Correlation Bounds Against Monotone NC^1

Benjamin Rossman

November 15, 2014

Abstract

This paper gives the first correlation bounds under product distributions (including the uniform distribution) against the class mNC^1 of $\mathrm{poly}(n)$-size $O(\log n)$-depth monotone circuits. Our main theorem, proved using the *pathset complexity* framework recently 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. As a corollary, via 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 $\varepsilon > 0$. (This bound is nearly tight, since every monotone function has correlation $\Omega(\frac{\log n}{\sqrt{n}})$ with a function in mNC^1 [44].)

Unlike previous lower bounds for monotone circuits (i.e. under non-product distributions), these correlation bounds extend smoothly to negation-limited circuits. By a simple argument using Holley's monotone coupling theorem [30], we show the following lemma: under 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, up to the same size and depth, for (non-monotone) boolean circuits with t negation gates. Our correlation bounds against mNC^1 thus extend to NC^1 circuits with $(\frac{1}{2} - \varepsilon)\log n$ negations. (This improves a previous $\frac{1}{6}\log\log n$ lower bound [7] on the negation-limited complexity of an explicit monotone function; by [21], NC^1 circuits with $\lceil \log(n+1) \rceil$ negations are equivalent to full NC^1.)

Contents

1 Introduction

The majority of research in Boolean Circuit Complexity has focused on restricted classes of circuits. Super-polynomial lower bounds have so far been achieved under two basic restrictions: *bounded depth* (essentially AC^0 [1, 24]) and the *monotone* setting (circuits without negation gates [51]). For another natural class, *deMorgan formulas* (circuits with fan-out 1), nearly cubic $n^{3-o(1)}$ lower bounds are known [28].

For bounded-depth circuits and deMorgan formulas, the state-of-the-art *worst-case* lower bounds (from 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],
- an explicit function (in P) is $\frac{1}{2} + 2^{-\Omega(r)}$ hard for deMorgan formulas of size $n^{3-o(1)}/r^2$ [40].

(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, the knowledge of worst-case lower bounds is much better: a long line of works [4, 8, 27, 37, 45, 48, 46, 50, 51] (among many others) have achieved nearly all separations between the monotone versions of 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), nothing has been known. It not known, for instance, whether any monotone function in NP is $1 - \frac{1}{\mathrm{poly}(n)}$ hard for polynomial-size monotone circuits.

This is a significant gap in our basic understanding of monotone computation. Product distributions are an important and natural setting for the average-case complexity of monotone functions. Both k-SAT and k-CLIQUE are believed to be hard-on-average under appropriate product distributions. Analysis of the threshold behavior, as well as the average-case performance of specific algorithms for these problems is an extremely active topic of research. Product distribution play a prominent role generally when it comes to monotone functions on the hypercube (see the FKG inequality [22], the Bollobás-Thomason theorem [18], Friedgut's threshold theorem [23], etc.) The special case of the uniform distribution is especially important for applications such as cryptography. For these reasons, average-case lower bounds under product distributions, even against monotone circuits, would be extremely interesting.

Despite the absence of results under product distributions, there is a history of correlation bounds against monotone circuits under *non-product* distributions. Consider the very first super-polynomial lower bound of Razborov [51] for the k-CLIQUE function. Although often stated as a worst-case lower bound, we can view this result (and the subsequent quantitative improvement of Alon and Boppana [4]) as a correlation bound under a particular distribution. This distribution, μ, is the following random graph on n vertices: half of the time, μ is a uniform random k-clique, and the other half of the time, μ is a uniform random $(k-1)$-coclique (i.e. complete $(k-1)$-partite graph). Stated as a correlation bound, the results of [4, 51] show:

- for all $3 \leq k \leq n^{1/4}$, the k-CLIQUE function on n-vertex graphs is $\frac{1}{2} + n^{-\Omega(\sqrt{k})}$ hard under μ for the class mP (poly(n)-size monotone circuits).

For a smaller range $k \leq \log n$, this hardness improves to (best possible) $\frac{1}{2} + n^{-\Omega(k)}$. (Correlation bounds under similar (non-product) distributions were recently [20, 25] obtained for monotone classes within mP, strengthening previous worst-case separations among these classes.)

In the context of monotone circuit 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). μ thus exploits monotonicity in the strongest possible way. On the other hand, 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 $\binom{k-1}{2}\left(\frac{n}{k-1}\right)^2$ ($\geq n^2/2$) of any 0-instance. This means that k-CLIQUE is equivalent under μ to the (anti-monotone) threshold function $\mathrm{THR}_{<n^2/2}$. Therefore, even though k-CLIQUE is hard under μ for monotone circuits, it is easy under μ for non-monotone circuits with a single negation gate. This discomfort is 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 values of $\ell \in \{k, \ldots, n\}$, 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 setting for the average-case analysis of k-CLIQUE is given by the Erdős-Rényi random graph $\mathrm{G}(n,p)$ for the unique $p = p(k,n)$ such that $\mathbb{P}[\,\mathrm{G}(n,p)$ contains a k-clique$\,] = \frac{1}{2}$. (That is, $\mathrm{G}(n,p)$ is the p-biased product distribution where $p = \Theta(n^{-2/(k-1)})$ for small $k \leq \log n$.) Karp [38] conjectured (in at least the special case $p = \frac{1}{2}$) that k-CLIQUE is hard-on-average under $\mathrm{G}(n,p)$. Work of the author [54] gave the first correlation bound for this problem in the restricted setting of AC^0 (polynomial-size constant-depth boolean circuits):

- for all $k \leq \log^{1/2} n$, k-CLIQUE is $\frac{1}{2} + n^{-\Omega(k)}$ hard under $\mathrm{G}(n,p)$ for AC^0.

Combining the technique of [54] with the "approximation method" framework of Razborov [51], follow-up work of the author [55] gave a correlation bound against monotone circuits under the following distribution ν: half of the time, ν is $\mathrm{G}(n,p)$ plus a uniform random planted k-clique; the other half of the time, ν is $\mathrm{G}(n,2p)$ conditioned on not having any k-clique. The result of [55] is

- for all $k \leq \log^{1/2} n$, k-CLIQUE is $\frac{1}{2} + n^{-\Omega(k)}$ hard under ν for mP.

Of course, we would really like to show that k-CLIQUE is $\frac{1}{2} + o(1)$ under $\mathrm{G}(n,p)$ for mP. (Under any product distribution, we cannot hope for hardness against mP better than $\frac{1}{2} + n^{-1/2}$, as we discuss momentarily.) While the result of [55] feels like progress (at least the distribution ν is realistically hard for non-monotone circuits), ν unfortunately suffers from the same shortcoming as μ: the 0-instances and 1-instances are separable by an anti-monotone threshold function (in this case $\mathrm{THR}_{<\frac{3}{2}\binom{n}{2}p}$).

In the present paper, we finally prove a correlation bound under $\mathrm{G}(n,p)$ in the monotone setting; however, not for k-CLIQUE and not for monotone circuits, but rather for k-CYCLE and for monotone formulas.

1.1 Our Results

The main theorem of this paper is a correlation bound for the average-case k-CYCLE problem against the class mNC^1 of $\mathrm{poly}(n)$-size $O(\log n)$-depth monotone circuits (equivalently: $\mathrm{poly}(n)$-size monotone formulas). We find it convenient to restrict attention to "C_k-partite" input graphs with kn vertices and kn^2 potential edges (Def. 4.2); however, our results hold in the setting of

$G(n,p)$. For the average-case analysis of k-CYCLE, we consider the (C_k-partite Erdős-Rényi) random graph, Γ, which includes each potential edge independently with probability p, where p is the unique threshold value such that $\mathbb{P}[\,\Gamma \text{ contains a } k\text{-cycle}\,] = \frac{1}{2}$. (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).

This lower bound is essentially tight: k-CYCLE is (worst-case) computable by monotone formulas of size $n^{O(\log k)}$, as well as by $\mathrm{poly}(n)$-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 strong average-case separation of mNC^1 from higher complexity classes.

Theorem 1.1 also implies (essentially optimal) correlation bounds against mNC^1 under the *uniform distribution*. Note that 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; the random graph Γ, although a product distribution, is not the uniform distribution. Nevertheless, using O'Donnell's hardness amplification theorem [43] (and a primitive device to generate Γ from uniform random bits), we get 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 direct product[1] TRIBES $\otimes$ $\log\log n$-CYCLE $\otimes$ p-BIAS (on $N = \mathrm{poly}(n)$ variables) where

- p-BIAS : $\{0,1\}^n \to \{0,1\}$ is a monotone function with $|p\text{-BIAS}^{-1}(1)| = \lceil p2^n \rceil$ (i.e. p-BIAS generates a p'-biased bit where $p \leq p' < p + 2^{-n}$ and $p \sim 1/n$ is the $\frac{1}{2}$-threshold for $\log\log n$-CYCLE),
- 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.

Since both p-BIAS (suitably defined)[2] and TRIBES are in mAC^0, this direct product remains in mSAC^1. See O'Donnell's paper [43] for details on the hardness amplification theorem which produces 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 $\mathrm{MAJ} \in \mathrm{mNC}^1$ [3, 63] (which is essential in the application of Implagliazzo's "hard-core set" theorem [31, 39]).

The correlation bound of Corollary 1.2 is nearly best possible under the uniform distribution. O'Donnell and Wimmer [44] showed that every monotone function $\{0,1\}^n \to \{0,1\}$ has agreement

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

[2] For every $m \in \{0, \ldots, 2^n\}$, there is an n-term monotone DNF with exactly m satisfying assignments.

$\frac{1}{2} + \Omega(\frac{\log n}{\sqrt{n}})$ with one of functions $0, 1, x_1, \ldots, x_n, \mathrm{MAJ}$. 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 negation-limited circuits, 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. *Suppose μ is a distribution which satisfies the FKG lattice condition (1) and f is 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 boolean 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 $\mathrm{MAJ} \in \mathrm{NC}^1$). This improves the 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 [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.)

1.2 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 function $f^{\vee\rho_0} \leq f^{\vee\rho_1} \leq \cdots \leq f^{\vee\rho_m}$ 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 a *d-persistent minterm* of f under $\vec{\rho}$ if it is a common minterm of $\binom{d+k-1}{k-1}$ of the functions $f^{\vee\rho_i}$.

Persistent minterms behave like ordinary minterms with respect to operations $\vee$ and $\wedge$ (Lemma 3.7). However, unlike ordinary minterms, persistent minterms are "noise-insensitive" in a certain sense. Suppose $\xi^{(1)}, \ldots, \xi^{(m)}$ are independent samples from a distribution of "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 in the sense of having survived ≥ 1 hit of monotone noise. This is advantageous for the following reason: whereas an arbitrary monotone function might (in the worst case) have $\binom{n}{k}$ ordinary minterms of size k, by choosing an appropriate of distribution of noise, we can ensure that every monotone function (with very high probability) has few persistent minterms of a given size.

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 "inside the variance" of the random graph Γ, in the sense that Γ and $\Gamma \cup \Xi_\ell$ have small total variation distance. Because of this, we are able to show the following (roughly speaking): 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 Ξ_ℓ-noise-invariant minterms of f (Lemma 4.5).

Our proof of Theorem 1.1 can be interpreted as showing that Ξ_ℓ is "hard-core noise" for the average-case k-CYCLE problem. In some sense, all of the action in our proof takes place within the noise Ξ_ℓ. Since ℓ is very small (we get non-trivial correlation bounds when ℓ is just $n^{O(1/\log k)}$), we are able to exploit monotonicity in a razor-thin way (i.e. less than the "variance" of the random graph Γ). This appears to be a very special property of the average-case k-CYCLE problem.[3]

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 pathsets with respect to the operations $\cup$ and $\bowtie$. Crucially, pathsets are subject to a collection of density constraints called *smallness*; this is responsible for the high cost of constructing pathsets beyond a certain density.

The pathset complexity framework was introduced in [56] with the purpose of separating formula-size and circuit-size within AC^0. The technique is highly specialized 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). (A reader's guide to [56] is given in Appendix A.)

Our correlation bound against mNC^1 (Theorem 1.1) is proved by reduction to this pathset complexity lower bound. Given a monotone formula with sufficiently large correlation with k-CYCLE, we define (random) pathsets at all gates using persistent minterms. We show (Lemma 5.13) that all of these pathsets satisfy the smallness condition (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.

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.

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 $x \in \{0,1\}^n$.

[3] In particular, consider the average-case k-CLIQUE problem (for small $k \leq \log n$) on the random graph $\mathrm{G}(n,p)$ at the critical threshold $p = \Theta(n^{-2/(k-1)})$. It is not clear whether k-CLIQUE admits a similar distribution of "hard-core noise" which lives inside the "variance" of $\mathrm{G}(n,p)$. Without such a distribution, the technique of this paper does not apply. It remains an open question whether average-case k-CLIQUE is $1 - 1/\mathrm{poly}(n)$ hard for mNC^1.

For $f \in \mathbb{B}_n^+$ and $x \in \{0,1\}^n$, we say that x 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 $f, g \in \mathbb{B}_n^+$,*

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

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 *(formula) 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 $\vec{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_m)$ 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 3.3. *For all $d, k \geq 1$ and $\vec{a} \in \mathrm{Seq}_{\leq}^k(\mathbb{R})$, 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 $(a_0, \ldots, a_{k-1}) \in \mathrm{Seq}_{\leq}^{k-1}(\mathbb{R})$. □

By the same basic induction, we have:

Lemma 3.4. *For all $d, m \geq 1$ and $\vec{x} \in \mathrm{Seq}_{\leq}^{m}(\{0,1\}^n)$, if $m \geq \left\langle {}^{d}_{|x_m|} \right\rangle$, then $x_s = x_t$ for some $0 \leq s \leq t \leq m$ with $t - s \geq \left\langle {}^{d-1}_{|x_s|} \right\rangle$.*

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

Definition 3.5 (Persistent Minterms)**.**
For $\vec{f} \in \mathrm{Seq}_{\leq}^{m}(\mathbb{B}_n^+)$ 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 $t - s \geq \left\langle {}^{d}_{|x|} \right\rangle$.
The set of d-persistent minterms of $\vec{f}$ is denoted by $\mathcal{M}_d(\vec{f})$.

We have defined persistent minterms in 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 3.6. For $f \in \mathbb{B}_n^+$ and $\vec{\rho} \in \mathrm{Seq}_{\leq}^{m}(\{0,1\}^n)$, let $\mathcal{M}_d^{\vec{\rho}}(f) := \mathcal{M}_d(f^{\vee\rho_0} \leq f^{\vee\rho_1} \leq \cdots \leq f^{\vee\rho_m})$.

Lemma 3.7 (Persistent Minterm Lemma)**.** *For all $f, g \in \mathbb{B}_n^+$ and $\vec{\rho} \in \mathrm{Seq}_{\leq}^{m}(\{0,1\}^n)$ and $d \geq 1$,*

$$\mathcal{M}_d^{\vec{\rho}}(f \vee g) \subseteq \mathcal{M}_{d-1}^{\vec{\rho}}(f) \cup \mathcal{M}_{d-1}^{\vec{\rho}}(g), \tag{2}$$

$$\mathcal{M}_d^{\vec{\rho}}(f \wedge g) \subseteq \{x \vee y : x \in \mathcal{M}_{d-1}^{\vec{\rho}}(f),\ y \in \mathcal{M}_{d-1}^{\vec{\rho}}(g)\}. \tag{3}$$

The proof, which we include in Appendix B, is straightforward (in particular, we show (3) using Lemma 3.4). 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 $\mathscr{G} \cong \{0,1\}^{k^2 n}$ of graphs (with kn^2 possible edges). Before defining $\mathscr{G}$, 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 $\mathscr{G}$ $(= \mathscr{G}(k,n))$ the set of *K-partite graphs* G satisfying

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

Here $v^{(i)}$ and $v^{(i)}w^{(j)}$ are just a friendly notation for ordered pairs (v,i) and $((v,i),(w,j))$.

Equivalently, $\mathscr{G}$ is the set of subgraphs of the product graph $K \times N$ where $N = ([n],[n]^2)$ is the complete directed graph (with a loop at every vertex).

In the context of functions $\mathscr{G} \to \{0,1\}$, we identify $\mathscr{G}$ with the hypercube $\{0,1\}^{kn^2}$.

Definition 4.3 (k-CYCLE)**.**

For $G \in \mathscr{G}$, 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 $\mathscr{G} \to \{0,1\}$ 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, $\circlearrowleft$ and Ξ_ℓ)**.**

Γ, $\circlearrowleft$ and Ξ_ℓ denote the following (independent) random graphs in $\mathscr{G}$:

- Let Γ be the (K-partite, Erdős-Rényi) random graph in $\mathscr{G}$ 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 critical threshold such that $\mathbb{P}[\,k\text{-CYCLE}(\Gamma) = 1\,] = \frac{1}{2}$.
- Let $\circlearrowleft$ be the uniform random k-cycle in $\mathscr{G}$. For $e \in E_K$, we write $\circlearrowleft^{-e}$ for the graph obtained from $\circlearrowleft$ by deleting the edge in $\circlearrowleft$ corresponding to e. Note that $\circlearrowleft^{-e}$ is a path of length $k-1$.
- For $\ell \in \mathbb{N}$, let Ξ_ℓ be the random graph $\circlearrowleft_1^{-e_1} \cup \cdots \cup \circlearrowleft_\ell^{-e_\ell}$ where $\circlearrowleft_1, \ldots, \circlearrowleft_\ell$ 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 $\circlearrowleft_1^{-e_1} \cup \cdots \cup \circlearrowleft_\ell^{-e_\ell}$ 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 Appendix C.

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

$$\mathop{\mathbb{P}}_{\Gamma}\left[\, f(\Gamma) = k\text{-CYCLE}(\Gamma)\,\right] \geq \frac{1}{2} + \frac{C(\ell+1)k^2}{\sqrt{n}} \tag{4}$$

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

$$\mathop{\mathbb{P}}_{\Xi_\ell}\left[\, \mathop{\mathbb{P}}_{\circlearrowleft}\left[\, \circlearrowleft \in \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G \cup \Xi_\ell})\,\right] \geq n^{-1/2}\right] \geq n^{-1/2}. \tag{5}$$

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 $\Xi^{(1)}_{\ell/m}, \dots, \Xi^{(m)}_{\ell/m}$. Writing ρ_s for the partial union $\Xi^{(1)}_{\ell/m} \cup \dots \cup \Xi^{(s)}_{\ell/m}$, we have a non-decreasing sequence $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\mathscr{G})$. 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]; a guide to the relevant sections in [56] is provided in Appendix A.

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 $|V_A| - |E_A|$ (i.e. the Euler characteristic of A). Note that every pattern graph, other than K itself, is a disjoint union of paths. Therefore,

$$A \neq K \;\Rightarrow\; |V_A| - |E_A| = |\{\text{connected components of } A\}|. \tag{6}$$

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 $A' \in \mathscr{G}$ 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 $\mathscr{G}_A$. As a matter of notation, we consistently write A-sections using primes (A', A'', etc.)

Every $A' \in \mathscr{G}_A$ is isomorphic to A via the projection $v^{(i)} \mapsto v$ (in this sense, A' is a "section" of the "product bundle" $A \times N \to A$ where $N = ([n], [n]^2)$ is the complete directed graph).

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 $\mathscr{G}$ (Def. 4.2).)

Definition 5.3 (Pathsets)**.**

For $A \subseteq K$, subsets of $\mathscr{G}_A$ (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

$$\mathsf{density}(\mathcal{A}) := |\mathcal{A}| \,/\, n^{|V_A|} = \mathop{\mathbb{P}}_{A' \in \mathscr{G}_A}[\, A' \in \mathcal{A}\,]. \tag{7}$$

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

$$\mathcal{A} \bowtie \mathcal{B} := \big\{C' \in \mathscr{G}_{A\cup B} : C' = A' \cup B' \text{ for some } A' \in \mathcal{A} \text{ and } B' \in \mathcal{B}\big\}. \tag{8}$$

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 $\overline{B}' \in \mathscr{G}_{A\setminus B}$ such that $\mathcal{B} = \{B' \in \mathscr{G}_B : B' \cup \overline{B}' \in \mathcal{A}\}$.

$\mathcal{B}$ 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

$$\mathsf{density}(\mathcal{B}) \leq \varepsilon^{|V_B|-|E_B|} \text{ for all } \mathcal{B} \preceq \mathcal{A}. \tag{9}$$

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

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 $\mathcal{A} \in \mathscr{P}_\varepsilon$, then $\mathcal{A}_0 \in \mathscr{P}_\varepsilon$ and $\mathcal{B} \in \mathscr{P}_\varepsilon$ 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 $\mathcal{A}, \mathcal{B} \in \mathscr{P}_\varepsilon$, then $\mathcal{A} \bowtie \mathcal{B} \in \mathscr{P}_\varepsilon$.

Definition 5.7 (Pathset Complexity)**.**

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

- (base case) If $|E_A| \leq 1$, then $\chi_\varepsilon(\mathcal{A}) := |E_A| \cdot |\mathcal{A}|$.

  That is, $\chi_\varepsilon(\emptyset) = \chi_\varepsilon(\{\varnothing\}) = 0$ and $\chi_\varepsilon(\mathcal{A}) = |\mathcal{A}|$ if A is a single edge.

- (induction case) If $|E_A| \geq 2$, then $\chi_\varepsilon(\mathcal{A}) := \min_{(\mathcal{B}_i, \mathcal{C}_i)_i} \sum_i \chi_\varepsilon(\mathcal{B}_i) + \chi_\varepsilon(\mathcal{C}_i)$

  where $(\mathcal{B}_i, \mathcal{C}_i)_i$ ranges over all sequences of ε-small pathsets $\mathcal{B}_i, \mathcal{C}_i \in \mathscr{P}_\varepsilon$ such that $B_i, C_i \subsetneq A$ 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 $\mathcal{A}_1 \subseteq \mathcal{A}_2 \in \mathscr{P}_\varepsilon$,
- (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 $\mathcal{A}_1 \cup \mathcal{A}_2 \in \mathscr{P}_\varepsilon$,
- (join inequality) $\chi_\varepsilon(\mathcal{A} \bowtie \mathcal{B}) \leq \chi_\varepsilon(\mathcal{A}) + \chi_\varepsilon(\mathcal{B})$ for all $\mathcal{A}, \mathcal{B} \in \mathscr{P}_\varepsilon$.

In fact, these three inequalities provide a *dual characterization* of pathset complexity: χ_ε is the unique pointwise maximal function $\mathscr{P}_\varepsilon \to \mathbb{N}$ 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,*

$$\chi_\varepsilon(\mathcal{K}) \geq (1/2)^{O(2^k)} \cdot (1/\varepsilon)^{\frac{1}{6}\log k} \cdot \mathsf{density}(\mathcal{K}). \tag{10}$$

Theorem 5.8 corresponds to Theorem 5.8 of [56] (see Appendix A for a reader's guide). We mention that the lower bound proved in [56] applies more broadly to pathsets $\mathcal{A} \in \mathscr{P}_\varepsilon$ over any pattern graph $A \subseteq K$:

$$\chi_\varepsilon(\mathcal{A}) \geq (1/2)^{O(2^{|E_A|})} \cdot (1/\varepsilon)^{\frac{1}{6}\log(\mathsf{length}(A)) + |V_A| - |E_A|} \cdot \mathsf{density}(\mathcal{A}) \tag{11}$$

where $\mathsf{length}(A)$ equals the number of edges in the largest connected component of A. In fact, (11) 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 (10) 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 $f : \mathscr{G} \to \{0,1\}$ and $A \subseteq K$, let $\mathcal{M}_A(f) := \mathscr{G}_A \cap \mathcal{M}(f)$ be the pathset of A-sections which are minterms of f.

For a monotone formula Φ and $\vec{\rho} \in \mathrm{Seq}^m_{\leq}(\mathscr{G})$ and $A \subseteq K$, the pathset $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ (over A) is defined by

$$\mathcal{P}_A^{\vec{\rho}}(\Phi) := \mathscr{G}_A \cap \mathcal{M}^{\vec{\rho}}_{\mathrm{depth}(\Phi)}(\Phi). \tag{12}$$

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

$$\mathcal{P}_A^{\vec{\rho}}(\Phi) = \bigcup_{0 \le s \le t \le m \,:\, t-s \ge \left\langle \frac{\mathrm{depth}(\Phi)}{|E_A|} \right\rangle} \big(\mathcal{M}_A(\Phi^{\cup \rho_s}) \cap \mathcal{M}_A(\Phi^{\cup \rho_t})\big) \tag{13}$$
$$\subseteq \bigcup_{0 \le s \le m-1} \big(\mathcal{M}_A(\Phi^{\cup \rho_s}) \cap \mathcal{M}_A(\Phi^{\cup \rho_{s+1}})\big).$$

The following lemma is a straightforward consequence of (13).

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 $\overline{B}' \in \mathscr{G}_{A \setminus B}$ such that $\mathcal{B} = \{B' \in \mathscr{G}_B : B' \cup \overline{B}' \in \mathcal{P}_A^{\vec{\rho}}(\Phi)\}$. 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 (13), hence $\mathcal{B} \subseteq \bigcup_{s=0}^{m-1} \{B' \in \mathscr{G}_B : B' \cup \overline{B}' \in \mathcal{A}_s\}$. It follows that there exists $s \in \{0, \ldots, m-1\}$ such that

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

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

We next restate the Persistent Minterm Lemma 3.7 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 $\vec{\rho} \in \mathrm{Seq}_{\le}^m(\mathscr{G})$ and $A \subseteq K$,*

$$\mathcal{P}_A^{\vec{\rho}}(\Phi \vee \Psi) \subseteq \mathcal{P}_A^{\vec{\rho}}(\Phi) \cup \mathcal{P}_A^{\vec{\rho}}(\Psi), \qquad \mathcal{P}_A^{\vec{\rho}}(\Phi \wedge \Psi) \subseteq \bigcup_{B,C \subseteq A \,:\, B \cup C = A} \mathcal{P}_B^{\vec{\rho}}(\Phi) \bowtie \mathcal{P}_C^{\vec{\rho}}(\Psi). \tag{14}$$

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 $\vec{\rho} \in \mathrm{Seq}_{\le}^m(\mathscr{G})$ such that pathsets $\mathcal{P}_A^{\vec{\rho}}(\Psi)$ are ε-small for all $\Psi \in \mathrm{Sub}(\Phi)$ and $A \subseteq K$. Then*

$$\chi_\varepsilon(\mathcal{P}_K^{\vec{\rho}}(\Phi)) \le 2^{O(k^2)} \cdot \mathrm{depth}(\Phi)^k \cdot \mathrm{size}(\Phi). \tag{15}$$

Although the statement of Lemma 5.12 might appear complicated, the proof is actually quite simple. The derivation of (15) 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 E.

5.3 Smallness Lemma

In the last subsection, we defined pathsets $\mathcal{P}_A^{\vec{\rho}}(\Phi)$ (for arbitrary sequences $\vec{\rho} \in \mathrm{Seq}_{\le}^m(\mathscr{G})$) and showed a relationship between pathset complexity and formula size, under condition that all of the relevant pathsets are ε-small. We now show how this ε-smallness condition can be achieved — with high probability — using random graphs Ξ_ℓ. The main technical lemma for this step is the following:

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

$$\mathop{\mathbb{P}}_{\Xi_\ell}\left[\, \mathcal{M}_A(f) \cap \mathcal{M}_A(f^{\cup \Xi_\ell}) \text{ is } \underline{\text{not}} \ \varepsilon\text{-small} \,\right] \le (2n)^k \cdot \exp\left(-\Omega(\varepsilon \ell / k^2)\right). \tag{16}$$

The proof, which uses Janson's Inequality [33] and the sunflower-plucking technique of Razborov [51], is included in Appendix D.

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

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

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

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

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 $\text{size}(\Phi) = n^{O(\log k)}$ and $\text{depth}(\Phi) = O(\log k \cdot \log n)$. However, for purposes of this proof, it is enough for us to assume much weaker upper bounds $\text{size}(\Phi) \le \exp(n^{1/k})$ and $\text{depth}(\Phi) \le n^{1/k}$. We also assume $c = \Omega(1/\log k)$, since otherwise there is nothing to prove.

We set parameters m, ℓ, ε as follows:

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

Note that $m = O(\text{depth}(\Phi))^k = n^{o(c)}$. We have $n^{-1/2+c} = \omega((m\ell+1)k^2/\sqrt{n})$, that is, Φ satisfies the hypothesis (4) of Lemma 4.5 (for all sufficiently large n). Therefore, by Lemma 4.5, there exists $G \in \mathscr{G}$ such that

$$\mathop{\mathbb{P}}_{\Xi_{\ell m}}\left[\, \mathop{\mathbb{P}}_{\circlearrowleft}\left[\, \circlearrowleft \in \mathcal{M}(\Phi^{\cup G}) \cap \mathcal{M}(\Phi^{\cup G \cup \Xi_{m\ell}}) \,\right] \ge n^{-1/2} \,\right] = \Omega(n^{-1/2}). \tag{17}$$

Fixing any such G, we now generate *random* $\vec{\boldsymbol{\rho}} \in \text{Seq}^m_{\le}(\mathscr{G})$ as follows:

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

By our choice of $m = \left\langle \tbinom{\text{depth}(\Phi)}{k} \right\rangle$ and Def. 5.9 of $\mathcal{P}_K^{\vec{\boldsymbol{\rho}}}(\Phi)$ (see (13)), 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 $\mathscr{G}_K$, it follows (by definition (7) 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 $\boldsymbol{\rho}_m \overset{\text{d}}{=} \Xi_{m\ell}$, we see that (17) is equivalent to

$$\mathop{\mathbb{P}}_{\vec{\boldsymbol{\rho}}}\left[\, \mathsf{density}(\mathcal{P}_K^{\vec{\boldsymbol{\rho}}}(\Phi)) \ge n^{-1/2} \,\right] = \Omega(n^{-1/2}). \tag{18}$$

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:

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

As the upshot of (18) and (19), (for all sufficiently large n) there exists $\vec{\rho}\in\mathrm{Seq}^m_{\le}(\mathscr{G})$ satisfying both

- **Dense**$(\vec{\rho})$, the event that $\mathsf{density}(\mathcal{P}_K^{\vec{\rho}}(\Phi))\ge n^{-1/2}$, 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 $k\le\log\log n$):

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

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

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 chenstein 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 Guide to the Lower Bound in [56]

The proof of Theorem 5.8 in the present paper (Theorem 5.8 in [56]) is found in Sections 8-11 of [56]. The exposition in these sections is self-contained and may be read independently from the parts of the paper which deal with bounded-depth boolean formulas. The reader may also wish to consult Section 4 ("Preliminaries") and Section 5 ("Pathset Complexity") and Appendices A-C which give background, motivation, key examples and upper bounds for pathset complexity.

One inconsequential difference which arises between the present paper and [56] is that we deal with the average-case k-CYCLE problem, whereas [56] considers with average-case DISTANCE-k STCONN problem. For all intents and purposes, these are the same problem. However, this results in one difference in definitions: the role of K (the k-cycle graph) in present paper is played by P_k (the path of length k) in [56]. Thus, *patterns graph* in [56] are subgraphs of P_k, rather than K (however this makes no difference in any of the results). Beside this difference, a few minor changes in notation are described below:

1. $|V_A| - |E_A|$ and $\mathsf{density}(\mathcal{A})$ are denoted by $\Delta(A)$ and $\delta(\mathcal{A})$ in [56]. Also, in [56] pattern graphs are consistently represented using letters G, H, while A, B, C are reserved for *patterns* (a notion which is crucial to the lower bound in [56], but which we do not require here).

2. The smallness parameter $\varepsilon > 0$ in the present paper corresponds to $1/\tilde{n}$ in [56]. That is, "small" and "critical" in [56] are equivalent to "$1/\tilde{n}$-small" and "$1/\tilde{n}$-critical" here. The pathset complexity lower bound in Sections 8-11 of [56] treat $\tilde{n}$ as an arbitrary parameter. (The setting $\tilde{n} = n^{1-1/\log k}$ is only used in the specific application of the Theorem 5.8 in Sections 6-7.)

3. In the present paper, a *pathset* $\mathcal{A}$ is defined as a subset of $\mathscr{G}_A$ (i.e. a set of A-section). In [56], a pathset $\mathcal{A}$ is defined as a subset of $[n]^{V_A}$ (i.e. a "V_A-ary relation" on $[n]$). These definitions are equivalent, since $\mathscr{G}_A$ and $[n]^{V_A}$ are in bijective correspondence (each $A' \in \mathscr{G}_A$ has $E_{A'} = \{v^{(\iota_v)}w^{(\iota_w)} : vw \in E_A\}$ for a unique $\iota \in [n]^{V_A}$). While the view of $\mathcal{A} \subseteq \mathscr{G}_A$ is natural for us here, the relational perspective was convenient in [56] (which dealt with *projections* of pathsets, in addition to joins and restrictions).

B Proof of Lemma 3.7 (Persistent Minterms Under $\vee$ and $\wedge$)

Proof of Lemma 3.7. To simplify notation, we write f_s for $f^{\vee\rho_s}$ and g_s for $g^{\vee\rho_s}$.

Proof of (2): 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 (3): 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 $y_0 \geq y_1 \geq \cdots \geq y_\ell$ and $z_0 \geq z_1 \geq \cdots \geq z_\ell$ 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 $y, z \leq x$. Note that $(f_s \wedge g_s)(y \vee z) = 1$ and $y \vee z \leq x$. 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 $f_{s+i-1} \leq f_{s+i}$ and $g_{s+i-1} \leq g_{s+i}$, 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 $y \leq y_{i-1}$ and $z \leq z_{i-1}$. Note that $(f_{s+i} \wedge g_{s+i})(y \vee z) = 1$ and $y \vee z \leq x$. 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 $\vec{y}, \vec{z} \in \mathrm{Seq}^{\ell}_{\geq}(\{0,1\}^n)$, we finish the proof using Lemma 3.4. Since $\ell \geq \left\langle {d \atop |x|} \right\rangle \geq \left\langle {d \atop |y_0|} \right\rangle$, we may apply Lemma 3.4 to the reversed sequence $(y_\ell, y_{\ell-1}, \ldots, y_0) \in \mathrm{Seq}^{\ell}_{\leq}(\{0,1\}^n)$; we get $0 \leq a \leq b \leq \ell$ such that $y_a = y_b$ and $b - a \geq \left\langle {d-1 \atop |y_a|} \right\rangle$. Therefore, $y_a \in \mathcal{M}^{\vec{\rho}}_{d-1}(f)$. Similarly, we get $z_c \in \mathcal{M}^{\vec{\rho}}_{d-1}(g)$ for some $0 \leq c \leq \ell$. Since $y_0 \leq y_a \leq y_\ell$ and $z_0 \leq z_c \leq z_\ell$ and $z_0 \vee y_0 = y_\ell \vee z_\ell = x$, we conclude that $y_a \vee z_c = x$. □

C Proof of Lemma 4.5 (Persistent k-Cycle Minterms)

Definition C.1. We define a Markov chain $\Gamma_0 \subseteq \Gamma_1 \subseteq \cdots$ in $\mathscr{G}$ by the following process:

- Let Γ_0 be the random graph Γ conditioned on k-CYCLE$(\Gamma) = 0$.
- Let $\circlearrowleft_1, \circlearrowleft_2, \cdots$ be independent uniform random k-cycles and let $\Gamma_t := \Gamma_0 \cup \circlearrowleft_1 \cup \cdots \cup \circlearrowleft_t$ $(= \Gamma_{t-1} \cup \circlearrowleft_t)$ for all $t \geq 1$.

By standard results in probability theory, the number of k-cycles in Γ is asymptotically Poisson with mean $\ln 2$. In particular, the probability of Γ having more than (say) $\log^2 n$ k-cycles is negligibly small. We will show that Γ is well-approximated by Γ_t for $t \sim \mathrm{Poisson}(\ln 2)$ (where "well-approximated" means total variation distance $O(1/n^{0.49})$). For this purpose, we only care about very small values of t (say $\leq \log^2 n$). For such t, note that $\circlearrowleft_1, \ldots, \circlearrowleft_t$ are very likely to be the only k-cycles Γ_t (where "very likely" means with probability better than $1 - O(1/n^{0.99})$).

Definition C.2. We define random variables τ and σ over $\mathbb{N}$:

- Let τ be Poisson with mean $\ln 2$. That is, $\mathbb{P}[\,\tau = t\,] = (\ln 2)^t / 2t!$ for all $t \in \mathbb{N}$. Note that $\mathbb{P}[\,\tau = 0\,] = \mathbb{P}[\,\tau \geq 1\,] = 1/2$.
- Let σ be the random variable with probability mass function[4]

$$\mathbb{P}[\,\sigma = s\,] = (\ln 2)^{-1}\, \mathbb{P}[\,\tau \geq s + 1\,].$$

[4]To see that $\sum_{s=0}^{\infty} \mathbb{P}[\,\sigma = s\,] = 1$, observe that (for any $\lambda > 0$)

$$\sum_{s=0}^{\infty} \mathbb{P}\left[\,\mathrm{Pois}(\lambda) \geq s + 1\,\right] = \sum_{t=1}^{\infty} t \cdot \mathbb{P}\left[\,\mathrm{Pois}(\lambda) = t\,\right] = \sum_{t=1}^{\infty} t \cdot \frac{\lambda^t e^{-\lambda}}{t!} = \lambda \sum_{t=0}^{\infty} \frac{\lambda^{t-1} e^{-\lambda}}{(t-1)!} = \lambda.$$

We denote by Γ_τ (resp. Γ_σ) the random graph Γ_t where $t \in \mathbb{N}$ is sampled according to τ (resp. σ).

The next lemma gives our reason for considering σ.

Lemma C.3. *For every monotone $f : \mathscr{G} \to \{0,1\}$,*

$$\mathbb{P}\left[\, f(\Gamma_\tau) = k\text{-CYCLE}(\Gamma_\tau) \,\right] - 1/2 = \ln 2 \cdot \mathbb{P}\left[\, f(\Gamma_\sigma) \neq f(\Gamma_\sigma \cup \circlearrowleft) \,\right]. \tag{20}$$

Note that (20) is an exact equality, which relates the correlation of f and k-CYCLE on Γ_τ to the probability that f distinguishes the two sides of the monotone coupling $(\Gamma_\sigma, \Gamma_\sigma \cup \circlearrowleft)$.

Proof. First, we have

$$\begin{aligned}
\mathbb{P}\left[\, f(\Gamma_\tau) = k\text{-CYCLE}(\Gamma_\tau) \,\right] &= \mathbb{P}\left[\, \tau = 0 \,\right] \mathbb{P}\left[\, f(\Gamma_0) = 0 \,\right] + \mathbb{P}\left[\, \tau \geq 1 \,\right] \mathbb{P}\left[\, f(\Gamma_\tau) = 1 \;\middle|\; \tau \geq 1 \,\right] \\
&= \frac{1}{2}\Big(\mathbb{P}\left[\, f(\Gamma_0) = 0 \,\right] + \mathbb{P}\left[\, f(\Gamma_\tau) = 1 \;\middle|\; \tau \geq 1 \,\right] \Big) \\
&= \frac{1}{2}\Big(1 - \mathbb{E}\left[\, f(\Gamma_0) \,\right] + \mathbb{E}\left[\, f(\Gamma_\tau) \;\middle|\; \tau \geq 1 \,\right] \Big) \\
&= \frac{1}{2}\Big(1 + \mathbb{E}\left[\, f(\Gamma_\tau) - f(\Gamma_0) \;\middle|\; \tau \geq 1 \,\right] \Big).
\end{aligned} \tag{21}$$

Using the fact that f is monotone and $\Gamma_0 \subseteq \Gamma_1 \subseteq \cdots$, we continue

$$\begin{aligned}
\mathbb{E}\left[\, f(\Gamma_\tau) - f(\Gamma_0) \;\middle|\; \tau \geq 1 \,\right] &= \mathbb{P}\left[\, f(\Gamma_0) \neq f(\Gamma_\tau) \;\middle|\; \tau \geq 1 \,\right] \\
&= \textstyle\sum_{t=1}^{\infty} \mathbb{P}\left[\, \tau = t \;\middle|\; \tau \geq 1 \,\right] \mathbb{P}\left[\, f(\Gamma_0) \neq f(\Gamma_t) \,\right] \\
&= 2 \textstyle\sum_{t=1}^{\infty} \mathbb{P}\left[\, \tau = t \,\right] \sum_{s=0}^{t-1} \mathbb{P}\left[\, f(\Gamma_s) \neq f(\Gamma_{s+1}) \,\right] \\
&= 2 \textstyle\sum_{s=0}^{\infty} \mathbb{P}\left[\, \tau \geq s+1 \,\right] \mathbb{P}\left[\, f(\Gamma_s) \neq f(\Gamma_{s+1}) \,\right] \\
&= 2 \ln 2 \textstyle\sum_{s=0}^{\infty} \mathbb{P}\left[\, \sigma = s \,\right] \mathbb{P}\left[\, f(\Gamma_s) \neq f(\Gamma_{s+1}) \,\right] \\
&= 2 \ln 2 \cdot \mathbb{P}\left[\, f(\Gamma_\sigma) \neq f(\Gamma_{\sigma+1}) \,\right].
\end{aligned} \tag{22}$$

To complete the proof, we plug (22) into (21) and observe that $(\Gamma_\sigma, \Gamma_{\sigma+1}) \overset{\mathrm{d}}{=} (\Gamma_\sigma, \Gamma_\sigma \cup \circlearrowleft)$. □

Next we state three lemmas giving bounds on the total variation distance between various random graphs which we consider.

Lemma C.4 (Total Variation Distance Bounds)**.**

$$d_{\mathrm{TV}}(\Gamma, \Gamma_\tau) = O(k/\sqrt{n}), \tag{23}$$

$$d_{\mathrm{TV}}(\Gamma, \Gamma \cup \circlearrowleft^{-e}) = O(k/\sqrt{n}), \tag{24}$$

$$d_{\mathrm{TV}}(\Gamma_\sigma, \Gamma_\sigma \cup \Xi_\ell) = O(\ell k/\sqrt{n}). \tag{25}$$

Proof. We first observe that (25) follows from (24) by the triangle inequality and standard Markov chain coupling arguments:[5]

$$\begin{aligned} d_{\mathrm{TV}}(\Gamma_\sigma, \Gamma_\sigma \cup \Xi_\ell) &\le d_{\mathrm{TV}}(\Gamma_0, \Gamma_0 \cup \Xi_\ell) \\ &\le \textstyle\sum_{i=1}^{\ell} d_{\mathrm{TV}}\big(\Gamma_0 \cup \Xi_1^{(1)} \cup \cdots \cup \Xi_1^{(i-1)},\, \Gamma_0 \cup \Xi_1^{(1)} \cup \cdots \cup \Xi_1^{(i)}\big) \\ &\le \ell \cdot d_{\mathrm{TV}}(\Gamma_0 \cup \Xi_1) \\ &\le 2\ell \cdot d_{\mathrm{TV}}(\Gamma \cup \circlearrowleft^{-e}). \end{aligned}$$

For (23) and (24), we use the Poisson approximation theorems of Arratia, Goldstein and Gordon [10]. (These results not only show that the number of k-cycles in Γ asymptotically Poisson with mean $\ln 2$, but give a tight approximation to the entire *process* of indicators of k-cycles.) The proof of (23) and (24) will be included in the full version of this paper. □

Proof of Lemma 4.5. Let $f : \mathscr{G} \to \{0,1\}$ be any monotone function and let

$$\delta := \mathbb{P}\left[\, f(\Gamma) = k\text{-CYCLE}(\Gamma) \,\right] - 1/2.$$

Note that $\circlearrowleft \notin \mathcal{M}(f^{\cup\Gamma_\sigma}) \cap \mathcal{M}(f^{\cup\Gamma_\sigma \cup \Xi_\ell})$ if, and only if, $f(\Gamma_\sigma \cup \circlearrowleft) = 0$ or there exists $e \in E_K$ such that $f(\Gamma_\sigma \cup \Xi_\ell \cup \circlearrowleft^{-e}) = 1$. We have

$$\begin{aligned} (26)\quad &\mathop{\mathbb{P}}_{\Gamma_\sigma, \Xi_\ell, \circlearrowleft}\left[\, \circlearrowleft \notin \mathcal{M}(f^{\cup\Gamma_\sigma}) \cap \mathcal{M}(f^{\cup\Gamma_\sigma \cup \Xi_\ell}) \,\right] \\ &\qquad = \mathbb{P}\left[\big(f(\Gamma_\sigma \cup \circlearrowleft) = 0\big) \vee \bigvee_{e \in E_K} \big(f(\Gamma_\sigma \cup \Xi_\ell \cup \circlearrowleft^{-e}) = 1\big) \right] \\ &\qquad \le \mathbb{P}\left[\big(f(\Gamma_\sigma \cup \circlearrowleft) = 0\big) \vee \big(f(\Gamma_\sigma) = 1\big) \vee \bigvee_{e \in E_K} \big(f(\Gamma_\sigma \cup \Xi_\ell \cup \circlearrowleft^{-e}) \ne f(\Gamma_\sigma)\big) \right] \\ &\qquad \le \mathbb{P}\left[\, \big(f(\Gamma_\sigma \cup \circlearrowleft) = 0\big) \vee \big(f(\Gamma_\sigma) = 1\big) \,\right] + \sum_{e \in E_K} d_{\mathrm{TV}}(\Gamma_\sigma, \Gamma_\sigma \cup \Xi_\ell \cup \circlearrowleft^{-e}) \\ &\qquad \le \mathbb{P}\left[\, \big(f(\Gamma_\sigma \cup \circlearrowleft) = 0\big) \vee \big(f(\Gamma_\sigma) = 1\big) \,\right] + O\left(\frac{(\ell+1)k^2}{\sqrt{n}}\right) \quad \text{(by (24) and (25))}. \end{aligned}$$

Since f is monotone and $\Gamma_\sigma \subseteq \Gamma_\sigma \cup \circlearrowleft$, we have $f(\Gamma_\sigma) = f(\Gamma_\sigma \cup \circlearrowleft)$ if, and only if, $f(\Gamma_\sigma \cup \circlearrowleft) = 0$ or $f(\Gamma_\sigma) = 1$. Therefore,

$$\begin{aligned} (27)\quad &\mathbb{P}\left[\, \big(f(\Gamma_\sigma \cup \circlearrowleft) = 0\big) \vee \big(f(\Gamma_\sigma) = 1\big) \,\right] \\ &\qquad = 1 - \mathbb{P}\left[\, f(\Gamma_\sigma) \ne f(\Gamma_\sigma \cup \circlearrowleft) \,\right] \\ &\qquad = 1 - \frac{1}{\ln 2}\left(\mathbb{P}\left[\, f(\Gamma_\tau) = k\text{-CYCLE}(\Gamma_\tau) \,\right] - \frac{1}{2} \right) \quad \text{(by Lemma C.3)} \\ &\qquad \le 1 - \frac{1}{\ln 2}\left(\mathbb{P}\left[\, f(\Gamma) = k\text{-CYCLE}(\Gamma) \,\right] - d_{\mathrm{TV}}(\Gamma, \Gamma_\tau) - \frac{1}{2} \right) \\ &\qquad \le 1 - \frac{\delta}{\ln 2} + O\left(\frac{k}{\sqrt{n}}\right) \quad \text{(by (23))}. \end{aligned}$$

[5] Let P be the transition matrix of a Markov chain with state space Ω. Then for all probability distributions μ and ν on Ω, $d_{\mathrm{TV}}(\mu P, \nu P) \le d_{\mathrm{TV}}(\mu, \nu)$.

We now have

$$\min_{G\in\mathscr{G}} \mathop{\mathbb{P}}_{\Xi_\ell,\circlearrowleft}\left[\circlearrowleft \notin \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G\cup\Xi_\ell})\right] \le \mathop{\mathbb{P}}_{\Gamma_\sigma,\Xi_\ell,\circlearrowleft}\left[\circlearrowleft \notin \mathcal{M}(f^{\cup\Gamma_\sigma}) \cap \mathcal{M}(f^{\cup\Gamma_\sigma\cup\Xi_\ell})\right]$$
$$\le 1 - \frac{\delta}{\ln 2} + O\left(\frac{(\ell+1)k^2}{\sqrt{n}}\right) \quad \text{(by (26) and (27))}. \tag{28}$$

Fixing an optimal G in (28), we have (by Markov's Inequality)

$$\begin{aligned}
&\mathop{\mathbb{P}}_{\Xi_\ell}\left[\mathop{\mathbb{P}}_{\circlearrowleft}\left[\circlearrowleft \in \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G\cup\Xi_\ell})\right] \ge \frac{1}{\sqrt{n}}\right] \\
&\qquad = 1 - \mathop{\mathbb{P}}_{\Xi_\ell}\left[1 - \mathop{\mathbb{P}}_{\circlearrowleft}\left[\circlearrowleft \notin \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G\cup\Xi_\ell})\right] \ge 1 - \frac{1}{\sqrt{n}}\right] \\
&\qquad \ge 1 - \frac{1}{1-1/\sqrt{n}}\left(1 - \mathop{\mathbb{P}}_{\Xi_\ell,\circlearrowleft}\left[\circlearrowleft \notin \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G\cup\Xi_\ell})\right]\right) \\
&\qquad \ge 1 - \mathop{\mathbb{P}}_{\Xi_\ell,\circlearrowleft}\left[\circlearrowleft \notin \mathcal{M}(f^{\cup G}) \cap \mathcal{M}(f^{\cup G\cup\Xi_\ell})\right] - \frac{1}{\sqrt{n}} \\
&\qquad \ge \frac{\delta}{\ln 2} - O\left(\frac{(\ell+1)k^2}{\sqrt{n}}\right) \quad \text{(by (28))}.
\end{aligned}$$

If δ is larger than $C(\ell+1)k^2/\sqrt{n}$ (for a constant C depending on the constant in the final big-O term), then the final bound is $\ge 1/\sqrt{n}$. Equation (5) follows, which completes the proof. □

D Proof of Lemma 5.13 (Smallness)

The following definition captures the minimal obstructions to ε-smallness.

Definition D.1. A pathset $\mathcal{A}$ is *ε-critical* if

- $\mathsf{density}(\mathcal{A}) > \varepsilon^{|V_A|-|E_A|}$ and
- $\mathsf{density}(\mathcal{B}) \le \varepsilon^{|V_B|-|E_B|}$ for every proper restriction $\mathcal{B} \prec \mathcal{A}$.

Note that a pathset $\mathcal{A}$ is <u>not</u> ε-small if, and only if, some restriction $\mathcal{B} \preceq \mathcal{A}$ is ε-critical. (This is immediate from Definition 5.6 of ε-smallness.)

The key technical step in the proof of Lemma 5.13 is the following lemma on ε-critical pathsets. This lemma is very similar to Lemma 7.3 of [56] and Theorem 4.4 ("Quasi-sunflower Lemma") of [55]. The proof uses Janson's Inequality [33] in an efficient way.

Lemma D.2. *Suppose $\mathcal{A}$ is an ε-critical pathset. Let $A_1,\dots,A_t$ where $t = |V_A| - |E_A|$ enumerate the connected components of A (in any order). Let $\mathbf{R}_1,\dots,\mathbf{R}_t$ be independent random pathsets $\mathbf{R}_i \subseteq_{p_i} \mathscr{G}_{A_i}$ (that is, $\mathbb{P}[A_i' \in \mathbf{R}_i] = p_i$ independently for all $A_i' \in \mathscr{G}_{A_i}$) where $p_i = \ell/n^{|V_{A_i}|}$ and $\ell \ge k/\varepsilon$. Then*

$$\mathbb{P}\left[\mathcal{A} \cap (\mathbf{R}_1 \bowtie \cdots \bowtie \mathbf{R}_t) = \emptyset\right] \le \exp(-\varepsilon\ell/2k).$$

Proof. Note that $1 \le t \le k/2$. For $S \subseteq [t]$, let $\overline{S} := [t] \setminus S$.

For every $A' \in \mathcal{A}$, let $\mathbf{I}_{A'} \in \{0,1\}$ be the indicator for the event that $A' \in \mathbf{R}_1 \bowtie \cdots \bowtie \mathbf{R}_t$. Thus,

$$\mathcal{A} \cap (\mathbf{R}_1 \bowtie \cdots \bowtie \mathbf{R}_t) = \emptyset \;\Leftrightarrow\; \sum_{A' \in \mathcal{A}} \mathbf{I}_{A'} = 0. \tag{29}$$

For $i \in [t]$, let A'_i be the subgraph of A' which projects to A_i (i.e. A'_i is the unique A_i-section which is a subgraph of A'). Note that A' is the edge-disjoint union of graphs A'_i.

For $S \subseteq [t]$, let $A_S := \bigcup_{i \in S} A_i$ and let $\mathcal{A}^2_S$ denote the set of pairs

$$\mathcal{A}^2_S := \left\{ (A', A'') \in \mathcal{A}^2 : \bigwedge_{i \in S} (A'_i \neq A''_i) \wedge \bigwedge_{i \in \overline{S}} (A'_i \neq A''_i) \right\}.$$

Note that $\mathcal{A}^2$ is the disjoint union of sets $\mathcal{A}^2_S$.

For $A', A'' \in \mathcal{A}$, observe that $\mathbf{I}_{A'}$ and $\mathbf{I}_{A''}$ are independent if, and only if, $(A', A'') \in \mathcal{A}^2_\emptyset$. Define

$$\lambda := \sum_{A' \in \mathcal{A}} \mathbb{E}[\,\mathbf{I}_{A'}\,], \qquad \Upsilon_S := \sum_{(A', A'') \in \mathcal{A}^2_S} \mathbb{E}[\,\mathbf{I}_{A'} \mathbf{I}_{A''}\,], \qquad \Upsilon := \sum_{\emptyset \subset S \subset [t]} \Upsilon_S.$$

In this context, Janson's Inequality [33] states

$$\mathbb{P}\left[\sum_{A' \in \mathcal{A}} \mathbf{I}_{A'} = 0 \right] \le \exp\left(-\min\left\{ \frac{\lambda}{2}, \frac{\lambda^2}{2\Upsilon} \right\} \right). \tag{30}$$

In light of (29), it suffices to prove that $\min\{\lambda, \lambda^2/\Upsilon\} \ge \varepsilon \ell / k$.

First we bound λ. For all $A' \in \mathcal{A}$, we have

$$\mathbb{E}[\,\mathbf{I}_{A'}\,] = \mathbb{P}[\,A' \in \mathbf{R}_1 \bowtie \cdots \bowtie \mathbf{R}_t\,] = \prod_{i \in [t]} [\,A'_i \in \mathbf{R}_i\,] = \prod_{i \in [t]} p_i = \frac{\ell^t}{n^{|V_A|}}.$$

Since $\mathcal{A}$ is ε-critical, we have $\mathsf{density}(\mathcal{A}) > \varepsilon^t$, hence $|\mathcal{A}| > \varepsilon^t \cdot n^{|V_A|}$. Therefore,

$$\lambda = |\mathcal{A}| \cdot \mathbb{E}[\,\mathbf{I}_{A'}\,] = \frac{|\mathcal{A}| \cdot \ell^t}{n^{|V_A|}} > (\varepsilon \ell)^t. \tag{31}$$

Next we bound Υ. For each $\emptyset \subset S \subset [t]$, we have

$$\begin{aligned}
|\mathcal{A}^2_S| &\le \sum_{\overline{B}' \in \mathscr{G}_{A_{\overline{S}}}} |\{B' \in \mathscr{G}_{A_S} : B' \cup \overline{B}' \in \mathcal{A}\}|^2 \\
&\le \sum_{\overline{B}' \in \mathscr{G}_{A_{\overline{S}}}} |\{B' \in \mathscr{G}_{A_S} : B' \cup \overline{B}' \in \mathcal{A}\}| \cdot \max_{\overline{B}' \in \mathscr{G}_{A_{\overline{S}}}} |\{B' \in \mathscr{G}_{A_S} : B' \cup \overline{B}' \in \mathcal{A}\}| \\
&= |\mathcal{A}| \cdot \max_{\overline{B}' \in \mathscr{G}_{A_{\overline{S}}}} |\{B' \in \mathscr{G}_{A_S} : B' \cup \overline{B}' \in \mathcal{A}\}| \\
&= |\mathcal{A}| \cdot n^{|V_{A_S}|} \cdot \max_{\overline{B}' \in \mathscr{G}_{A_{\overline{S}}}} \mathsf{density}(\{B' \in \mathscr{G}_{A_S} : B' \cup \overline{B}' \in \mathcal{A}\}) \\
&\le |\mathcal{A}| \cdot n^{|V_{A_S}|} \cdot \varepsilon^{|V_{A_S}| - |E_{A_S}|} && \text{(by } \varepsilon\text{-criticality of } \mathcal{A}) \\
&= |\mathcal{A}| \cdot n^{|V_{A_S}|} \cdot \varepsilon^{|S|} && \text{(since } |V_{A_S}| - |E_{A_S}| = |S|).
\end{aligned} \tag{32}$$

For all $(A', A'') \in \mathcal{A}_S^2$, we have

$$\mathbb{E}[\,\mathbf{I}_{A'}\mathbf{I}_{A''}\,] = \prod_{i\in[t]}[\,A'_i \in \mathbf{R}_i\,] \cdot \prod_{j\in S}[\,A''_j \in \mathbf{R}_j\,] = \prod_{i\in[t]} p_i \cdot \prod_{j\in S} p_j = \frac{\ell^{t+|S|}}{n^{|V_A|+|V_{A_S}|}}. \tag{33}$$

Therefore,

$$\begin{aligned}
\Upsilon_S = |\mathcal{A}_S^2| \cdot \mathbb{E}[\,\mathbf{I}_{A'}\mathbf{I}_{A''}\,] &= \frac{|\mathcal{A}_S^2| \cdot \ell^{t+|S|}}{n^{|V_A|+|V_{A_S}|}} && \text{(by (33))} \\
&\le \frac{|\mathcal{A}| \cdot \varepsilon^{|S|} \cdot \ell^{t+|S|}}{n^{|V_A|}} && \text{(by (32))} \\
&= \lambda(\varepsilon\ell)^{|S|} && \text{(by (31))}.
\end{aligned}$$

Since $\varepsilon\ell \ge k \ge 2t$,

$$\Upsilon = \sum_{\emptyset\subset S\subset[t]} \Upsilon_S \le \sum_{i=1}^{t-1} \binom{t}{i} \cdot \lambda(\varepsilon\ell)^i \le \lambda t(\varepsilon\ell)^{t-1} \cdot \sum_{i=1}^{t-1} \left(\frac{t}{\varepsilon\ell}\right)^{i-1} \le \lambda k(\varepsilon\ell)^{t-1}.$$

Using this upper bound on Υ and the lower bound $\lambda > (\varepsilon\ell)^t$ (31),

$$\frac{\lambda^2}{\Upsilon} \ge \frac{\lambda}{k(\varepsilon\ell)^{t-1}} > \frac{\varepsilon\ell}{k}.$$

Plugging the bounds on λ and λ^2/Υ into (30) completes the proof. □

Lemma D.3. *For every ε-critical pathset $\mathcal{A}$ and $\ell \ge 2k^2/\varepsilon$,*

$$\mathop{\mathbb{P}}_{\Xi_\ell}\Big[\bigwedge_{A'\in\mathcal{A}} A' \not\subseteq \Xi_\ell\Big] \le \exp(-\varepsilon\ell/4k^2).$$

Proof. Let $A = A_1 \uplus \cdots \uplus A_t$ as in Lemma D.2. Fix distinct edges $\eta_1, \ldots, \eta_t \in E_K$ such that $\eta_i \notin E_{A_i}$ for all $i \in [t]$. The random graph Ξ_ℓ is equivalent to $\circlearrowleft_1^{-e_1} \cup \cdots \cup \circlearrowleft_\ell^{-e_\ell}$ where $\circlearrowleft_1, \ldots, \circlearrowleft_\ell$ are independent uniform k-cycles and $e_1, \ldots, e_\ell$ are independent uniform edges in E_K. For $i \in [t]$, let $\mathbf{S}_i \subseteq \mathscr{G}_{A_i}$ be the A_i-pathset consisting of the A_i-subsections of all $\circlearrowleft_j^{-e_j}$ such that $e_j = \eta_i$. Intuitively, while not a product distribution, $\mathbf{S}_i$ is similar to a random subset of $\mathscr{G}_{A_i}$ of size ℓ/k.

Let $\mathbf{R}_i \subseteq_{p_i} \mathscr{G}_{A_i}$ for $p_i = \ell/2kn^{|V_{A_i}|}$ (i.e. half the expected density of $\mathbf{S}_i$). It is easy to show that $\mathbf{S}_i$ stochastically dominates $\mathbf{R}_i$. Therefore, $\mathbf{S}_1 \bowtie \ldots \bowtie \mathbf{S}_t$ stochastically dominates $\mathbf{R}_1 \bowtie \ldots \bowtie \mathbf{R}_t$. By Lemma D.2,

$$\begin{aligned}
\mathop{\mathbb{P}}_{\Xi_\ell}\Big[\bigwedge_{A'\in\mathcal{A}} A' \not\subseteq \Xi_\ell\Big] &= \mathop{\mathbb{P}}_{\mathbf{S}_1,\ldots,\mathbf{S}_t}\big[\,\mathcal{A} \cap (\mathbf{S}_1 \bowtie \ldots \bowtie \mathbf{S}_t) = \emptyset\,\big] \\
&\le \mathop{\mathbb{P}}_{\mathbf{R}_1,\ldots,\mathbf{R}_t}\big[\,\mathcal{A} \cap (\mathbf{R}_1 \bowtie \ldots \bowtie \mathbf{R}_t) = \emptyset\,\big] \\
&\le \exp(-\varepsilon\ell/4k^2).
\end{aligned}$$

□

Proof of Lemma 5.13. Let $f : \mathscr{G} \to \{0,1\}$ be a monotone function, let $A \subset K$, let $\varepsilon > 0$, and let $\ell \in \mathbb{N}$. Our goal is to show

$$\mathop{\mathbb{P}}_{\Xi_\ell}\big[\,\mathcal{M}_A(f) \cap \mathcal{M}_A(f^{\cup\Xi_\ell}) \text{ is not } \varepsilon\text{-small}\,\big] \le (2n)^k \exp\big(-\Omega(\varepsilon\ell/k^2)\big). \tag{34}$$

We choose a sequence $f =: f_0 < \cdots < f_r$ by the following process:

- If $\mathcal{M}_A(f_{i-1})$ is ε-small, then halt (i.e. $r = i-1$).
- If $\mathcal{M}_A(f_{i-1})$ is not ε-small, then pick any ε-critical restriction $\mathcal{B}_i \preceq \mathcal{M}_A(f_{i-1})$ and set

$$f_i := f_{i-1} \vee \mathrm{Ind}_{C_i'}$$

where $C_i' \in \mathscr{G}_{A \setminus B_i}$ is the unique $A \setminus B_i$-section such that $\mathcal{B}_i = \{B_i' \in \mathscr{G}_{B_i} : B_i' \cup C_i' \in \mathcal{M}_A(f_{i-1})\}$ and $\mathrm{Ind}_{C_i'} : \mathscr{G} \to \{0,1\}$ is the indicator function which takes value 1 on G if and only if $C_i' \subseteq G$.

(In the language of (quasi-)sunflowers, we are plucking the petals in $\mathcal{B}_i$ and adding a minterm at the core C_i'.)

Note that $r \leq (2n)^k$, since each C_i' shows up at most once in the construction of f_r and there are at most $(2n)^k$ possible C_i' (i.e. 2^{E_A} $(\leq 2^k)$ possibilities for $C_i \subseteq A$ and $n^{V_{C_i}}$ $(\leq n^k)$ possibilities for $C_i' \in \mathscr{G}_{C_i}$).

Claim D.4. *Let $H \in \mathscr{G}$ and $i \in \{1, \ldots, r\}$ and suppose there exists $B_i' \in \mathcal{B}_i$ such that $B_i' \subseteq H$. Then*

$$\mathcal{M}_A(f_{i-1}) \cap \mathcal{M}_A(f_{i-1}^{\cup H}) \subseteq \mathcal{M}_A(f_i) \cap \mathcal{M}_A(f_i^{\cup H}). \tag{35}$$

To prove the claim, consider any $X \in \mathcal{M}_A(f_{i-1}) \cap \mathcal{M}_A(f_{i-1}^{\cup H})$. We must show that $X \in \mathcal{M}_A(f_i) \cap \mathcal{M}_A(f_i^{\cup H})$. Since $f_i \leq f_i^{\cup H}$, this is equivalent to showing that $f_i(X) = 1$ and $f_i(Y \cup H) = 0$ for all $Y \subset X$. Since X is a minterm of f_{i-1} and $f_{i-1} \leq f_i$, we have $f_i(X) = 1$.

Now consider any $Y \subset X$. Since X is a minterm of $f_{i-1}^{\cup H}$, we have $f_{i-1}(Y \cup H) = 0$. Since $f_i = f_{i-1} \vee \mathrm{Ind}_{C_i'}$, it remains to show that $\mathrm{Ind}_{C_i'}(Y \cup H) = 0$, that is, $C_i' \nsubseteq Y \cup H$. This is easiest to argue by contradiction. Assume (for contradiction) that $C_i' \subseteq Y \cup H$. Since $B_i' \subseteq H$, we have $B_i' \cup C_i' \subseteq Y \cup H$. On the other hand, since $B_i' \cup C_i' \in \mathcal{M}_A(f_{i-1})$, we have $f_{i-1}(B_i' \cup C_i') = 1$ and hence $f_{i-1}(Y \cup H) = 1$. This contradicts the fact that $f_{i-1}(Y \cup H) = 0$ (as we already noted, since X is a minterm of $f_{i-1}^{\cup H}$).

From Claim D.4, we have following implication: for all $H \in \mathscr{G}$ and $A \subseteq K$, if $\mathcal{M}_A(f) \cap \mathcal{M}_A(f^{\cup H})$ is not ε-small, then there exist $i \in \{1, \ldots, r\}$ such that $B_i' \nsubseteq H$ for all $B_i' \in \mathcal{B}_i$. Equation (34) now follows by Lemma D.3:

$$\begin{aligned} \mathop{\mathbb{P}}_{\Xi_\ell}\left[\mathcal{M}_A(f) \cap \mathcal{M}_A(f^{\cup \Xi_\ell}) \text{ is not } \varepsilon\text{-small}\right] &\leq \sum_{1 \leq i \leq r} \mathop{\mathbb{P}}_{\Xi_\ell}\Big[\bigwedge_{B_i' \in \mathcal{B}_i} B_i' \nsubseteq \Xi_i\Big] \\ &\leq (2n)^k \exp\left(-\Omega(\varepsilon \ell / k^2)\right). \end{aligned}$$

□

E Proofs of Lemma 5.12 (Pathset Complexity and Formula Size)

Proof of Lemma 5.12. Assume Φ is a monotone formula and $\vec{\rho} \in \mathrm{Seq}_{\leq}^m(\mathscr{G})$ 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 $\sum_{A \subseteq K} |\mathcal{P}_A^{\vec{\rho}}(\phi)| \leq 1$.

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 $\mathcal{P}^{\vec{\rho}}_A(\Psi) \subseteq \mathcal{P}^{\vec{\rho}}_A(\Psi_1) \cup \mathcal{P}^{\vec{\rho}}_A(\Psi_2)$. By properties (monotonicity) and (sub-additivity) of χ_ε, it follows that

$$\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi)) \leq \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_1)) + \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_2)). \tag{36}$$

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

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

(This expression extracts from (14) 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,

$$\begin{aligned} \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi)) &\leq \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_1)) + \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_2)) + \sum_{B,C \subsetneq A \,:\, B \cup C = A} \Big(\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_B(\Psi_1)) + \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_C(\Psi_2))\Big) \\ &\leq \Big(\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_1)) + 2^k \sum_{B \subsetneq A} \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_B(\Psi_1))\Big) + \Big(\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_A(\Psi_2)) + 2^k \sum_{B \subsetneq A} \chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_B(\Psi_2))\Big). \end{aligned} \tag{37}$$

If we now start with $\chi_\varepsilon(\mathcal{P}^{\vec{\rho}}_K(\Phi))$ and repeatedly expand according to (37) and (36) 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 certain coefficients $c_{\phi,A} \in \mathbb{N}$. 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 (37)). From this explanation, we extract an 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 $\phi \in \mathrm{Leaves}(\Phi)$ (which we established earlier), together with $\mathrm{size}(\Phi) = |\mathrm{Leaves}(\Phi)|$ (here is where we use the fact that Φ is a formula), we conclude

$$\mathcal{P}^{\vec{\rho}}_K(\Phi) \leq 2^{O(k^2)} \cdot \mathrm{depth}(\Phi)^k \cdot \mathrm{size}(\Phi). \qquad \square$$

F 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 F.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 F.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* mon-pairs$(\mathfrak{C}) \subseteq \bigcup_{i=1}^{t'}$ 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 $(x, y) \in$ 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 $(x, y) \in$ mon-pairs$(\mathfrak{M}_\alpha)$. Therefore, mon-pairs$(\mathfrak{C}) \subseteq \bigcup\{$mon-pairs$(\mathfrak{M}_\alpha) : j \in [t+1], \alpha \in \{0,1\}^{j-1}\}$. □

Lemma F.3 ([30])**.** *Let μ_0, μ_1 be two strictly positive probability distributions on $\{0,1\}^n$ which satisfy the "Holley condition"*

$$\mu_0(x)\mu_1(y) \leq \mu_0(x \wedge y)\mu_1(x \vee y) \qquad \text{for all } x, y. \tag{38}$$

Then there exists a probability distribution ν on $\{0,1\}^n \times \{0,1\}^n$ such that

$$\textstyle\sum_y \nu(x, y) = \mu_0(x) \qquad \text{for all } x, \tag{39}$$

$$\textstyle\sum_x \nu(x, y) = \mu_1(y) \qquad \text{for all } y, \tag{40}$$

$$\nu(x, y) = 0 \qquad \text{unless } x \leq y. \tag{41}$$

We call ν satisfying (39), (40), (41) a *monotone coupling* of μ_0 and μ_1. The elementary proof of Lemma F.3 given by Holley [30] uses a Markov chain coupling argument. We remark that Lemma F.3 also follows from an earlier (and much more general) monotone coupling theorem of Strassen [60].

Lemma F.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*

$$\mu_b(x) := \begin{cases} \mu(x)/(1 - \mathbb{E}_\mu(f)) & \text{if } f(x) = b = 0, \\ \mu(x)/\mathbb{E}_\mu(f) & \text{if } f(x) = b = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

Then the pair μ_0, μ_1 satisfy the Holley condition (38).

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

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

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,

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

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

Define μ_0 and μ_1 by (42) as in Lemma F.4. By Lemma F.3 there exists a distribution ν, supported on mon-pairs(f), satisfying (39), (40), (41). Note that, for every monotone function $h : \{0,1\}^n \to \{0,1\}$, we have

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

It follows from Lemma F.2 that there exists a monotone circuit $\mathfrak{M}$, of the same size and depth as $\mathfrak{C}$, such that

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

We finish the proof using two applications of (43):

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