mathcal{A}) := |\mathcal{A}| \,/\, n^{|V_A|} = \mathop{\mathbb{P}}_{A' \in \mathcal{G}_A}[\, A' \in \mathcal{A}\,]. \tag{8}$$

▶ **Definition 5.4** (Joins). For any two pathsets $\mathcal{A}$ and $\mathcal{B}$, the *join* $\mathcal{A} \bowtie \mathcal{B}$ is the pathset (over $A \cup B$) defined by

$$\mathcal{A} \bowtie \mathcal{B} := \big\{C' \in \mathcal{G}_{A \cup B} : C' = A' \cup B' \text{ for some } A' \in \mathcal{A} \text{ and } B' \in \mathcal{B}\big\}. \tag{9}$$

Note that $\bowtie$ is an associative, commutative and idempotent operation on pathsets. Moreover, $\emptyset$ and $\{\varnothing\}$ act as the zero and identity: $\mathcal{A} \bowtie \emptyset = \emptyset$ and $\mathcal{A} \bowtie \{\varnothing\} = \mathcal{A}$. (Taking the view of a pathset $\mathcal{A}$ as a "V_A-ary relation" (i.e. a subset of $[n]^{V_A}$), $\bowtie$ is the standard relational join operation.)

▶ **Definition 5.5** (Restrictions). For pathsets $\mathcal{A}$ and $\mathcal{B}$, we say that $\mathcal{B}$ is a *restriction* of $\mathcal{A}$, denoted $\mathcal{B} \preceq \mathcal{A}$, if $B \subseteq A$ and there exists $\overline{B}' \in \mathcal{G}_{A \setminus B}$ such that $\mathcal{B} = \{B' \in \mathcal{G}_B : B' \cup \overline{B}' \in \mathcal{A}\}$. $\mathcal{B}$ is a *proper restriction* of $\mathcal{A}$, denoted $\mathcal{B} \prec \mathcal{A}$, if $\mathcal{B} \preceq \mathcal{A}$ and $\mathcal{B} \neq \mathcal{A}$.

▶ **Definition 5.6** (Smallness). For $\varepsilon > 0$, a pathset $\mathcal{A}$ is *ε-small* if it satisfies

$$\mathsf{density}(\mathcal{B}) \leq \varepsilon^{|V_B|-|E_B|} \text{ for all } \mathcal{B} \preceq \mathcal{A}. \tag{10}$$

The set of ε-small pathsets (over all pattern graphs) is denoted by $\mathcal{P}_\varepsilon$.

Note that every pathset over $\varnothing$ or K is ε-small, since $|V_\varnothing| - |E_\varnothing| = |V_K| - |E_K| = 0$. ε-smallness is obviously preserved under subsets, as well as under restrictions: if $\mathcal{A} \in \mathcal{P}_\varepsilon$, then $\mathcal{A}_0 \in \mathcal{P}_\varepsilon$ and $\mathcal{B} \in \mathcal{P}_\varepsilon$ for every $\mathcal{A}_0 \subseteq \mathcal{A}$ and $\mathcal{B} \preceq \mathcal{A}$. Somewhat less obvious is the fact that ε-smallness is also preserved under joins (Lemma 5.5 of [56]): if $\mathcal{A}, \mathcal{B} \in \mathcal{P}_\varepsilon$, then $\mathcal{A} \bowtie \mathcal{B} \in \mathcal{P}_\varepsilon$.

▶ **Definition 5.7** (Pathset Complexity). For any $\varepsilon > 0$ ("smallness parameter"), *pathset complexity* is the function $\chi_\varepsilon : \mathcal{P}_\varepsilon \to \mathbb{N}$ defined inductively as follows:

- (base case) $\chi_\varepsilon(\emptyset) = \chi_\varepsilon(\{\varnothing\}) = 0$ and $\chi_\varepsilon(\mathcal{A}) = |\mathcal{A}|$ if $|E_A| = 1$,
- (induction case) if $|E_A| \geq 2$, then

$$\chi_\varepsilon(\mathcal{A}) := \min_{(\mathcal{B}_i, \mathcal{C}_i)_i} \textstyle\sum_i \chi_\varepsilon(\mathcal{B}_i) + \chi_\varepsilon(\mathcal{C}_i)$$

where $(\mathcal{B}_i, \mathcal{C}_i)_i$ ranges over all sequences of ε-small pathsets $\mathcal{B}_i, \mathcal{C}_i \in \mathcal{P}_\varepsilon$ such that $B_i, C_i \subsetneqq A$ and $B_i \cup C_i = A$ and $\mathcal{A} \subseteq \bigcup_i \mathcal{B}_i \bowtie \mathcal{C}_i$.

In other words, for the (induction case) we consider all possible *coverings* of $\mathcal{A}$ by joins of *ε-small* pathsets over *proper* subgraphs of A.

It is clear from this definition that pathset complexity satisfies the following inequalities:

- (monotonicity) $\chi_\varepsilon(\mathcal{A}_1) \leq \chi_\varepsilon(\mathcal{A}_2)$ for all $\mathcal{A}_1 \subseteq \mathcal{A}_2 \in \mathcal{P}_\varepsilon$,
- (sub-additivity) $\chi_\varepsilon(\mathcal{A}_1 \cup \mathcal{A}_2) \leq \chi_\varepsilon(\mathcal{A}_1) + \chi_\varepsilon(\mathcal{A}_2)$ for all $\mathcal{A}_1, \mathcal{A}_2$ such that $\mathcal{A}_1 \cup \mathcal{A}_2 \in \mathcal{P}_\varepsilon$,
- (join inequality) $\chi_\varepsilon(\mathcal{A} \bowtie \mathcal{B}) \leq \chi_\varepsilon(\mathcal{A}) + \chi_\varepsilon(\mathcal{B})$ for all $\mathcal{A}, \mathcal{B} \in \mathcal{P}_\varepsilon$.

In fact, these three inequalities provide a *dual characterization* of pathset complexity: χ_ε is the unique pointwise maximal function $\mathcal{P}_\varepsilon \to \mathbb{N}$ which satisfies (base case), (monotonicity), (sub-additivity) and (join inequality).

The following lower bound on pathset complexity was shown in [56]:

▶ **Theorem 5.8** (Pathset Complexity Lower Bound). *For every pathset $\mathcal{K}$ over K,*

$$\chi_\varepsilon(\mathcal{K}) \geq (1/2)^{O(2^k)} \cdot (1/\varepsilon)^{\frac{1}{6}\log k} \cdot \mathsf{density}(\mathcal{K}). \tag{11}$$

Theorem 5.8 corresponds to Theorem 5.8 of [56]. We remark that the lower bound proved in [56] applies more broadly to pathsets $\mathcal{A} \in \mathcal{P}_\varepsilon$ over any pattern graph $A \subseteq K$:

$$\chi_\varepsilon(\mathcal{A}) \geq (1/2)^{O(2^{|E_A|})} \cdot (1/\varepsilon)^{\frac{1}{6}\log(\mathsf{length}(A)) + |V_A| - |E_A|} \cdot \mathsf{density}(\mathcal{A}) \tag{12}$$

where $\mathsf{length}(A)$ equals the number of edges in the largest connected component of A. In fact, (12) follows from an even more general lower bound for *pathset complexity with respect to patterns* (Theorem 8.3 of [56]). However, for the application in this paper, we only require the bound (11) for pathsets over K.

5.2 Pathsets of Persistent Minterms

In order to prove formula-size lower bounds using pathset complexity, we associate pathsets with all monotone formulas on kn^2 variables. The pathsets need to satisfy certain consistency conditions; moreover, these (random) pathsets must be ε-small (with high probability). Persistent minterms and random restrictions Ξ_ℓ accomplish both of these goals. In this subsection, we show how to define appropriate pathsets using persistent minterms; we deal with ε-smallness in the next subsection.

▶ **Definition 5.9** (Pathsets $\mathcal{M}_A(f)$ and $\mathcal{P}_A^{\vec{\rho}}(\Phi)$)**.** For a monotone function $f : \mathcal{G} \to \{0,1\}$ and $A \subseteq K$, let $\mathcal{M}_A(f) := \mathcal{G}_A \cap \mathcal{M}(f)$ be the pathset of A-sections which are minterms of f. For a monotone formula Φ and $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\mathcal{G})$ and $A \subseteq K$, the pathset $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ (over A) is defined by

$$\mathcal{P}_A^{\vec{\rho}}(\Phi) := \mathcal{G}_A \cap \mathcal{M}^{\vec{\rho}}_{\mathrm{depth}(\Phi)}(\Phi). \tag{13}$$

That is, the $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ is the set of A-sections which are $\mathrm{depth}(\Phi)$-persistent minterms of Φ under the sequence $\vec{\rho}$.

Unpacking definitions, for all $A \neq \varnothing$, we have the expression

$$\mathcal{P}_A^{\vec{\rho}}(\Phi) = \bigcup_{\substack{0\leq s\leq t\leq m:\\ t-s\geq \left\langle \frac{\mathrm{depth}(\Phi)}{|E_A|}\right\rangle}} \left(\mathcal{M}_A(\Phi^{\cup\rho_s}) \cap \mathcal{M}_A(\Phi^{\cup\rho_t})\right) \subseteq \bigcup_{0\leq s\leq m-1} \left(\mathcal{M}_A(\Phi^{\cup\rho_s}) \cap \mathcal{M}_A(\Phi^{\cup\rho_{s+1}})\right). \tag{14}$$

The following lemma is a straightforward consequence of (14).

▶ **Lemma 5.10.** *If $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ is <u>not</u> ε-small, then $\mathcal{M}_A(\Phi^{\cup\rho_s}) \cap \mathcal{M}_A(\Phi^{\cup\rho_{s+1}})$ is <u>not</u> (ε/m)-small for some $s \in \{0,\ldots,m-1\}$.*

Proof. Assume $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ is not ε-small. By Def. 5.6, there exists a restriction $\mathcal{B} \preceq \mathcal{P}_A^{\vec{\rho}}(\Phi)$ such that $\mathsf{density}(\mathcal{B}) > \varepsilon^{|V_B|-|E_B|}$. (Note that $A, B \notin \{\varnothing, K\}$.) By Def. 5.5, there exists an $(A \setminus B)$-section $\overline{B}' \in \mathcal{G}_{A\setminus B}$ such that $\mathcal{B} = \{B' \in \mathcal{G}_B : B' \cup \overline{B}' \in \mathcal{P}_A^{\vec{\rho}}(\Phi)\}$. Writing $\mathcal{A}_s$ for $\mathcal{M}_A(\Phi^{\cup\rho_s}) \cap \mathcal{M}_A(\Phi^{\cup\rho_{s+1}})$, we have $\mathcal{P}_A^{\vec{\rho}}(\Phi) \subseteq \bigcup_{s=0}^{m-1} \mathcal{A}_s$ by (14), hence $\mathcal{B} \subseteq \bigcup_{s=0}^{m-1}\{B' \in \mathcal{G}_B : B' \cup \overline{B}' \in \mathcal{A}_s\}$. It follows that there exists $s \in \{0,\ldots,m-1\}$ such that

$$\mathsf{density}(\{B' \in \mathcal{G}_B : B' \cup \overline{B}' \in \mathcal{A}_s\}) \geq \mathsf{density}(\mathcal{B})/m > \varepsilon^{|V_B|-|E_B|}/m \geq (\varepsilon/m)^{|V_B|-|E_B|}.$$

Since $\{B' \in \mathcal{G}_B : B' \cup \overline{B}' \in \mathcal{A}_s\} \preceq \mathcal{A}_s$, we conclude that $\mathcal{A}_s$ is not (ε/m)-small. ◀

We next restate the Persistent Minterm Lemma 3.4 in terms of pathsets $\mathcal{P}_A^{\vec{\rho}}(\Phi)$.

▶ **Lemma 5.11.** *For all monotone functions f, g and monotone formulas Φ, Ψ and $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\mathcal{G})$ and $A \subseteq K$,*

$$\mathcal{P}_A^{\vec{\rho}}(\Phi \vee \Psi) \subseteq \mathcal{P}_A^{\vec{\rho}}(\Phi) \cup \mathcal{P}_A^{\vec{\rho}}(\Psi), \tag{15}$$

$$\mathcal{P}_A^{\vec{\rho}}(\Phi \wedge \Psi) \subseteq \bigcup_{B,C\subseteq A\,:\,B\cup C=A} \mathcal{P}_B^{\vec{\rho}}(\Phi) \bowtie \mathcal{P}_C^{\vec{\rho}}(\Psi). \tag{16}$$

The main lemma of this subsection gives the key relationship between pathset complexity and formula size and depth.

▶ **Lemma 5.12.** *Suppose Φ is a monotone formula and $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\mathcal{G})$ such that pathsets $\mathcal{P}_A^{\vec{\rho}}(\Psi)$ are ε-small for all $\Psi \in \mathrm{Sub}(\Phi)$ and $A \subseteq K$. Then*

$$\chi_\varepsilon(\mathcal{P}_K^{\vec{\rho}}(\Phi)) \leq 2^{O(k^2)} \cdot \mathrm{depth}(\Phi)^k \cdot \mathrm{size}(\Phi). \tag{17}$$

Although the statement of Lemma 5.12 might appear complicated, the proof is actually quite simple. The derivation of (17) uses only Lemma 5.11 and the key properties (monotonicity), (sub-additivity) and (join inequality) of pathset complexity. The proof of Lemma 5.12, which is essentially the same as Lemma 6.7 in [56], is included in Appendix B.

5.3 Smallness Lemma

In the last subsection, we defined pathsets $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ (for an arbitrary sequence $\vec{\rho} \in \mathrm{Seq}_{\leq}^m(\mathcal{G})$) and showed a relationship between pathset complexity and formula size under the condition that all of the relevant pathsets are ε-small. The next lemma give a means of establishing ε-smallness.

▶ **Lemma 5.13.** *For every monotone function $f : \mathcal{G} \to \{0,1\}$ and $A \subseteq K$ and $\ell \in \mathbb{N}$ and $\varepsilon > 0$,*

$$\mathop{\mathbb{P}}_{\Xi_\ell}\big[\,\mathcal{M}_A(f) \cap \mathcal{M}_A(f^{\cup \Xi_\ell}) \text{ is } \underline{\text{not}}\ \varepsilon\text{-small}\,\big] \leq (2n)^k \cdot \exp\big(-\Omega(\varepsilon\ell/k^2)\big). \tag{18}$$

The main tools in the proof of Lemma 5.13 (included in the full version of this paper) are Janson's Inequality [33] and the sunflower-plucking technique of Razborov [51].

6 Proof of Theorem 1.1 (Correlation Bound for k-CYCLE)

Proof of Theorem 1.1. Let $k \leq \log\log n$ and suppose Φ is a monotone formula such that

$$\mathop{\mathbb{P}}_{\Gamma}\big[\,f(\Gamma) = k\text{-CYCLE}(\Gamma)\,\big] = \tfrac{1}{2} + n^{-1/2+c}.$$

Our goal is to show the lower bound $\mathrm{size}(\Phi) = n^{\Omega(c \log k)}$.

Using the fact that $n^{O(\log k)}$ is an upper bound on the size of monotone formulas for k-CYCLE (together with the "formula balancing lemma" [59, 66]: every monotone formula of size S is equivalent to a monotone formula of depth $O(\log S)$) we may assume that $\mathrm{size}(\Phi) = n^{O(\log k)}$ and $\mathrm{depth}(\Phi) = O(\log k \cdot \log n)$. However, for purposes of this proof, it is enough for us to assume much weaker upper bounds $\mathrm{size}(\Phi) \leq \exp(n^{1/k})$ and $\mathrm{depth}(\Phi) \leq n^{1/k}$. We also assume $c = \Omega(1/\log k)$, since otherwise there is nothing to prove.

We set parameters m, ℓ, ε as follows:

$$m := \left\langle {\mathrm{depth}(\Phi) \atop k} \right\rangle \left(= \tbinom{\mathrm{depth}(\Phi)+k-1}{k-1}\right), \qquad \ell := n^{c/2}, \qquad \varepsilon := n^{-c/4}. \tag{19}$$

Note that $m = O(\mathrm{depth}(\Phi))^k = n^{o(c)}$. We have $n^{-1/2+c} = \omega((m\ell+1)k^2/\sqrt{n})$, that is, Φ satisfies the hypothesis (5) of Lemma 4.5 (for all sufficiently large n). Therefore, by Lemma 4.5, there exists $G \in \mathcal{G}$ such that

$$\mathop{\mathbb{P}}_{\Xi_{\ell m}}\Big[\,\mathop{\mathbb{P}}_{\circlearrowleft}\big[\,\circlearrowleft \in \mathcal{M}(\Phi^{\cup G}) \cap \mathcal{M}(\Phi^{\cup G \cup \Xi_{m\ell}})\,\big] \geq n^{-1/2}\,\Big] = \Omega(n^{-1/2}). \tag{20}$$

Fixing any such G, we now generate *random* $\vec{\boldsymbol{\rho}} \in \mathrm{Seq}_{\leq}^m(\mathcal{G})$ as follows:

- Let $\Xi_\ell^{(1)}, \ldots, \Xi_\ell^{(m)}$ be independent random copies of Ξ_ℓ.
- For $s \in \{0, \ldots, m\}$, let $\boldsymbol{\rho}_s := G \cup (\Xi_\ell^{(1)} \cup \cdots \cup \Xi_\ell^{(s)})$.

By our choice of $m = \left\langle {\mathrm{depth}(\Phi) \atop k} \right\rangle$ and Def. 5.9 of $\mathcal{P}_K^{\vec{\boldsymbol{\rho}}}(\Phi)$ (see (14)), we have

$$\mathcal{P}_K^{\vec{\boldsymbol{\rho}}}(\Phi) = \mathcal{M}_K(\Phi^{\cup \boldsymbol{\rho}_0}) \cap \mathcal{M}_K(\Phi^{\cup \boldsymbol{\rho}_m}).$$

Since $\circlearrowleft$ is uniform in $\mathcal{G}_K$, it follows (by definition (8) of $\mathsf{density}(\cdot)$) that

$$\mathsf{density}(\mathcal{P}_K^{\vec{\boldsymbol{\rho}}}(\Phi)) = \mathop{\mathbb{P}}_{\circlearrowleft}\left[\circlearrowleft \in \mathcal{M}(\Phi^{\cup\boldsymbol{\rho}_0}) \cap \mathcal{M}(\Phi^{\cup\boldsymbol{\rho}_m})\right].$$

Since $\boldsymbol{\rho}_0 = G$ and $\boldsymbol{\rho}_m \stackrel{\mathrm{d}}{=} G \cup \Xi_{m\ell}$, we see that (20) is equivalent to

$$\mathop{\mathbb{P}}_{\vec{\boldsymbol{\rho}}}\left[\mathsf{density}(\mathcal{P}_K^{\vec{\boldsymbol{\rho}}}(\Phi)) \geq n^{-1/2}\right] = \Omega(n^{-1/2}). \tag{21}$$

We next observe that, with all-but-negligible probability $1 - n^{-\omega(1)}$, pathsets $\mathcal{P}_A^{\vec{\boldsymbol{\rho}}}(\Psi)$ are all ε-small:

$$\begin{aligned}
&\mathop{\mathbb{P}}_{\vec{\boldsymbol{\rho}}}\Big[\bigvee_{\Psi\in\mathrm{Sub}(\Phi)}\bigvee_{\varnothing\subset A\subset K}\mathcal{P}_A^{\vec{\boldsymbol{\rho}}}(\Psi) \text{ is } \underline{\text{not}}\ \varepsilon\text{-small}\Big] && \text{(Lemma 5.10)} \qquad (22)\\
&\quad\leq \sum_{\Psi\in\mathrm{Sub}(\Phi)}\sum_{\varnothing\subset A\subset K}\sum_{0\leq s\leq m-1}\mathop{\mathbb{P}}_{\vec{\boldsymbol{\rho}}}\left[\mathcal{M}_A(\Psi^{\cup\boldsymbol{\rho}_s})\cap\mathcal{M}_A(\Psi^{\cup\boldsymbol{\rho}_{s+1}}) \text{ is } \underline{\text{not}}\ (\varepsilon/m)\text{-small}\right]\\
&\quad\leq \mathrm{size}(\Phi)\cdot 2^k\cdot m\cdot\exp\left(-\Omega(\varepsilon\ell/k^2m)\right) && \text{(Lemma 5.13)}\\
&\quad= \exp(O(n^{1/k}))\cdot\exp(-n^{c/4-o(c)}) && (\mathrm{size}(\Phi)\leq\exp(n^{1/k}))\\
&\quad= n^{-\omega(1)} && (c=\Omega(1/\log k)).
\end{aligned}$$

As the upshot of (21) and (22), (for all sufficiently large n) there exists $\vec{\rho}\in\mathrm{Seq}_{\leq}^{m}(\mathcal{G})$ satisfying both

- **Dense**$(\vec{\rho})$, the event that $\mathsf{density}(\mathcal{P}_K^{\vec{\rho}}(\Phi)) \geq n^{-1/2}$, and
- **Small**$(\vec{\rho})$, the event that pathsets $\mathcal{P}_A^{\vec{\rho}}(\Psi)$ are ε-small for all $\Psi\in\mathrm{Sub}(\Phi)$ and $A\subseteq K$.

Fixing any such $\vec{\rho}$, we complete the reduction to our pathset complexity lower bound (using $k\leq\log\log n$):

$$\begin{aligned}
\mathrm{size}(\Phi) &\geq \mathrm{depth}(\Phi)^{-k}\cdot 2^{-O(k^2)}\cdot\chi_\varepsilon(\mathcal{P}_K^{\vec{\rho}}(\Phi)) && (\textbf{Small}(\vec{\rho}) \text{ and Lemma 5.12})\\
&\geq n^{-O(1)}\cdot\chi_\varepsilon(\mathcal{P}_K^{\vec{\rho}}(\Phi)) && (\mathrm{depth}(\Phi)\leq n^{1/k})\\
&\geq n^{-O(1)}\cdot 2^{-O(2^k)}\cdot(1/\varepsilon)^{\frac{1}{6}\log k}\cdot\mathsf{density}(\mathcal{P}_K^{\vec{\rho}}(\Phi)) && \text{(Theorem 5.8)}\\
&= n^{(c/24)\log k - O(1)} && (\textbf{Dense}(\vec{\rho})).
\end{aligned}$$

Therefore, $\mathrm{size}(\Phi) = n^{\Omega(c\log k)}$ as required. ◀

Acknowledgements. I would like to thank Li-Yang Tan for valuable feedback on various technical aspects of this paper, Rahul Santhanam and Osamu Watanabe for helpful discussions, Siyao Guo for debugging a confusing typo, and the anonymous referees for many useful suggestions.

References

1 Miklós Ajtai. Σ_1^1 formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

2 Miklós Ajtai and Yuri Gurevich. Monotone versus positive. *J. ACM*, 34:1004–1015, 1987.

3 Miklós Ajtai, János Komlós, and Endre Szemerédi. An $O(n\log n)$ sorting network. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing*, pages 1–9. ACM, 1983.

4 Noga Alon and Ravi B. Boppana. The monotone circuit complexity of boolean functions. *Combinatorica*, 7(1):1–22, 1987.

5 Noga Alon and Joel Spencer. *The Probabilistic Method, 3rd Edition*. John Wiley, 2008.

6 Kazuyuki Amano and Akira Maruoka. Potential of the approximation method. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 431–440. IEEE, 1996.

7 Kazuyuki Amano and Akira Maruoka. A superpolynomial lower bound for a circuit computing the clique function with at most $(1/6)\log\log n$ negation gates. *SIAM Journal on Computing*, 35(1):201–216, 2005.

8 Alexander E. Andreev. On a method for obtaining lower bounds for the complexity of individual monotone functions. 31(3):530–534, 1985.

9 Alexander E. Andreev, Andrea E.F. Clementi, and José D.P. Rolim. Optimal bounds for the approximation of boolean functions and some applications. *Theoretical Computer Science*, 180(1):243–268, 1997.

10 Richard Arratia, Larry Goldstein, and Louis Gordon. Poisson approximation and the chen-stein method. *Statistical Science*, pages 403–424, 1990.

11 Robert Beals, Tetsuro Nishino, and Keisuke Tanaka. On the complexity of negation-limited boolean networks. *SIAM Journal on Computing*, 27(5):1334–1347, 1998.

12 Shai Ben-David, Benny Chor, Oded Goldreich, and Michel Luby. On the theory of average case complexity. *Journal of Computer and system Sciences*, 44(2):193–219, 1992.

13 Michael Ben-Or and Nathan Linial. Collective coin flipping. *Randomness and Computation*, 5:91–115, 1990.

14 Stuart J. Berkowitz. On some relationships between monotone and nonmonotone circuit complexity. Technical report, Department of Computer Science, University of Toronto, Canada, Toronto, Canada, 1982.

15 Avrim Blum, Carl Burch, and John Langford. On learning monotone boolean functions. In *Foundations of Computer Science, 1998. Proceedings. 39th Annual Symposium on*, pages 408–415. IEEE, 1998.

16 Norbert Blum. On negations in boolean networks. In *Efficient Algorithms*, pages 18–29. Springer, 2009.

17 Andrej Bogdanov and Luca Trevisan. Average-case complexity. *Theoretical Computer Science*, 2(1):1–106, 2006.

18 Béla Bollobás and A.G. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

19 Nader H. Bshouty and Christino Tamon. On the Fourier spectrum of monotone functions. *Journal of the ACM (JACM)*, 43(4):747–770, 1996.

20 Yuval Filmus, Toniann Pitassi, Robert Robere, and Stephen A Cook. Average case lower bounds for monotone switching networks. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 20, page 54, 2013.

21 Michael J. Fischer. The complexity of negation-limited networks – a brief survey. In *Automata Theory and Formal Languages 2nd GI Conference Kaiserslautern, May 20–23, 1975*, pages 71–82. Springer, 1975.

22 Cees M. Fortuin, Pieter W. Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22(2):89–103, 1971.

23 Ehud Friedgut. Sharp thresholds of graph properties, and the k-SAT problem. *J. Amer. Math. Soc*, 12:1017–1054, 1998.

24 Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17:13–27, 1984.

25 Mika Göös and Toniann Pitassi. Communication lower bounds via critical block sensitivity. *arXiv preprint arXiv:1311.2355*, 2013.

26 Michelangelo Grigni and Michael Sipser. Monotone complexity. *Boolean function complexity*, 169:57–75, 1992.

27 Michelangelo Grigni and Michael Sipser. Monotone separation of logarithmic space from logarithmic depth. *Journal of Computer and System Sciences*, 50(3):433–437, 1995.

28 Johan Håstad. The shrinkage exponent of de Morgan formulas is 2. *SIAM Journal on Computing*, 27(1):48–64, 1998.

29 Johan Håstad. On the correlation of parity and small-depth circuits. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 19, page 137, 2012.

30 Richard Holley. Remarks on the FKG inequalities. *Communications in Mathematical Physics*, 36(3):227–231, 1974.

31 Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on*, pages 538–545. IEEE, 1995.

32 Kazuo Iwama and Hiroki Morizumi. An explicit lower bound of $5n - o(n)$ for boolean circuits. In *Mathematical foundations of computer science 2002*, pages 353–364. Springer, 2002.

33 Svante Janson. Poisson approximation for large deviations. *Random Structures & Algorithms*, 1(2):221–229, 1990.

34 Stasys Jukna. On the minimum number of negations leading to super-polynomial savings. *Information processing letters*, 89(2):71–74, 2004.

35 Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers*, volume 27. Springer-Verlag Berlin Heidelberg, 2012.

36 George Karakostas, Jeff Kinne, and Dieter van Melkebeek. On derandomization and average-case complexity of monotone functions. *Theoretical Computer Science*, 434:35–44, 2012.

37 Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM Journal on Discrete Mathematics*, 3(2):255–265, 1990.

38 Richard M. Karp. The probabilistic analysis of some combinatorial search algorithms. *Algorithms and complexity: New directions and recent results*, 1:19, 1976.

39 Adam Klivans and Rocco A. Servedio. Boosting and hard-core set construction. *Machine Learning*, 51(3):217–238, 2003.

40 Ilan Komargodski, Ran Raz, and Avishay Tal. Improved average-case lower bounds for demorgan formula size. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 588–597. IEEE, 2013.

41 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size[AC^0] circuits with $n^{1-o(1)}$ symmetric gates. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 640–651. Springer, 2011.

42 A.A. Markov. On the inversion complexity of a system of functions. *Journal of the ACM (JACM)*, 5(4):331–334, 1958.

43 Ryan O'Donnell. Hardness amplification within NP. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 751–760. ACM, 2002.

44 Ryan O'Donnell and Karl Wimmer. KKL, Kruskal-Katona, and monotone nets. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 725–734. IEEE, 2009.

45 Aaron Potechin. Bounds on monotone switching networks for directed connectivity. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 553–562. IEEE, 2010.

46 Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. In *Foundations of Computer Science, 1997. Proceedings., 38th Annual Symposium on*, pages 234–243. IEEE, 1997.

47 Ran Raz and Avi Wigderson. Probabilistic communication complexity of boolean relations. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 562–567. IEEE Comput. Soc. Press, 1989.

48 Ran Raz and Avi Wigderson. Monotone circuits for matching require linear depth. *Journal of the ACM (JACM)*, 39(3):736–744, 1992.

49 Alexander Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold cicuits with AND gates at the bottom. *Information Processing Letters*, 45(6):303–307, 1993.

50 Alexander A. Razborov. Lower bounds on monotone complexity of the logical permanent. *Mathematical Notes*, 37(6):485–493, 1985.

51 Alexander A. Razborov. Lower bounds on the monotone complexity of some boolean functions. *Doklady Akademii Nauk SSSR*, 281:798–801, 1985. English translation in Soviet Math. Doklady 31 (1985), 354–357.

52 Alexander A Razborov. On the method of approximations. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 167–176. ACM, 1989.

53 Alexander A. Razborov and Steven Rudich. Natural proofs. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 204–213. ACM, 1994.

54 Benjamin Rossman. On the constant-depth complexity of k-clique. In *Proceedings of the 40th annual ACM symposium on Theory of computing*, pages 721–730. ACM, 2008.

55 Benjamin Rossman. The monotone complexity of k-clique on random graphs. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 193–201. IEEE, 2010.

56 Benjamin Rossman. Formulas vs. circuits for small distance connectivity. *arXiv preprint arXiv:1312.0355*, 2013.

57 Rocco A. Servedio. Monotone boolean formulas can approximate monotone linear threshold functions. *Discrete Applied Mathematics*, 142(1):181–187, 2004.

58 Alexander A. Sherstov. Communication complexity under product and nonproduct distributions. In *Computational Complexity, 2008. CCC'08. 23rd Annual IEEE Conference on*, pages 64–70. IEEE, 2008.

59 P.M. Spira. On time-hardware complexity tradeoffs for boolean functions. In *Proceedings of the 4th Hawaii Symposium on System Sciences*, pages 525–527, 1971.

60 Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, pages 423–439, 1965.

61 Éva Tardos. The gap between monotone and non-monotone circuit complexity is exponential. *Combinatorica*, 8(1):141–142, 1988.

62 Prasoon Tiwari and Martin Tompa. A direct version of Shamir and Snir's lower bounds on monotone circuit depth. *Information Processing Letters*, 49(5):243–248, 1994.

63 Leslie G. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5(3):363–366, 1984.

64 Leslie G. Valiant. Negation is powerless for boolean slice functions. *SIAM Journal on Computing*, 15(2):531–535, 1986.

65 Emanuele Viola. Correlation bounds for polynomials over {0,1}. *ACM SIGACT News*, 40(1):27–44, 2009.

66 Ingo Wegener. Relating monotone formula size and monotone depth of boolean functions. *Information Processing Letters*, 16(1):41–42, 1983.

67 Ingo Wegener. More on the complexity of slice functions. *Theoretical computer science*, 43:201–211, 1986.

A Proof of Lemma 3.4 (Persistent Minterms Under $\vee$ and $\wedge$)

To simplify notation, we write f_s for $f^{\vee\rho_s}$ and g_s for $g^{\vee\rho_s}$.

Proof of (3): Consider any $x \in \mathcal{M}_d^{\vec{\rho}}(f \vee g)$. Fix $0 \le s \le t \le m$ such that $t - s \ge \left\langle {d \atop |x|} \right\rangle$ and $x \in \mathcal{M}(f_s \vee g_s) \cap \mathcal{M}(f_t \vee g_t)$. Since x is a minterm of $f_s \vee g_s$, we have $f_s(x) = 1$ or $g_s(x) = 1$. Without loss of generality, assume $f_s(x) = 1$. We claim that x is also a minterm of f_t. Clearly $f_t(x) = 1$ since $f_s \le f_t$. It suffices to show that $f_t(y) = 0$ for all $y < x$. This follows from the fact that x is a minterm of $f_t \vee g_t$, hence $(f_t \vee g_t)(y) = 0$ for all $y < x$. Therefore, $x \in \mathcal{M}(f_s) \cap \mathcal{M}(f_t)$. Since $t - s \ge \left\langle {d \atop j} \right\rangle \ge \left\langle {d-1 \atop j} \right\rangle$, we conclude that $x \in \mathcal{M}_{d-1}^{\vec{\rho}}(f)$.

Proof of (4): Consider any $x \in \mathcal{M}_d^{\vec{\rho}}(f \wedge g)$. Fix $0 \le s \le t \le m$ such that $t - s \ge \left\langle {d \atop |x|} \right\rangle$ and $x \in \mathcal{M}(f_s \wedge g_s) \cap \mathcal{M}(f_t \wedge g_t)$. Let $\ell := t - s$. We will construct, by induction on $i = 0, 1, \ldots, \ell$, two sequences $y_0 \ge y_1 \ge \cdots \ge y_\ell$ and $z_0 \ge z_1 \ge \cdots \ge z_\ell$ such that $y_i \in \mathcal{M}(f_{s+i})$ and $z_i \in \mathcal{M}(g_{s+i})$ and $y_i \vee z_i = x$:

- For the base case $i = 0$, since x is a minterm of $f_s \wedge g_s$, we have $f_s(x) = g_s(x) = 1$. Therefore, there exist $y \in \mathcal{M}(f_s)$ and $z \in \mathcal{M}(g_s)$ such that $y, z \le x$. Note that $(f_s \wedge g_s)(y \vee z) = 1$ and $y \vee z \le x$. Again using the fact that x is a minterm of $f_s \wedge g_s$, it follows that $y \vee z = x$. These are the starting terms of our sequence: $y_0 = y$ and $z_0 = z$.
- For the induction step, suppose we have chosen $y_{i-1} \in \mathcal{M}(f_{s+i-1})$ and $z_{i-1} \in \mathcal{M}(g_{s+i-1})$ such that $y_{i-1} \vee z_{i-1} = x$. Since $f_{s+i-1} \le f_{s+i}$ and $g_{s+i-1} \le g_{s+i}$, we have $f_{s+i}(y_{i-1}) = g_{s+i}(z_{i-1}) = 1$. Therefore, there exist $y \in \mathcal{M}(f_{s+i})$ and $z \in \mathcal{M}(g_{s+i})$ such that $y \le y_{i-1}$ and $z \le z_{i-1}$. Note that $(f_{s+i} \wedge g_{s+i})(y \vee z) = 1$ and $y \vee z \le x$. Since x is a minterm of $f_{s+i} \wedge g_{s+i}$, it follows that $y \vee z = x$. These are the next terms in our sequence: $y_i = y$ and $z_i = z$.

Having constructed sequences $\vec{y}, \vec{z} \in \mathrm{Seq}_{\ge}^{\ell}(\{0,1\}^n)$, we finish the proof using Lemma 3.2. Since $\ell \ge \left\langle {d \atop |x|} \right\rangle \ge \left\langle {d \atop |y_0|} \right\rangle$, we may apply Lemma 3.2 to the reversed sequence $(y_\ell, y_{\ell-1}, \ldots, y_0) \in \mathrm{Seq}_{\le}^{\ell}(\{0,1\}^n)$; we get $0 \le a \le b \le \ell$ such that $y_a = y_b$ and $b - a \ge \left\langle {d-1 \atop |y_a|} \right\rangle$. Therefore, $y_a \in \mathcal{M}_{d-1}^{\vec{\rho}}(f)$. Similarly, we get $z_c \in \mathcal{M}_{d-1}^{\vec{\rho}}(g)$ for some $0 \le c \le \ell$. Since $y_0 \le y_a \le y_\ell$ and $z_0 \le z_c \le z_\ell$ and $z_0 \vee y_0 = y_\ell \vee z_\ell = x$, we conclude that $y_a \vee z_c = x$. ◀

B Proof of Lemma 5.12 (Pathset Complexity and Formula Size)

Assume Φ is a monotone formula and $\vec{\rho} \in \mathrm{Seq}_{\le}^{m}(\mathcal{G})$ such that $\mathcal{P}_A^{\vec{\rho}}(\Psi)$ is ε-small for every subformula Ψ of Φ and every $A \subseteq K$.

Consider any $\phi \in \mathrm{Leaves}(\Phi)$ labeled by the indicator variable for a potential edge $v^{(i)}w^{(j)}$. Clearly $\mathcal{P}_A^{\vec{\rho}}(\phi) = \emptyset$ for all $A \subseteq K$ except possibly when $E_A = \{vw\}$, in which case the only possibility for $\mathcal{P}_A^{\vec{\rho}}(\phi)$ other than $\emptyset$ is the singleton pathset $\{A'\}$ where A' is the A-section with $E_{A'} = \{v^{(i)}w^{(j)}\}$. It follows that $\sum_{A \subseteq K} |\mathcal{P}_A^{\vec{\rho}}(\phi)| \le 1$.

Next, consider $\Psi \in \mathrm{Sub}(\Phi)$ with an $\vee$-gate on top: $\Psi = \Psi_1 \vee \Psi_2$. For all $A \subseteq K$, by Lemma 5.11, we have $\mathcal{P}_A^{\vec{\rho}}(\Psi) \subseteq \mathcal{P}_A^{\vec{\rho}}(\Psi_1) \cup \mathcal{P}_A^{\vec{\rho}}(\Psi_2)$. By properties (monotonicity) and (sub-additivity) of χ_ε, it follows that

$$\chi_\varepsilon(\mathcal{P}_A^{\vec{\rho}}(\Psi)) \le \chi_\varepsilon(\mathcal{P}_A^{\vec{\rho}}(\Psi_1)) + \chi_\varepsilon(\mathcal{P}_A^{\vec{\rho}}(\Psi_2)). \tag{23}$$

Now consider $\Psi = \Psi_1 \wedge \Psi_2 \in \mathrm{Sub}(\Phi)$. By Lemma 5.11, we have

$$\mathcal{P}_A^{\vec{\rho}}(\Psi) \subseteq \mathcal{P}_A^{\vec{\rho}}(\Psi_1) \cup \mathcal{P}_A^{\vec{\rho}}(\Psi_2) \cup \bigcup_{B,C \subsetneq A \,:\, B \cup C = A} \mathcal{P}_B^{\vec{\rho}}(\Psi_1) \bowtie \mathcal{P}_C^{\vec{\rho}}(\Psi_2). \tag{24}$$

(This expression extracts from (16) the case where $B = A$, noting that $\mathcal{P}^{\vec{\rho}}_A(\Psi_1) \bowtie \mathcal{P}^{\vec{\rho}}_C(\Psi_2) \subseteq \mathcal{P}^{\vec{\rho}}_A(\Psi_1)$; and similarly the case where $C = A$.) By properties (monotonicity), (sub-additivity) and (join inequality) of χ_ε,

$$\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi)) \leq \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_1)) + \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_2)) + \sum_{B,C \subsetneq A \,:\, B \cup C = A} \big(\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_B(\Psi_1)) + \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_C(\Psi_2))\big)$$

$$\leq \Big(\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_1)) + 2^k \sum_{B \subsetneq A} \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_B(\Psi_1))\Big) + \Big(\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_2)) + 2^k \sum_{B \subsetneq A} \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_B(\Psi_2))\Big). \quad (25)$$

If we now start with $\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_K(\Phi))$ and repeatedly expand according to (25) and (23) down to the leaves of Φ, we get a bound of the form

$$\mathcal{P}^{\vec{\rho}}_K(\Phi) \leq \sum_{\phi \in \mathrm{Leaves}(\Phi)} \sum_{A \subseteq K} c_{\phi,A} \cdot \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\phi))$$

for some $c_{\phi,A} \in \mathbb{N}$. For $\phi \in \mathrm{Leaves}(\Phi)$ at depth d ($\leq \mathrm{depth}(\Phi)$), the coefficient $c_{\phi,A}$ equals the sum, over all chains $K = B_0 \supset B_1 \supset \cdots \supset B_t = A$, of 2^{kt} times the binomial coefficient $\binom{d}{t}$ (counting the locations of the $\wedge$-gates above ϕ where branching occurred in the expansion of (25)). Thus, we have the upper bound $c_{\phi,A} \leq 2^{O(k^2)} \cdot \mathrm{depth}(\Phi)^k$. Using the fact that $\sum_{A \subseteq K} |\mathcal{P}^{\vec{\rho}}_A(\phi)| \leq 1$ for all $\phi \in \mathrm{Leaves}(\Phi)$ and the definition $\mathrm{size}(\Phi) = |\mathrm{Leaves}(\Phi)|$, we conclude that $\mathcal{P}^{\vec{\rho}}_K(\Phi) \leq 2^{O(k^2)} \cdot \mathrm{depth}(\Phi)^k \cdot \mathrm{size}(\Phi)$. ◀

C Proof of Lemma 1.3 (Negation-Limited Circuits)

Our proof of Lemma 1.3 combines a monotone coupling theorem of Holley [30] (which is the main ingredient in the proof of his generalization the FKG inequalities [22]) with an observation about negations in circuits due to Amano and Maruoka [7]. We require one definition:

▶ **Definition 3.1.** For a boolean (not necessarily monotone) function $h : \{0,1\}^n \to \{0,1\}$, let

$$\text{mon-pairs}(h) := \big\{(x,y) \in \{0,1\}^n \times \{0,1\}^n : h(x) = 0 \text{ and } h(y) = 1 \text{ and } x < y\big\}.$$

The following lemma and its proof are adapted from Theorem 3.2 of [7]. The only difference is that we consider all monotone pairs, rather than only the monotone boundary (i.e. only monotone pairs (x,y) with $|y| - |x| = 1$).

▶ **Lemma 3.2.** *For every circuit $\mathfrak{C}$ with t negation gates, there exist $t' = 2^{t+1} - 1$ monotone circuits $\mathfrak{M}_1, \ldots, \mathfrak{M}_{t'}$ of the same size and depth such that* $\text{mon-pairs}(\mathfrak{C}) \subseteq \bigcup_{i=1}^{t'} \text{mon-pairs}(\mathfrak{M}_i)$.

Proof. Let $\mathfrak{C}_1, \ldots, \mathfrak{C}_t$ be the sub-circuits of $\mathfrak{C}$ which feed directly into negation gates, listed in "topological order" such that $i < j$ whenever $\mathfrak{C}_i$ is a sub-circuit of $\mathfrak{C}_j$. Also, let $\mathfrak{C}_{t+1}$ be $\mathfrak{C}$ itself. For every $j \in \{1, \ldots, t+1\}$ and $\alpha \in \{0,1\}^{j-1}$, let $\mathfrak{M}_\alpha$ be the monotone circuit obtained from $\mathfrak{C}_j$ by, for each $i \in \{1, \ldots, j-1\}$ such that $\mathfrak{C}_i$ is a sub-circuit of $\mathfrak{C}_j$, replacing the negation gate above $\mathfrak{C}_i$ with the constant α_i. The number of these monotone circuits is $\sum_{j=1}^{t+1} 2^{j-1} = 2^{t+1} - 1$. To finish the argument, consider any $(x,y) \in \text{mon-pairs}(\mathfrak{C})$. Let j be the first index such that $\mathfrak{C}_j(x) \neq \mathfrak{C}_j(y)$, and let $\alpha \in \{0,1\}^{j-1}$ be the element $\alpha_i := \mathfrak{C}_i(x) = \mathfrak{C}_i(y)$. Then $(x,y) \in \text{mon-pairs}(\mathfrak{M}_\alpha)$. We conclude that $\text{mon-pairs}(\mathfrak{C}) \subseteq \bigcup_{j \in [t+1]} \bigcup_{\alpha \in \{0,1\}^{j-1}} \text{mon-pairs}(\mathfrak{M}_\alpha)$. ◀

▶ **Lemma 3.3** (Holley [30]). *Let μ_0, μ_1 be two strictly positive probability distributions on $\{0,1\}^n$ which satisfy the "Holley condition"*

$$\mu_0(x)\mu_1(y) \le \mu_0(x \wedge y)\mu_1(x \vee y) \qquad \text{for all } x, y. \tag{26}$$

Then there exists a probability distribution ν on $\{0,1\}^n \times \{0,1\}^n$ ("monotone coupling of μ_0 and μ_1") such that

- $\sum_y \nu(x,y) = \mu_0(x)$ *for all x,*
- $\sum_x \nu(x,y) = \mu_1(y)$ *for all y,*
- $\nu(x,y) > 0 \Rightarrow x \le y$ *for all x, y.*

Holley's proof of Lemma 3.3 uses a Markov chain coupling argument. We remark that Lemma 3.3 also follows from an earlier (and much more general) monotone coupling theorem of Strassen [60].

▶ **Lemma 3.4.** *Let μ be a distribution on $\{0,1\}^n$ which satisfies the FKG lattice condition* (1)*, and let $f : \{0,1\}^n \to \{0,1\}$ be a monotone function such that $\mathbb{E}_\mu(f) \in (0,1)$. For $b \in \{0,1\}$, define the distribution μ_b on $\{0,1\}^n$ by*

$$\mu_b(x) := \begin{cases} \mu(x)/(1 - \mathbb{E}_\mu(f)) & \text{if } f(x) = b = 0, \\ \mu(x)/\mathbb{E}_\mu(f) & \text{if } f(x) = b = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

Then the pair μ_0, μ_1 satisfy the Holley condition (26)*.*

Proof. We simply observe:

- If $f(x) = 1$, then $\mu_0(x) = \mu_0(x \wedge y) = 0$.
- If $f(y) = 0$, then $\mu_1(y) = \mu_1(x \vee y) = 0$.
- If $f(x) = 0$ and $f(y) = 1$, then

$$\mu_0(x)\mu_1(y) = \frac{\mu(x)\mu(y)}{\mathbb{E}_\mu(f)(1 - \mathbb{E}_\mu(f))} \le \frac{\mu(x \wedge y)\mu(x \vee y)}{\mathbb{E}_\mu(f)(1 - \mathbb{E}_\mu(f))} = \mu_0(x \wedge y)\mu_1(x \vee y). \qquad \blacktriangleleft$$

Proof of Lemma 1.3. Let μ be a distribution on $\{0,1\}^n$ which satisfies the FKG lattice condition (1), and suppose $f \in \mathbb{B}_n^+$ such that $\mathbb{E}_\mu(f) = 1/2$ (i.e. f is balanced with respect to μ). We prove the contrapositive statement to Lemma 1.3. Assume $\mathfrak{C}$ is a monotone circuit which computes f on μ with advantage δ, that is,

$$\mathop{\mathbb{P}}_{x \sim \mu}\big[\, \mathfrak{C}(x) = f(x) \,\big] = \frac{1}{2} + \delta.$$

We will show that f is computed with advantage $\ge \delta/(2^{t+1} - 1)$ by a monotone circuit of the same size and depth.

Define μ_0, μ_1 by (27) as in Lemma 3.4. By Lemma 3.3 there is a monotone coupling ν of μ_0, μ_1, which is supported on mon-pairs(f). For every monotone function $h \in \mathbb{B}_n^+$, we have

$$\begin{aligned} \nu\big(\text{mon-pairs}(h)\big) &= \mathop{\mathbb{E}}_{(x,y)\sim\nu}\big[\, h(y) - h(x) \,\big] \\ &= \mathop{\mathbb{E}}_{(x,y)\sim\nu}\big[\, h(y) \,\big] - \mathop{\mathbb{E}}_{(x,y)\sim\nu}\big[\, h(x) \,\big] \\ &= \mathop{\mathbb{P}}_{(x,y)\sim\nu}\big[\, h(y) = 1 \,\big] + \mathop{\mathbb{P}}_{(x,y)\sim\nu}\big[\, h(x) = 0 \,\big] - 1 \\ &= 2\Big(\mathop{\mathbb{P}}_{y\sim\mu}\big[\, h(y) = f(y) = 1 \,\big] + \mathop{\mathbb{P}}_{x\sim\mu}\big[\, h(x) = f(x) = 0 \,\big] \Big) - 1 \\ &= 2 \mathop{\mathbb{P}}_{x\sim\mu}\big[\, h(x) = f(x) \,\big] - 1. \end{aligned} \tag{28}$$

It follows from Lemma 3.2 that there exists a monotone circuit $\mathfrak{M}$, of the same size and depth as $\mathfrak{C}$, such that

$$\nu\big(\text{mon-pairs}(\mathfrak{M})\big) \geq \frac{1}{2^{t+1}-1}\nu\big(\text{mon-pairs}(\mathfrak{C})\big).$$

We complete the proof by two applications of (28):

$$\begin{aligned}
\mathop{\mathbb{P}}_{x\sim\mu}\big[\,\mathfrak{M}(x) = f(x)\,\big] &= \frac{1}{2}\Big(1 + \nu\big(\text{mon-pairs}(\mathfrak{M})\big)\Big)\\
&\geq \frac{1}{2}\Big(1 + \frac{1}{2^{t+1}-1}\nu\big(\text{mon-pairs}(\mathfrak{C})\big)\Big)\\
&= \frac{1}{2}\Big(1 + \frac{1}{2^{t+1}-1}\Big(2\mathop{\mathbb{P}}_{x\sim\mu}\big[\,\mathfrak{C}(x) = f(x)\,\big] - 1\Big)\Big)\\
&= \frac{1}{2} + \frac{\delta}{2^{t+1}-1}.
\end{aligned}$$

◀