

# Polylogarithmic independence can fool DNF formulas \*

Louay M.J. Bazzi †

## Abstract

We show that any  $k$ -wise independent probability distribution on  $\{0, 1\}^n$  of size  $O(m^{2.2}2^{-\sqrt{k}/10})$  fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables. Thus, for each constant  $e > 0$ , there is a constant  $c > 0$  such that any boolean function computable by an  $m$ -clause DNF (or CNF) formula is  $m^{-e}$ -fooled by any  $c \log^2 m$ -wise probability distribution. This resolves up to an  $O(\log m)$  factor the depth-2 circuits case of a conjecture due to Linial and Nisan (1990). The result is equivalent to a new characterization of DNF (or CNF) formulas by low degree polynomials. It implies a similar statement for probability distributions with the small bias property. Using known explicit constructions of small probability spaces having the limited independence property or the small bias property, we directly obtain a large class of explicit PRG's of  $O(\log^2 m \log n)$ -seed length for  $m$ -clause DNF (or CNF) formulas on  $n$  variables, improving previously known seed lengths.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Dual problem . . . . .	4
1.2	Paper outline . . . . .	5
<b>2</b>	<b>Some direct applications</b>	<b>5</b>
2.1	Extension to small bias probability spaces . . . . .	6
2.2	A large class of PRG's for DNF formulas . . . . .	7
2.3	Patterns in binary linear codes . . . . .	8
<b>3</b>	<b>Fourier transform preliminaries</b>	<b>9</b>
<b>4</b>	<b>LP duality perspective</b>	<b>10</b>

---

\*An extended abstract of this paper appeared in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 63-73, 2007.

†Department of Electrical and Computer Engineering, American University of Beirut, Beirut, Lebanon.  
E-mail: Louay.Bazzi@aub.edu.lb.

<b>5</b>	<b>Outline of proof</b>	<b>11</b>
5.1	Simplifications and notation . . . . .	11
5.2	Approximation notions used in the proof . . . . .	12
5.3	Main steps in the proof . . . . .	13
5.4	Ignoring large clauses . . . . .	13
5.5	From bias to zero-energy . . . . .	14
5.6	Construction overview . . . . .	16
5.7	Skin and cover auxiliary functions . . . . .	18
5.8	From zero-energy to the energies of auxiliary functions . . . . .	19
5.9	Back to DNF formulas . . . . .	20
5.10	Summary . . . . .	21
5.11	Backtracking . . . . .	22
<b>6</b>	<b>Möbius and Fourier analysis of DNF formulas and auxiliary functions</b>	<b>23</b>
6.1	Posets preliminaries . . . . .	23
6.2	Poset $B_n$ . . . . .	23
6.3	Monotone DNF formulas and auxiliary functions . . . . .	25
6.4	The poset $B_n^{(2)}$ . . . . .	28
6.5	General DNF formulas and auxiliary functions . . . . .	30
6.6	Miscellaneous remarks . . . . .	32
<b>7</b>	<b>From zero-energy to energies of auxiliary functions</b>	<b>33</b>
7.1	Monotone case error analysis . . . . .	34
7.2	Discussion . . . . .	37
7.3	General case construction . . . . .	39
7.4	Bounds . . . . .	42
<b>8</b>	<b>Back to DNF formulas</b>	<b>45</b>
8.1	Proof of Theorem 8.1 . . . . .	48
<b>9</b>	<b>A sharper bound</b>	<b>50</b>
<b>10</b>	<b>Optimal solution</b>	<b>54</b>
<b>11</b>	<b>Concluding remarks</b>	<b>57</b>
<b>A</b>	<b>LP duality calculations appendix</b>	<b>60</b>
<b>B</b>	<b>What will not work appendix</b>	<b>62</b>

# 1 Introduction

If  $\mu$  is a probability distribution on  $\{0, 1\}^n$  and  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  is a boolean function, we say that  $\mu$   $\epsilon$ -fools  $g$  [BM82, Yao82] if

$$|Pr_{x \sim \mu}[g(x) = 1] - Pr_{x \in \{0, 1\}^n}[g(x) = 1]| \leq \epsilon,$$

where the second probability is with respect to the uniform probability distribution on  $\{0, 1\}^n$ .

Let  $\mu$  be a probability distribution on  $\{0, 1\}^n$  and let  $k \geq 0$  be an integer. We say that  $\mu$  is  $k$ -wise independent (e.g., [Lub85, Vaz86]) if any  $k$  or less of the underlying  $n$  binary random variables are statistically independent and each is equally likely to be zero or one<sup>1</sup>.

A *DNF (Disjunctive Normal Form) formula* on  $n$  variables  $x_1, \dots, x_n$  is an OR of AND gates, called *clauses*, on the *literals*  $x_1, \neg x_1, \dots, x_n, \neg x_n$ . Similarly, a *CNF (Conjunctive Normal Form) formula* is an AND of OR gates.

We consider in this paper the following problem: how large should  $k$  be in terms of  $\epsilon, n$ , and  $m$  so that any  $k$ -wise independent probability distribution on  $\{0, 1\}^n$   $\epsilon$ -fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables?

The main contribution of this paper is the following theorem.

**Theorem 1.1** *Any  $k$ -wise independent probability distribution on  $\{0, 1\}^n$  ( $16m^{2.2}2^{-\sqrt{k}/10}$ )-fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables.*

The proof is based on harmonic and poset analysis techniques. It uses Hastad's switching Lemma [Has86] indirectly via Linial-Mansour-Nisan energy bound [LMN93], applied to many DNF formulas derived from the DNF formula under consideration. The proof can be regarded as a sequence of reductions between some  $L_1$  and  $L_2$  approximations of DNF formulas and auxiliary functions by low degree polynomials with real coefficients.

**Corollary 1.2** *For each constant  $e > 0$ , there is a constant  $c > 0$  such that any boolean function computable by an  $m$ -clause DNF (or CNF) formula is  $m^{-e}$ -fooled by any  $c \log^2 m$ -wise probability distribution.*

The above problem was first proposed by Linial and Nisan [LN90]. Motivated by this problem, they derived a general bound on approximate inclusion-exclusion from which they concluded that any boolean function computable by an  $m$ -clause DNF (or CNF) formula is  $o(1)$ -fooled by any  $\lfloor \sqrt{m} \log m \rfloor$ -wise independent probability distribution. They conjectured that any boolean function computable by a size- $M$  depth- $d$  unbounded-fanin AND/OR circuit is  $\epsilon$ -fooled by any  $\log^{d-1} M$ -wise independent probability distribution, where  $\epsilon = 0.1$ . Corollary 1.2 resolves up to an  $O(\log m)$  factor this conjecture for depth-2 circuits, reducing the  $\sqrt{m} \log m$  bound of [LN90] to  $O(\log^2 m)$ . Note that the conjecture's strict parameters are not correct: Luby and Velickovic [LV96] reported a counter example which exhibits for

---

<sup>1</sup>For technical convenience, we allow  $k = 0$  in the sense that any probability distribution on  $\{0, 1\}^n$  is 0-wise independent.

each power  $m$  of 2 and for all  $n \geq \log m$  a function  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  computable by an  $m$ -clause DNF formula, and a  $\log m$ -wise independent probability distribution  $\mu$  on  $\{0, 1\}^n$  such that  $\mu$  does not  $\frac{1}{2}$ -fool  $f$ . Theorem 1.1 leaves the region between  $O(\log m)$  and  $o(\log^2 m)$  open for depth-2 circuits.

Next we explain the LP-dual of Theorem 1.1 which is a new approximation of DNF (or CNF) formulas by low degree polynomials and we compare it with the related literature.

## 1.1 Dual problem

Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  be a boolean function,  $k \geq 0$  an integer, and  $\epsilon \geq 0$ . Then saying that

“any  $k$ -wise independent probability distribution  $\epsilon$ -fools  $g$ ”

is equivalent to saying “there exist  $g_l, g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  such that:

- **(low degree<sup>2</sup>)**  $\deg(g_l) \leq k$  and  $\deg(g_u) \leq k$
- **(sandwiching polynomials)**  $g_l \leq g \leq g_u$
- **(small  $L_1$ -approximation error)**  $E(g - g_l) \leq \epsilon$  and  $E(g_u - g) \leq \epsilon$ ,  
where the expectation is over the uniform probability distribution”.

We show that in Section 4 (see Theorem 4.2). Thus the dual problem is about an  $L_1$ -approximation of DNF (or CNF) formulas by low-degree sandwiching polynomials with real coefficients. Via this duality, Theorem 1.1 is (up to a constant factor) equivalent to:

**Theorem 1.3** *Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  be a boolean function computable by an  $m$ -clause DNF (or CNF) formula, and let  $k \geq 0$  be an integer. Then there exist two real-valued functions  $g_l, g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  each of degree at most  $k$  such that  $g_l \leq g \leq g_u$ ,  $E(g - g_l) = O(m^{2.2}2^{-\sqrt{k}/10})$ , and  $E(g_u - g) = O(m^{2.2}2^{-\sqrt{k}/10})$ , where the expectation is over the uniform probability distribution.*

We will actually establish the dual statement.

Theorem 1.3 implies the following weaker  $L_1$ -approximation.

**Corollary 1.4** *Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  be a boolean function computable by an  $m$ -clause DNF (or CNF) formula, and let  $k \geq 0$  be an integer. Then there exists a real-valued function  $p : \{0, 1\}^n \rightarrow \mathbb{R}$  of degree at most  $k$  such that  $E|g - p| = O(m^{2.2}2^{-\sqrt{k}/10})$ .*

**Proof:** Set  $p = g_l$  or  $g_u$ . ■

Small constant-depth unbounded-fanin AND/OR circuits can be approximated by low degree polynomials with real coefficients in different ways [ABFR94, BRS91, LMN93]. They can

---

<sup>2</sup>The degree of a function  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  is the smallest degree of a polynomial  $p \in \mathbb{R}[x_1, \dots, x_n]$  such that  $p(x) = f(x)$  for all  $x \in \{0, 1\}^n$ .

be also approximated by low degree polynomials with coefficients over finite fields [Raz87]. We compare our sandwiching  $L_1$ -approximation with [ABFR94, BRS91, LMN93] specialized to depth-2 circuits. The approximation in [LMN93] is an  $L_2$ -approximation based on Hastad Switching Lemma:

**Theorem 1.5** [LMN93] *Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  be computable by an  $m$ -clause DNF (or CNF) formula, and let  $k \geq 0$  be an integer. Then there exists a real-valued function  $p : \{0, 1\}^n \rightarrow \mathbb{R}$  of degree at most  $k$  such that  $E(g - p)^2 \leq 2m2^{-\sqrt{k}/20}$ .*

The proof of Theorem 1.1 uses a variation of this  $L_2$ -approximation (see Theorem 9.1) applied to many DNF formulas derived from the DNF formula under consideration.

The approximation in [ABFR94, BRS91] does not use Hastad Switching Lemma and is in terms of probabilistic polynomials:

**Theorem 1.6** [ABFR94, BRS91] *Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  be computable by an  $m$ -clause DNF (or CNF) formula. For all  $\epsilon > 0$ , there exists a (finite) family of polynomials  $\{p_\alpha\}_\alpha$ , where each  $p_\alpha$  is a real polynomial (with integer coefficients) on the variables  $x_1, \dots, x_n$  of degree  $O(\log^2(m/\epsilon) \log^2 m \log^2 n)$  such that  $\Pr_x[p_\alpha(x) \neq g(x)] \leq \epsilon$  for all  $x \in \{0, 1\}^n$ . Thus, in particular,  $\Pr_x[p_\alpha(x) \neq g(x)] \leq \epsilon$  for some  $\alpha$ .*

The polynomials do not give a good  $L_1$ -approximation ( $E_x|p_\alpha(x) - g(x)|$  is potentially as large as  $2^{O(\log^2(m/\epsilon) \log^2 m \log^2 n)}$ ), but we believe that they probably can be used indirectly to establish a weak version of Theorem 1.1 which is naturally extensible to  $AC_0$  circuits. See [Baz03] (Sections 5.7 and 5.8) for a work in this direction.

A final remark is that one cannot hope to obtain a good  $L_\infty$ -approximation of DNF formulas by low degree polynomials. This follows from [NS94].

## 1.2 Paper outline

In Section 2, we give direct applications of Theorem 1.1. Section 3 contains Fourier transform preliminaries. Section 4 highlights the dual characterization of the class of boolean functions that are fooled by the limited independence property. The remainder of the paper is about the proof of Theorem 1.1, starting with the proof outline in Section 5. The proof consists of Sections 5, 6, 7, 8, and 9. The proof depends on Sections 3 and 4, but it does not use Section 2 or Section 10, which branches from the proof and ends up with an open problem.

## 2 Some direct applications

This section can be skipped without loss of continuity. In Section 2.1, we conclude from Theorem 1.1 that probability spaces with quasi-polynomially small <sup>3</sup> bias also fool all polynomial size DNF (or CNF) formulas. Using known explicit constructions of small probability spaces having the limited independence property or the small bias property, we directly obtain in

---

<sup>3</sup>By quasi-polynomially small we mean  $2^{-\log^{\Theta(1)}(n)}$ .

Section 2.2 a large class of explicit PRG's of  $O(\log^2 m \log n)$ -seed length for  $m$ -clause DNF (or CNF) formulas on  $n$  variables, improving previously known (unconditional) seed lengths. Finally, we highlight in Section 2.3 a direct application of Theorem 1.1 to the distribution of patterns in linear codes.

## 2.1 Extension to small bias probability spaces

We can conclude from Theorem 1.1 that probability spaces with quasi-polynomially small bias also fool all polynomial size DNF (or CNF) formulas.

Let  $\mu$  be a probability distribution on  $\{0, 1\}^n$ ,  $k \geq 0$  be an integer, and  $\delta > 0$ . We say that  $\mu$  is  $(\delta, k)$ -biased [NN93] if  $\mu$   $\delta$ -fools all parity functions on  $k$  or fewer of the  $n$  binary variables. This is a relaxation of the  $k$ -wise independent property since the latter is equivalent to the  $(0, k)$ -bias property. If  $\mu$  is  $(\delta, n)$ -biased, it is called  $\delta$ -biased [NN93].

We need the following relation:

**Theorem 2.1** [AGM02] *Any  $(\delta, k)$ -biased probability distribution  $\mu$  on  $\{0, 1\}^n$  is  $n^k \delta$ -close to a  $k$ -wise independent probability distribution  $\mu'$  on  $\{0, 1\}^n$  in the sense that  $|\mu(A) - \mu'(A)| \leq n^k \delta$ ,  $\forall A \subset \{0, 1\}^n$ . Thus if  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  is such that any  $k$ -wise independent probability distribution on  $\{0, 1\}^n$   $\epsilon$ -fools  $f$ , then any  $(\delta, k)$ -biased probability distribution on  $\{0, 1\}^n$   $(\epsilon + \delta n^k)$ -fools  $f$ .*

Using Theorem 1.1 and Theorem 2.1, we get:

**Corollary 2.2** *Any  $(\delta, k)$ -biased probability distribution on  $\{0, 1\}^n$   $(16m^{2.2}2^{-\sqrt{k}/10} + \delta n^k)$ -fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables.*

**Corollary 2.3** *There is a function  $\delta(m, n, \epsilon) = 2^{-\Theta(\log^2 \frac{m}{\epsilon} \log n)}$  such that for all positive integers  $m$  and  $n$ , and all  $0 < \epsilon < 1$ , any  $\delta(m, n, \epsilon)$ -biased probability distribution on  $\{0, 1\}^n$   $\epsilon$ -fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables.*

**Proof:** Set  $k = \left\lceil \left(10 \log \frac{32m^{2.2}}{\epsilon}\right)^2 \right\rceil$  in Corollary 2.2 so that  $16m^{2.2}2^{-\sqrt{k}/10} \leq \epsilon/2$  and  $\delta n^k \leq \epsilon/2$  if  $\delta \leq \epsilon 2^{-\left\lceil \left(10 \log \frac{32m^{2.2}}{\epsilon}\right)^2 \right\rceil \log n - 1} \stackrel{\text{def}}{=} \delta(n, n, \epsilon)$ . ■

Related previously known bounds in [AW85, LV96, Tre04] work for DNF formulas with very small fanins. Call a DNF formula an  $s$ -DNF if each clause contains at most  $s$  literal. Call a probability distribution on  $\{0, 1\}^n$   $k$ -wise  $\delta$ -dependent if it  $\delta$ -fools all AND gates on at most  $k$  literals. Ajtai and Wigderson [AW85], and Luby and Velickovic [LV96] show that for all integers  $1 \leq s \leq n$ , any  $k$ -wise  $\delta$ -dependent probability distribution  $\mu$  on  $\{0, 1\}^n$   $(e^{-k/(s2^s)} + 2^s \delta)$ -fools all boolean functions computable by  $s$ -DNF (or  $s$ -CNF) formulas on  $n$  variables. This implies that there is a function  $\delta(s, \epsilon) = 2^{-O(s2^s \log(1/\epsilon))}$  such that for all  $0 < \epsilon < 1$ , and all integers  $1 \leq s \leq n$ , any  $\delta(s, \epsilon)$ -biased probability distribution  $\mu$  on  $\{0, 1\}^n$   $\epsilon$ -fools all boolean functions computable by  $s$ -DNF (or  $s$ -CNF) formulas on  $n$  variables [Tre04]. The bound is good for  $s = O(1)$  and is nontrivial only if the condition  $k > s2^s$  is satisfied. For general DNF formulas, we can restrict our attention to the case when  $s = \log m + O(\log \frac{1}{\epsilon})$  (see Section 5.4), but that will not help here since then the condition  $k > s2^s$  would require  $k > m \log m$ .

## 2.2 A large class of PRG's for DNF formulas

The problem of derandomizing  $AC_0$  circuits (polynomial-size constant-depth unbounded-fanin AND/OR circuits) was first studied by Ajtai and Wigderson [AW85]. Using Hastad's parity lower bound [Has86], Nisan [Nis91] constructed a quasi-polynomial complexity<sup>4</sup> PRG for  $AC_0$  circuits of seed length  $O(\log^{2d+6} n)$ , where  $d$  is the circuit depth, and  $n$  is the input length. This initiated the hardness-versus-randomness approach which was developed in [NW88], [IW97], and others. Nisan's Generator was optimized by Luby, Velickovic, and Wigderson [LVW93] for depth-2 circuits reducing the seed length from  $O(\log^{10} n)$  to  $O(\log^4 mn)$ , where  $m$  is the number of clauses of the DNF (or CNF) formula. Using classical linear-code-based constructions of small probability spaces having the  $k$ -wise independent or the  $\delta$ -bias property, we directly obtain from Theorem 1.1 a large class of explicit PRG's for depth-2 circuits of seed length  $O(\log^3 mn)$ .

Small probability spaces having the  $k$ -wise independence property can be constructed from linear codes. The following construction is folklore. If  $C \subset \{0, 1\}^n$  is a binary linear code whose dual  $C^\perp \stackrel{\text{def}}{=} \{x \in \{0, 1\}^n : \sum_i x_i y_i = 0 \pmod{2} \text{ for all } y \in C\}$  has minimum distance greater than  $k$ , then the uniform distribution on the codewords of  $C$  is  $k$ -wise independent as a probability distribution on  $\{0, 1\}^n$ . Classical linear codes explicit constructions achieve  $|C| = n^{\Theta(k)}$ .

**Corollary 2.4** *For all positive integers  $m, n$ , and every  $\epsilon > 0$ , there is an integer  $t = O(\log^2 \frac{m}{\epsilon} \log n)$  and an explicit generator  $G : \{0, 1\}^t \rightarrow \{0, 1\}^n$ , constructible in  $\text{poly}(n)$ -time and computable in  $O(tn)$ -time, such that for every boolean function  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables, we have*

$$|Pr_{x \in \{0, 1\}^t} [g(G(x)) = 1] - Pr_{x \in \{0, 1\}^n} [g(x) = 1]| \leq \epsilon.$$

**Proof:** Without loss of generality, assume that  $\log^2 \frac{m}{\epsilon} \log n = o(n)$  (Otherwise, set  $t = n$  and let  $G$  be the identity map). Given  $n, m$ , and  $\epsilon$ , let

$$k \stackrel{\text{def}}{=} \left\lceil \left( 10 \log \frac{16m^{2.2}}{\epsilon} \right)^2 \right\rceil,$$

thus  $16m^{2.2} 2^{-\sqrt{k}/10} \leq \epsilon$ . Construct in  $\text{poly}(n, k)$ -time the parity check matrix  $H_{t \times n}$  of an explicit binary linear code  $D$  of block length  $n$ , message length  $n - t$ , and minimum distance at least  $k + 1$ , where  $t = O(k \log n) = O(\log^2 \frac{m}{\epsilon} \log n)$ . We can achieve  $t = O(k \log n)$  using, for instance, a Reed-Solomon code or an algebraic geometry code reduced to a binary code by plain binarization or by concatenation, and punctured if necessary. Thus the probability distribution on  $\{0, 1\}^n$  resulting from choosing a random codeword from the dual  $C = D^\perp$  of  $D$  is  $k$ -wise independent. This distribution is induced by the uniform distribution on  $\{0, 1\}^t$  via the  $\mathbb{F}_2$ -linear map  $G : \{0, 1\}^t \rightarrow \{0, 1\}^n$  defined by  $G(x) = xH$ . ■

---

<sup>4</sup>By quasi-polynomial complexity we mean  $2^{\log^{\Theta(1)}(n)}$ .

Probability distributions with the  $\delta$ -bias property can be explicitly constructed from linear codes with support size  $\left(\frac{n}{\delta}\right)^{\Theta(1)}$  [NN93, AGHP92]. Using those constructions and Corollary 2.3, we get a variation of Corollary 2.4 of asymptotically the same seed length, i.e.,  $t = O(\log^2 \frac{m}{\epsilon} \log n)$  (see also Corollary 11.2). Explicit constructions of  $(\delta, k)$ -biased probability distributions of support size  $\left(\frac{k \log n}{\delta}\right)^{\Theta(1)}$  [NN93, AGHP92] can be also used via Corollary 2.2 to achieve asymptotically the same seed length for  $k = \left\lceil \left(10 \log \frac{32m^{2.2}}{\epsilon}\right)^2 \right\rceil$  and  $\delta = \frac{\epsilon}{2n^k}$ .

For  $\epsilon = n^{-O(1)}$ , Corollary 2.4 and its variations give us PRG's of  $O(\log^2 m \log n)$ -seed length for  $m$ -clause DNF (or CNF) formulas.

Note that, when  $\epsilon = n^{-O(1)}$ , each of the above PRG's leads to a  $2^{O(\log^2 m \log n)}$ -time algorithm for the *DNF formula approximate counting problem* (given a DNF formula and  $\epsilon > 0$ , approximate the fraction of its satisfying assignments within  $\pm\epsilon$  additive error). We should mention here that this does not improve the best known time for the DNF formula approximate counting problem; the algorithm of Luby and Velickovic [LV96], which is not based on a PRG, solves this problem in  $(m \log n)^{\exp(O(\sqrt{\log \log n}))}$  time when  $\epsilon$  is constant.

The problems of constructing a logarithmic seed-length (unconditional) PRG for DNF formulas or finding a polynomial time algorithm for the DNF formulas approximate counting problem remain open.

## 2.3 Patterns in binary linear codes

If  $I \subset [n]$  and  $\alpha \in \{0, 1\}^I$ , we call the pair  $(I, \alpha)$  an *n-pattern*. We say that a string  $x \in \{0, 1\}^n$  *contains* an *n-pattern*  $p = (I, \alpha)$  if  $x_i = \alpha_i$  for all  $i \in I$ .

If  $C \subset \{0, 1\}^n$  is a binary linear code whose dual has minimum distance greater than  $k$ , then the uniform distribution on the codewords of  $C$  is  $k$ -wise independent as a probability distribution on  $\{0, 1\}^n$ . Specialized to such  $k$ -wise independent distributions, Theorem 1.1 can be rephrased as an estimate of the probability that a random codeword of  $C$  contains a pattern from a given set of patterns.

**Corollary 2.5** *Let  $A$  be a set consisting of  $m$   $n$ -patterns, and let  $C \subset \{0, 1\}^n$  be a linear code whose dual has minimum distance greater than  $k$ . Then*

$$\left| \Pr_{x \in C} \left[ \exists p \in A \text{ s.t. } x \text{ contains } p \right] - \Pr_{x \in \{0, 1\}^n} \left[ \exists p \in A \text{ s.t. } x \text{ contains } p \right] \right| \leq 16m^{2.2} 2^{-\sqrt{k}/10}.$$

**Proof:** Let  $\mu$  be the  $k$ -wise independent probability distribution resulting from choosing a random codeword of  $C$ , and let  $F$  be the DNF formula whose clauses correspond to the patterns in  $A$ , i.e.,

$$F = \bigvee_{(I, \alpha) \in A} \left( \bigwedge_{i \in I: \alpha_i = 1} x_i \wedge \bigwedge_{i \in I: \alpha_i = 0} \neg x_i \right).$$

Apply Theorem 1.1 on  $\mu$  and  $F$ . ■

For instance, consider the following concrete case.

**Corollary 2.6** *Let  $C \subset \{0, 1\}^n$  be a linear code whose dual has minimum distance greater than  $k$ , and let  $1 \leq t \leq n$  be an integer. Then*

$$\left| \Pr_{x \in C} \left[ \begin{array}{c} x \text{ contains } t \\ \text{consecutive ones} \end{array} \right] - \Pr_{x \in \{0,1\}^n} \left[ \begin{array}{c} x \text{ contains } t \\ \text{consecutive ones} \end{array} \right] \right| \leq 16(n-t+1)^{2.2} 2^{-\sqrt{k}/10}.$$

Note that if the maximum size  $s$  of a pattern in  $A$  in Corollary 2.5 is  $o(\sqrt{k})$ , the bound can be improved via Corollary 9.6.

### 3 Fourier transform preliminaries

The study of boolean function using harmonic analysis methods dates back to the late 60's. See for instance [Lec71],[KKL88], and [LMN93].

We assemble below some needed preliminaries: Fourier transform definition, Parseval's equality, the degree of a boolean function, and the Fourier truncation operator.

We identify the hypercube  $\{0, 1\}^n$  with the group  $\mathbb{Z}_2^n \stackrel{\text{def}}{=} (\mathbb{Z}/2\mathbb{Z})^n$ . The *characters* of the abelian group  $\mathbb{Z}_2^n$  are  $\{\mathcal{X}_y\}_{y \in \mathbb{Z}_2^n}$ , where  $\mathcal{X}_y(x) \stackrel{\text{def}}{=} (-1)^{\sum_{i=1}^n x_i y_i}$ . Those characters form an orthogonal basis of the space of real-valued functions defined on  $\mathbb{Z}_2^n$ . They are orthogonal with respect the uniform distribution, i.e.,  $E\mathcal{X}_y \mathcal{X}_{y'} = 0$  if  $y \neq y'$  and 1 otherwise. If  $g$  is a real-valued function on  $\mathbb{Z}_2^n$ , we denote by  $\hat{g}$  the *Fourier transform* of  $g$  with respect to the characters of the abelian group  $\mathbb{Z}_2^n$ . That is, if  $g : \{0, 1\}^n \rightarrow \mathbb{R}$ , its Fourier transform  $\hat{g} : \{0, 1\}^n \rightarrow \mathbb{R}$  is given by the coefficients of the expansion of  $g$  in terms of the  $\{\mathcal{X}_z\}_z$  basis:

$$g(x) = \sum_y \hat{g}(y) \mathcal{X}_y(x) \quad \text{and} \quad \hat{g}(y) = \frac{1}{2^n} \sum_x g(x) \mathcal{X}_y(x).$$

*Parseval's equality* relates the expected value of the square of a boolean function  $g : \{0, 1\}^n \rightarrow \mathbb{R}$  to the  $L_2$ -norm of its Fourier transform:

$$Eg^2 = \sum_y \hat{g}(y)^2 = \|\hat{g}\|_2^2.$$

The equality follows from the orthogonality of the characters  $\{\mathcal{X}_y\}_y$ .

If  $x \in \mathbb{Z}_2^n$ , the *weight* of  $x$ , which we denote by  $|x|$ , is the number of nonzero coordinates of  $x$ . The *degree* of  $g : \{0, 1\}^n \rightarrow \mathbb{R}$  is the smallest degree of a polynomial  $p \in \mathbb{R}[x_1, \dots, x_n]$  such that  $p(x) = g(x)$  for all  $x \in \{0, 1\}^n$ . Equivalently, in terms of the basis  $\{\mathcal{X}_y\}_y$ , the *degree* of  $g$  is equal to the maximal weight of  $y \in \{0, 1\}^n$  such that  $\hat{g}(y) \neq 0$ .

Finally, we define the Fourier truncation operator. Denote by  $L(\{0, 1\}^n)$  the space of real-valued functions on  $\{0, 1\}^n$ . If  $t \geq 0$  is an integer, define the *Fourier truncation operator*  $\text{Trn}_t : L(\{0, 1\}^n) \rightarrow L(\{0, 1\}^n)$  by

$$\text{Trn}_t g = \sum_{y:|y| \leq t} \hat{g}(y) \mathcal{X}_y.$$

The truncation operator  $\text{Trn}_t$  kills the high frequencies of  $g$  and produces a function  $\text{Trn}_t g$  of degree at most  $t$ . Equivalently,  $\text{Trn}_t g$  can be defined as the optimal solution  $f^*$  of the following  $L_2$ -approximation problem: given  $g$  and  $t$ , minimize  $E(g - f)^2$  over the choice of  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  of degree at most  $t$ . This follows from solving the underlying least-square problem in the orthogonal basis  $\{\mathcal{X}_y\}_y$ . Note that, by Parseval's equality, the smallest  $L_2$ -approximation error is

$$E(g - \text{Trn}_t g)^2 = \sum_{y:|y|>k} \widehat{g}(y)^2.$$

## 4 LP duality perspective

**Definition 4.1** *We say that a distribution property  $\epsilon$ -fools a function  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  if any probability distribution on  $\{0, 1\}^n$  with this property  $\epsilon$ -fools  $g$ .*

We note that linear programming duality gives a purely analytical characterization of the class of boolean functions that are fooled by the  $k$ -wise independence property. The characterization is in terms of  $L_1$ -approximability by sandwiching polynomials of degree at most  $k$ . In particular, we show that:

**Theorem 4.2** *Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$ ,  $k \geq 0$  an integer, and  $\epsilon \geq 0$ . Then the  $k$ -wise independence property  $\epsilon$ -fools  $g$  if and only if there exist  $g_l, g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  such that:*

- i) (low degree)  $\deg(g_l) \leq k$  and  $\deg(g_u) \leq k$*
- ii) (sandwiching polynomials)  $g_l \leq g \leq g_u$*
- iii) (small  $L_1$ -approximation error)  $E(g - g_l) \leq \epsilon$  and  $E(g_u - g) \leq \epsilon$ , where the expectation is over the uniform probability distribution.*

We show the LP duality calculations in Appendix A in the more general context of the  $(\delta, k)$ -bias property. Since the  $k$ -wise independence property is the  $(0, k)$ -bias property, Theorem 4.2 follows from Theorem A.1 in Appendix A by setting  $\delta = 0$

The proof of the main result of this paper in Theorem 1.1 depends only on the if part of Theorem 4.2. We give in this section a direct verification of the if part which does not involve LP duality calculations.

In terms of the  $\{\mathcal{X}_y\}_y$  basis, the definition of  $k$ -wise independence can be rephrased as follows. Let  $\mu$  be a probability distribution on  $\{0, 1\}^n$  and let  $k \geq 0$  be an integer. Then the following are equivalent:

- a)  $\mu$  is  $k$ -wise independent
- b)  $E_\mu \mathcal{X}_y = 0$  for each nonzero  $y$  in  $\{0, 1\}^n$  whose weight is less than or equal to  $k$
- c)  $E_\mu p = E p$  for each  $p : \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $\deg(p) \leq k$ , where the second expectation is with respect to the uniform probability distribution.

The equivalence between (a) and (b) is immediate. To relate to (c), write  $p$  as  $p = \sum_{y:|y|\leq k} \hat{p}(y)\mathcal{X}_y$ , thus  $E_\mu p = \hat{p}(0) + \sum_{y\neq 0:|y|\leq k} \hat{p}(y)E_\mu \mathcal{X}_y$ .

This equivalence is the key relation between  $k$ -wise independent probability distributions and polynomials of degree at most  $k$ . Using this relation, we establish below the if part of Theorem 4.2. Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$ ,  $k \geq 0$  an integer, and  $\epsilon \geq 0$ . Assume the existence of sandwiching polynomials  $g_l$  and  $g_u$  satisfying (i), (ii), and (iii). Let  $\mu$  be a  $k$ -wise independent probability measure. We want to show that  $\mu$   $\epsilon$ -fools  $g$ . Since  $g_u$  has degree at most  $k$ ,  $E_\mu g_u = E g_u$ . Hence

$$\begin{aligned} Pr_{x\sim\mu}[g(x) = 1] - Pr_{x\in\{0,1\}^n}[g(x) = 1] &= E_\mu g - E g \\ &= E_\mu(g - g_u) + E(g_u - g) \leq E(g_u - g) \leq \epsilon. \end{aligned}$$

where the first inequality follows from the fact that  $g_u \geq g$ . Similarly, using  $g_l$ , we get

$$\begin{aligned} -Pr_{x\sim\mu}[g(x) = 1] + Pr_{x\in\{0,1\}^n}[g(x) = 1] &= -E_\mu g + E g \\ &= E_\mu(g_l - g) + E(g - g_l) \leq E(g - g_l) \leq \epsilon. \end{aligned}$$

That is,  $\mu$   $\epsilon$ -fools  $g$ .

## 5 Outline of proof

We outline and overview in this section the proof of Theorem 1.1, restated below.

**Theorem 1.1** *Any  $k$ -wise independent probability distribution on  $\{0, 1\}^n$  ( $16m^{2.2}2^{-\sqrt{k}/10}$ )-fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables.*

The proof is based on harmonic and poset analysis techniques. It uses Hastad's switching Lemma [Has86] indirectly via Linial-Mansour-Nisan energy bound [LMN93], applied to many DNF formulas derived from original the DNF formula. The proof can be regarded as sequence of reductions between some  $L_1$  and  $L_2$  approximations of DNF formulas and auxiliary functions by low degree polynomials with real coefficients. After some simplifications in Section 5.1, we define those approximation notions in Section 5.2, then we outline the main steps in the proof in Section 5.3.

### 5.1 Simplifications and notation

Without loss of generality, we restrict our attention to DNF formulas since any CNF formula is the negation of a DNF formula with the same number of clauses, and a probability distribution  $\epsilon$ -fools a boolean function if and only if it  $\epsilon$ -fools its negation.

To avoid degenerate cases, we assume that the DNF formula has at least one clause and that each clause has at least one literal. We can do this without loss of generality since any

probability distribution 0-fools the identically one boolean function and the identically zero boolean function.

A final notational technicality is that if  $F$  is a DNF formula, we abuse notation and also denote by  $F$  the boolean function computed by  $F$ . That is, if  $A_1, \dots, A_m$  are the clauses of  $F$ , we will denote by  $F$  also the boolean function  $F : \{0, 1\}^n \rightarrow \{0, 1\}$  given by  $F(x) = \bigvee_{c=1}^m A_c(x)$ .

## 5.2 Approximation notions used in the proof

The proof uses the following three approximation notions of real-valued functions on the hypercube by low degree polynomials with real coefficients.

**Definition 5.1 (Sandwiched  $L_1$ -approximation: bias)** *If  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  is a boolean function and  $k \geq 0$  is an integer, define the  $k$ -bias of  $g$ , denoted by  $\text{bias}(g; k)$ , to be the minimum value of  $\epsilon$  such that there exist  $g_l, g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  each of degree at most  $k$  such that  $g_l \leq g \leq g_u$ ,  $E(g_u - g) \leq \epsilon$ , and  $E(g - g_l) \leq \epsilon$ . We call  $g_l$  and  $g_u$  sandwiching polynomials of  $g$ .*

*Equivalently, by LP-duality (Theorem 4.2),  $\text{bias}(g; k)$  is the minimum value of  $\epsilon$  such that any  $k$ -wise independent probability distribution  $\mu$  on  $\{0, 1\}^n$   $\epsilon$ -fools  $g$ .*

The term bias should not be confused with its other various meanings in the literature.

**Definition 5.2 ( $L_2$ -approximation: energy)** *If  $g : \{0, 1\}^n \rightarrow \mathbb{R}$  is a real-valued function and  $t \geq 0$  is an integer, define the  $t$ -energy of  $g$  to be  $\text{energy}(g; t) = \min_f E(g - f)^2$  over the choice of a polynomial  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  of degree at most  $t$ .*

*Equivalently (see Section 3),*

$$\text{energy}(g; t) = \sum_{y: |y| > t} \widehat{g}(y)^2.$$

*That is,  $\text{energy}(g; t)$  is the high energy content of  $g$  at frequencies above  $t$ , and hence the “energy” terminology.*

We are allowing  $g$  to take real values since eventually we will be working with nonboolean-valued functions derived from DNF formulas.

**Definition 5.3 (Constrained  $L_2$ -approximation: zero-energy)** *If  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  is a boolean function and  $t \geq 0$  is an integer, define the  $t$ -zero-energy of  $g$  to be  $\text{zeroEnergy}(g; t) = \min_f E(g - f)^2$  over the choice of a polynomial  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  of degree at most  $t$  satisfying the **zeros-constraint**:  $f = 0$  whenever  $g = 0$ , i.e.,  $f(x) = 0$  for each  $x \in \{0, 1\}^n$  such that  $g(x) = 0$ .*

The zero-energy has no natural interpretation in the Fourier domain. The terminology is motivated by the above energy terminology and the zeros-constraint.

### 5.3 Main steps in the proof

Given an  $m$ -clause DNF formula  $F$  and  $k \geq 0$ , we want to bound its sandwiched  $L_1$ -approximation error  $\text{bias}(F; k)$ . The bound claimed in Theorem 1.1 is  $\text{bias}(F; k) \leq 16m^{2.2}2^{-\sqrt{k}/10}$ .

To get a concrete sense of the parameters, keep in mind the typical case when  $k = \Theta(\log^2 m)$ , and note that, although the number  $n$  of variables of  $F$  does not appear in the above bound or the subsequent ones, we care about the typical case when  $m$  is polynomial in  $n$ .

The proof can be regarded as a sequence of reductions between the above approximation notions in the context of DNF formulas and auxiliary functions. At a high level, we reduce the DNF sandwiched  $L_1$ -approximation problem to the DNF  $L_2$ -approximation problem, to which we apply Linial-Mansour-Nisan (LMN) energy bound. The DNF constrained  $L_2$ -approximation problem serves as an intermediate problem in this reduction.

The LMN energy bound [LMN93] says that for each  $m$ -clause DNF formula  $F$  and each integer  $t \geq 0$ ,  $\text{energy}(F; t) \leq 2m2^{-\sqrt{t}/20}$ .

To get started, we restrict our attention to DNF formulas with at most  $s$  literals per clause, where  $s = \Theta(\sqrt{k})$ . We call such DNF formulas  $s$ -DNF formulas. We can do that without loss of generality by paying a small additive error as noted in Section 5.4 of this outline. Note that  $s = \Theta(\log m)$  in the typical case when  $k = \Theta(\log^2 m)$ .

The first step of the proof reduces the problem of estimating the  $k$ -bias of an  $s$ -DNF formula to that of estimating its  $t$ -zero-energy, where  $t = \lfloor (k - s)/2 \rfloor \approx k/2$ . The argument is short and it is in Section 5.5 of this outline. Hence  $s = \Theta(\sqrt{t})$ , and  $t = \Theta(\log^2 m)$  in the typical case when  $k = \Theta(\log^2 m)$ .

The second and more difficult step of the proof is estimating the zero-energy of an  $s$ -DNF formula, i.e., the  $s$ -DNF constrained  $L_2$ -approximation problem. We reduce this problem to the  $s$ -DNF  $L_2$ -approximation problem. The argument is long and it involves two intermediate reductions to  $L_2$ -approximation problems of auxiliary nonboolean-valued functions associated with DNF formulas. First, we reduce the problem of bounding the  $t$ -zero-energy of an  $s$ -DNF formula  $F$  to that of bounding the  $t'$ -energies of auxiliary real-valued functions associated with DNF formulas derived from  $F$ , where  $t' = t - s \approx t$ . We overview this in Sections 5.6, 5.7, and 5.8. Then, we reduce the problem of bounding the  $t'$ -energies of each of those auxiliary functions to that of bounding the  $t'$ -energies of additional derived DNF formulas, which finally enables us to use the LMN energy bound. We give an overview of this in Section 5.9. We conclude in Sections 5.10 and 5.11.

The arguments in the second step are based on harmonic and poset analysis machinery, which we develop in Section 6. In this outline section, we overview the underlying constructions and the end results without using the language of Section 6.

### 5.4 Ignoring large clauses

If  $s \geq 1$  is an integer, we call a DNF formula  $F$  an  $s$ -DNF if each clause of  $F$  contains at most  $s$  literals. By paying a small additive error, we can assume without loss of generality that the DNF formula does not contain very large clauses.

**Lemma 5.4** *Let  $k \geq s \geq 1$  be integers, and  $\epsilon \geq 0$ . If  $\text{bias}(F; k) \leq \epsilon$  for each  $m$ -clause  $s$ -DNF formula  $F$ , then  $\text{bias}(F; k) \leq \epsilon + m2^{-s}$  for each  $m$ -clause DNF formula  $F$ .*

At the end, we will set  $s = \Theta(\sqrt{k})$ . Hence  $s = \Theta(\log m)$  in the typical case when  $k = \Theta(\log^2 m)$ .

**Proof:** The argument is easy. It is more direct in this lemma to work with the primal definition of the bias. That is, we argue on probability distribution and not on sandwiching polynomials.

Assume that the hypothesis is correct. Let  $F$  be an  $m$ -clause DNF formula on  $n$  variables, and let  $\mu$  be a  $k$ -wise independent probability distribution on  $\{0, 1\}^n$ . We want to show that  $|Pr_\mu[F = 1] - Pr[F = 1]| \leq \epsilon + m2^{-s}$ .

Let  $A_1, \dots, A_m$  be the clauses of  $F$ , thus  $F = \bigvee_{c=1}^m A_c$ . Let  $C' \subset [m]$  be the set of indices of clauses each containing at most  $s$  literals,  $F' = \bigvee_{c \in C'} A_c$ ,  $C'' = [m] \setminus C'$ , and  $F'' = \bigvee_{c \in C''} A_c$ .

Since  $F'$  is an  $s$ -DNF formula, we have  $|Pr_\mu[F' = 1] - Pr[F' = 1]| \leq \epsilon$  because  $\text{bias}(F'; k) \leq \epsilon$  by the lemma hypothesis.

Since  $F = F' \vee F''$ , we have  $Pr_\mu[F' = 1] \leq Pr_\mu[F = 1] \leq Pr_\mu[F' = 1] + Pr_\mu[F'' = 1]$  and  $-Pr[F' = 1] - Pr[F'' = 1] \leq -Pr[F = 1] \leq -Pr[F' = 1]$ . Thus

$$-\epsilon - Pr[F'' = 1] \leq Pr_\mu[F = 1] - Pr_\mu[F' = 1] \leq \epsilon + Pr_\mu[F'' = 1].$$

The lemma then follows from the inequalities  $Pr[F'' = 1] \leq |C''|2^{-(s+1)} \leq m2^{-s}$  and  $Pr_\mu[F'' = 1] \leq |C''|2^{-s} \leq m2^{-s}$ . The first inequality is immediate since each clause of  $F''$  contains at least  $s + 1$  literals. To verify the second inequality, construct another DNF formula  $G$  from  $F''$  by arbitrarily removing literals from each clause of  $F''$  to make its size equal to  $s$ . By construction,  $G$  is satisfied by all the satisfying assignments of  $F$ , hence  $Pr_\mu[F'' = 1] \leq Pr_\mu[G = 1]$ . Since  $\mu$  is  $k$ -wise independent and  $k \geq s$ , each clause of  $G$  is satisfied with a probability exactly  $2^{-s}$  with respect to  $\mu$ . Thus  $Pr_\mu[G = 1] \leq |C''|2^{-s}$ . ■

## 5.5 From bias to zero-energy

In this section, we reduce the  $s$ -DNF sandwiched  $L_1$ -approximation problem to the  $s$ -DNF constrained  $L_2$ -approximation problem. In particular, we reduce the problem of estimating the  $k$ -bias of an  $s$ -DNF formula to that of estimating its  $t$ -zero-energy, where  $t = \lfloor \frac{k-s}{2} \rfloor$ .

To justify this move from  $L_1$  to  $L_2$ , we briefly mention in Appendix B natural  $L_1$ -approaches which fall short of bounding the  $k$ -bias of  $s$ -DNF formulas.

We show that:

**Lemma 5.5 (bias  $\approx$  zero-energy)** *Let  $F$  be an  $m$ -clause  $s$ -DNF formula and let  $k \geq s$  be an integer. Then  $\text{bias}(F; k) \leq m \times \text{zeroEnergy}(F; t)$ , where  $t = \lfloor \frac{k-s}{2} \rfloor$ .*

Note that  $t \approx k/2$  when  $s = \Theta(\sqrt{k})$ . Moreover,  $t = \Theta(\log^2 m)$  in the typical case when  $k = \Theta(\log^2 m)$ .

**Proof:** Assume that we have  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $\text{deg}(f) \leq \lfloor \frac{k-s}{2} \rfloor$ , and  $f$  satisfies the zeros-constraint:  $f(x) = 0$  for each  $x \in \{0, 1\}^n$  such that  $F(x) = 0$ .

The approach is to construct the sandwiching polynomials  $f_l$  and  $f_u$  of  $F$  as

$$f_l \stackrel{\text{def}}{=} 1 - (1 - f)^2, \text{ and}$$

$$f_u \stackrel{\text{def}}{=} 1 - \left(1 - \sum_{c=1}^m A_c\right)(1 - f)^2,$$

where  $A_1, \dots, A_m$  are the clauses of  $F$  realized as polynomials on the variables  $x_1, \dots, x_n$ . Since  $F$  is an  $s$ -DNF, the degree of each  $A_c$  is at most  $s$ . Hence, by construction,  $\deg(f_l), \deg(f_u) \leq k$ .

We want to show that:

- i)  $f_l \leq F \leq f_u$ , and
- ii)  $E(F - f_l) \leq mE(F - f)^2$  and  $E(f_u - F) \leq mE(F - f)^2$ .

To establish (i), let  $x \in \{0, 1\}^n$ , and consider two cases depending on whether  $F(x) = 0$  or 1.

If  $F(x) = 0$ , then  $f(x) = 0$  by the zeros-constraint on  $f$ . Moreover, none of the clauses of  $F$  are satisfied by  $x$ , i.e.,  $A_c(x) = 0$  for each clause  $A_c$  of  $F$ . Hence  $f_l(x) = 1 - (1 - 0)^2 = 0$  and  $f_u(x) = 1 - (1 - 0)(1 - 0)^2 = 0$ . That is, (i) holds with equality when  $F(x) = 0$ .

If  $F(x) = 1$ , there exists at least one clause  $c$  such that  $A_c(x) = 1$ , hence  $1 - \sum_c A_c(x) \leq 0$ . It follows that

$$f_u(x) = 1 - \left(1 - \sum_c A_c(x)\right)(1 - f(x))^2 \geq 1 = F(x).$$

Moreover,  $f_l(x) = 1 - (1 - f(x))^2 \leq 1 = F(x)$ , which verifies (i) when  $F(x) = 1$ .

To establish (ii), it is enough to argue that  $E(f_u - f_l) \leq mE(F - f)^2$  since (by (i))  $E(f_u - f_l)$  is an upper bound on both  $E(F - f_l)$  and  $E(f_u - F)$ .

We have

$$f_u(x) - f_l(x) = \sum_c A_c(x)(1 - f(x))^2 = \sum_c A_c(x)(F(x) - f(x))^2.$$

To verify the second equality, consider two cases depending on whether  $F(x) = 1$  or 0. The  $F(x) = 1$  case is immediate. If  $F(x) = 0$ , then  $A_c(x) = 0$  for each  $c$ , and hence the equality holds because both terms are zeros. It follows that

$$E(f_u - f_l) = E\left(\sum_{c=1}^m A_c\right)(F - f)^2 \leq mE(F - f)^2.$$

■

We derive in Section 10 a compact form of the optimal solution of the least square problem underlying the definition of the  $t$ -zero-energy of an  $s$ -DNF formula (we focus on the monotone case for simplicity). Unable to estimate the optimal solution, we leave the problem open, and we construct next a suboptimal solution.

## 5.6 Construction overview

Let  $F$  be an  $m$ -clause  $s$ -DNF formula and  $t \geq s$  be an integer. We want to upper bound the  $t$ -zero-energy of  $F$ , i.e., construct  $f : \{0, 1\}^n \rightarrow \mathbb{R}$  of degree at most  $t$  such that the mean square error  $E(F - f)^2$  is small, and  $f$  satisfies the zeros-constraint:  $f = 0$  whenever  $F = 0$ .

In this section we explain a construction of a function  $f$  satisfying the zeros-constraints. We do not analyze the corresponding mean square error, but we give some intuition why one would speculate that it is small. The mean square error analysis is fully presented in Section 7. In a first reading, the reader may skip this overview section without loss of formal continuity and move to the end results in Sections 5.7 and 5.8.

The zeros-constraint is behind the difficulty of the problem. It is worth mentioning that it excludes setting  $f$  to the truncation  $\text{Trn}_t F$  of  $F$  by the Fourier truncation operator  $\text{Trn}_t$  (defined in Section 3). This choice minimizes the mean square error, but it is not an option for us because  $\text{Trn}_t F$  typically violates the zeros-constraint as it is rarely equal to 0 (or 1). We will not truncate the formula  $F$ , but we will apply truncation to carefully chosen components arising from rewriting the formula using inclusion-exclusion as we explain next.

For simplicity, we assume in this overview section that the DNF formula  $F$  is monotone. A DNF formula is called *monotone* if none of its clauses contain a negated variable. Represent  $F$  by a bipartite graph  $F = (C, [n], N)$  between the set  $C = [m]$  of clauses and the set  $[n]$  of variables indices. For each clause  $c \in C$ ,  $N(c)$  is the neighborhood of  $c$  consisting of the indices of the variables in  $c$ . If  $S \subset C$  is a set of clauses, let  $N(S)$  be the neighborhood of  $S$ , i.e.,  $N(S) \stackrel{\text{def}}{=} \cup_{c \in S} N(c)$ . If  $z \subset [n]$  is a set of variable indices, denote the corresponding monotone AND gate by  $AND_z$ , i.e.,

$$AND_z(x) \stackrel{\text{def}}{=} \bigwedge_{i \in z} x_i = \prod_{i \in z} x_i,$$

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

$$F(x) = \bigvee_{c \in C} AND_{N(c)}(x).$$

To construct  $f$ , expand  $F$  as follows:

$$\begin{aligned} F(x) &= 1 - \prod_{c \in C} (1 - \prod_{i \in N(c)} x_i) \\ &= \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} \prod_{i \in N(S)} x_i \\ &= \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} AND_{N(S)}(x). \end{aligned} \tag{5.1}$$

A direct way to obtain from this summation a low degree function which satisfies the zeros-constraint is to throw away the terms where  $N(S)$  is larger than  $t$ . This reduces the degree to  $t$  and satisfies the zeros-constraint (since if  $F = 0$ , then  $AND_{N(c)} = 0$  for each  $c \in C$ , and hence  $AND_{N(S)} = 0$  for each  $S \neq \emptyset \subset C$ ). But this does not work since the resulting mean square error may grow exponentially as  $t$  grows (the simplest example is when the

DNF formula consists of a single OR gate on the  $n$  variables, i.e.,  $C = [n]$  and  $N(i) = \{i\}$  for each  $i \in C$ .

Instead of throwing away the large terms, we will modify them while guaranteeing that each is still zero on the zeros of  $F$ . Let  $S \subset C$  such that  $S \neq \emptyset$ , and consider the term corresponding to  $S$ . For each  $c \in S$ , we have

$$AND_{N(S)} = AND_{N(c)} AND_{N(S) \setminus N(c)}.$$

Averaging over all  $c \in S$ , we trivially get

$$AND_{N(S)} = E_{c \in S} AND_{N(c)} AND_{N(S) \setminus N(c)},$$

hence we can express  $F$  as

$$F = \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} E_{c \in S} AND_{N(c)} AND_{N(S) \setminus N(c)}.$$

Consider constructing  $f$  by truncating each  $AND_{N(S) \setminus N(c)}$  to a degree- $(t - |N(c)|)$  polynomial via the Fourier truncation operator  $\text{Trn}_{t-|N(c)|}$ . That is, define

$$f \stackrel{\text{def}}{=} \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} E_{c \in S} AND_{N(c)} \text{Trn}_{t-|N(c)|} AND_{N(S) \setminus N(c)}.$$

The degree of each  $\text{Trn}_{t-|N(c)|} AND_{N(S) \setminus N(c)}$  is at most  $t - |N(c)|$  and the degree of each  $AND_{N(c)}$  is  $|N(c)|$ . Thus, by construction,  $\deg(f) \leq t$ .

The key point is that: if  $F(x) = 0$ , then  $AND_{N(c)}(x) = 0$  for each clause  $c \in C$ , and hence  $f(x) = 0$ . That is,  $f$  satisfies the zeros-constraint.

To sum up, let  $F$  be a monotone  $s$ -DNF formula and  $t \geq s$  be an integer. Then we have the bound

$$\text{zeroEnergy}(F; t) \leq E(F - f)^2 = \|\widehat{F - f}\|_2^2, \quad (5.2)$$

where  $F - f$  is the construction error term given by

$$F - f = \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} E_{c \in S} AND_{N(c)} (1 - \text{Trn}_{t-|N(c)|}) AND_{N(S) \setminus N(c)}. \quad (5.3)$$

Needless to say, the signs  $(-1)^{|S|}$  in the above summation are critical. That is, using a triangular inequality to upper bound  $E(F - f)^2$  does not give a nontrivial bound since we have exponentially many terms whose values are not small enough to make the sum value less than 1.

Estimating the mean square error  $E(F - f)^2 = \|\widehat{F - f}\|_2^2$  as  $t$  grows is difficult. Before moving into that, we give below some intuition why one would speculate that the mean square error of this construction decays as  $t$  grows.

It can be shown that  $\widehat{f}(y) = \widehat{\text{Trn}_t F}(y)$  if  $|y| \leq t - s$  and  $\widehat{f}(y) = \widehat{\text{Trn}_t F}(y) = 0$  if  $|y| > t$  (see Section 7 for a verification). In the region  $t - s < |y| \leq t$ ,  $\widehat{f}(y)$  behaves oddly. Since  $\widehat{f}(y)$  and  $\widehat{\text{Trn}_t F}(y)$  are equal outside this region, and since we know from LMN energy bound (Theorem

5.8) that  $\text{energy}(F; t) = \|\widehat{F - \text{Trn}_t F}\|_2^2$  decays quickly as  $t$  grows, we can hope that  $\widehat{f}(y)$  is not too bad in the region  $t - s < |y| \leq t$  and hence speculate that  $E(F - f)^2 = \|\widehat{F - f}\|_2^2$  decays also in some way with  $t$ . This makes  $f$  a potential candidate to bound the  $t$ -zero-energy of  $F$ , but unfortunately this intuition does not help in the analysis as the frequencies in the region  $t - s < |y| \leq t$  are too many to handle separately by trivial bounds.

Using analytical means, we bound  $\|\widehat{F - f}\|_2^2$  in Section 7 in terms of the energies of auxiliary functions associated with DNF formulas derived from  $F$ , which reduces via (5.2) the problem of bounding the  $t$ -zero-energy of  $F$  to that of bounding the energies of those functions. After defining the auxiliary functions in Section 5.7 below, we state the reduction in Section 5.8 without going into the above construction of  $f$ .

## 5.7 Skin and cover auxiliary functions

The proof uses the following auxiliary functions associated with DNF formulas.

**Definition 5.6 (Skin and cover auxiliary functions)** *Let  $G$  be a DNF formula on  $n$  variables whose clauses are  $A_1, \dots, A_m$ .*

- **(Skin)** *If  $u \geq 0$ , define the  $u$ -skin of  $G$  to be the real-valued function  $\text{skin}_{G,u} : \{0, 1\}^n \rightarrow \mathbb{R}$  given by*

$$\text{skin}_{G,u}(x) \stackrel{\text{def}}{=} 1 - (1 - e^{-u})^{\sum_{c=1}^m A_c(x)}.$$

*Note that  $\sum_{c=1}^m A_c(x)$  is the the number of clauses of  $G$  satisfied by  $x$ . The function  $\text{skin}_{G,u}$  is extended to  $u = 0$  by continuity, i.e.,  $\text{skin}_{G,0} = G$ .*

- **(Cover)** *Define the cover of  $G$  to be the real-valued function  $\text{cover}_G : \{0, 1\}^n \rightarrow \mathbb{R}$  given by*

$$\text{cover}_G(x) \stackrel{\text{def}}{=} \int_0^\infty \text{skin}_{G,u}(x) e^{-u} du = 1 - \frac{1}{1 + \sum_{c=1}^m A_c(x)}.$$

To evaluate the integral, note that

$$\int_0^\infty (1 - (1 - e^{-u})^a) e^{-u} du = \int_0^\infty e^{-u} du - \int_0^1 (1 - e^{-u})^a d(1 - e^{-u}) = 1 - \frac{1}{1 + a}$$

for all  $a \neq -1$ .

**Remark 5.7** The function  $\text{skin}_{G,u}$  converges to  $G$  from above as  $u$  approaches 0, and hence the name  $u$ -skin of  $G$ . Moreover,  $\text{cover}_G \geq G$ , and hence the name cover of  $G$ . The fact that both functions are greater than  $G$  is not used in the proof. Note that we defined  $\text{cover}_G$  as an integral. The fact that this integral evaluates to  $1 - 1/(1 + \sum_{c=1}^m A_c)$  is not used either in the proof (it is needed, however, to justify the “cover” name). The proof is based on those functions in the Fourier domain. The origin of the skin and cover functions is the construction overviewed in Section 5.6 above. We show in Section 7 that the Fourier transform of the cover function naturally appears when analyzing the Fourier transform of the construction error term given in (5.3), and the skin function naturally appears when trying to recover the cover function from its Fourier transform.

## 5.8 From zero-energy to the energies of auxiliary functions

In this section, we state the end results of Section 7 in which we reduce the problem of bounding the  $t$ -zero-energy of an  $s$ -DNF formula  $F$  to that of bounding the  $(t - s)$ -energies of cover and skin auxiliary functions associated with DNF formulas derived from  $F$ .

First, we reduce the problem of bounding the  $t$ -zero-energy of an  $s$ -DNF formula  $F$  to that of bounding the  $(t - s)$ -energies of the cover functions of DNF formulas derived from  $F$  as follows.

For simplicity, we start by stating the reduction in the context of monotone DNF formulas.

**Theorem 7.3 (zero-energy  $\leq$  energy of cover)** *Let  $F$  be an  $s$ -DNF formula on the variables  $x_1, \dots, x_n$  and let  $t \geq s$  be an integer. Let  $A_1, \dots, A_m$  be the clauses of  $F$  and let  $C = [m]$  be the set of indices of the clauses of  $F$*

- a) **(Monotone case)** *Assume that  $F$  is monotone and  $m \geq 2$ . For each clause index  $c \in C$ , let  $F_c$  be the DNF formula on the variables  $x_1, \dots, x_n$  whose clauses are  $\{A_c \wedge A_d\}_{d \in C \setminus \{c\}}$ . That is,  $F_c$  is the formula resulting from removing from  $F$  the clause  $A_c$  and adding the variables of  $A_c$  to each of the remaining clauses. Then*

$$\text{zeroEnergy}(F; t) \leq m^2 \max_{c \in C} \text{energy}(\text{cover}_{F_c}; t - s).$$

- b) **(General case)** *We call two clauses  $A_c$  and  $A_d$  consistent if they have a common satisfying assignment. If  $c \in C$ , let  $C_c$  be the set of indices of clauses other than  $A_c$  which are consistent with  $A_c$ , i.e.,  $C_c = \{d \in C \setminus \{c\} : A_c \text{ and } A_d \text{ are consistent}\}$ . Let  $C_{\text{main}}$  be the set of indices of the clauses of  $F$  which are consistent with at least one clause of  $F$  other than themselves, i.e.,  $C_{\text{main}} = \{c \in C : C_c \neq \emptyset\}$ .*

*For each clause index  $c \in C_{\text{main}}$ , let  $F_c$  be the DNF formula on the variables  $x_1, \dots, x_n$  whose clauses are  $\{A_c \wedge A_d\}_{d \in C_c}$ . That is,  $F_c$  is the formula resulting from removing from  $F$  the clause  $A_c$  and all the clauses not consistent with  $A_c$ , and adding the literals of  $A_c$  to each of the remaining clauses. Then*

$$\text{zeroEnergy}(F; t) \leq |C_{\text{main}}|^2 \max_{c \in C_{\text{main}}} \text{energy}(\text{cover}_{F_c}; t - s).$$

Thus if  $F$  is monotone, then  $C_c = C \setminus \{c\}$  for each  $c \in C$ , and  $C_{\text{main}} = C$ .

The proof of Theorem 7.3 is in Section 7 and it uses the machinery developed in Section 6. The construction underlying the reduction is overviewed in Section 5.6.

Note that each  $F_c$  is a  $2s$ -DNF formula on  $n$  variables with at most  $m - 1$  clauses and at least one clause (by the definition of  $C_{\text{main}}$ ). That is, the complexity of each  $F_c$  is in the worst case comparable to that of  $F$ . Recall from Section 5.5 that we care about the case when  $s = \Theta(\sqrt{t})$  (since  $t = \lfloor \frac{k-s}{2} \rfloor$  and  $s = \Theta(\sqrt{k})$ ), hence  $t - s \approx t$ . Moreover, typically  $t = \Theta(\log^2 m)$  (in the typical case when  $k = \Theta(\log^2 m)$ ).

Therefore, in general, we can now focus on estimating the  $t$ -energy of the cover of a DNF formula  $G$ , where  $t \geq 0$  is an integer. Unable to argue directly on the cover function, we

move to the  $u$ -skin function. The fact that  $\text{cover}_G = \int_0^\infty \text{skin}_{G,u} e^{-u} du$  immediately reduces the problem of estimating the  $t$ -energy of the cover function of a DNF to that of estimating the  $t$ -energies of its  $u$ -skin functions, for all  $u \geq 0$ . In particular, we have the following bound, which is verified via Cauchy-Schwarz inequality in Section 7.4.

**Lemma 7.4 (energy of cover  $\preceq$  energy of skin)** *Let  $G$  be a DNF formula and let  $t \geq 0$  be an integer. Then*

$$\text{energy}(\text{cover}_G; t) \leq \sup_{u \geq 0} \text{energy}(\text{skin}_{G,u}; t)$$

## 5.9 Back to DNF formulas

Let  $G$  be a DNF formula,  $u \geq 0$ , and  $t \geq 0$  be an integer. We want to estimate the  $t$ -energy of the  $u$ -skin of  $G$ . We bound in Section 8 the  $t$ -energy of the  $u$ -skin of  $G$  by the  $t$ -energies of DNF formulas derived from  $G$  by adding new auxiliary variables, which enables us to use LMN energy bound.

First we state the reduction in the special case when there exists a nonnegative integer  $v$  such that  $e^{-u} = 2^{-v}$ . Construct from  $G$  a new DNF formula  $G_v$  by adding  $v$  auxiliary new nonnegated variables to each clause of  $G$ . Thus, the total number of variables of  $G_v$  is  $n + mv$ . We show in Section 8 that  $\text{energy}(\text{skin}_{G,u}; t) \leq \text{energy}(G_v; t)$ . The proof is based on examining the Fourier transforms of  $\text{skin}_{G,u}$  and uses the machinery developed in Section 6.

In general, if  $v$  is not necessarily an integer, we show in Section 8 that:

**Theorem 8.1 (energy of skin  $\preceq$  energy)** *Let  $G$  be a DNF formula whose clauses are  $A_1, \dots, A_m$ , and let  $t \geq 0$  be an integer. If  $d \in \mathbb{N}^m$ , construct from  $G$  a new DNF formula  $G_d$  by adding  $d_c$  auxiliary new nonnegated variables to each clause  $A_c$ . That is, the clauses of  $G_d$  are  $\{A_c \wedge \bigwedge_{i=1}^{d_c} \tilde{x}_{ci}\}_{c=1}^m$ , where  $\{\tilde{x}_{ci}\}_{i=1}^{d_c}$  are the auxiliary new variables added to clause  $A_c$ .*

*Let  $u \geq 0$  and let  $v \geq 0$  such that  $e^{-u} = 2^{-v}$ , i.e.,  $v = u / \ln 2$ . Then*

$$\text{energy}(\text{skin}_{G,u}; t) \leq \max_{d \in \{\lfloor v \rfloor, \lceil v \rceil\}^m} \text{energy}(G_d; t).$$

Note that for each  $d \in \{\lfloor v \rfloor, \lceil v \rceil\}^m$ , the formula  $G_d$  is an  $m$ -clause DNF on  $n + \sum_c d_c$  variables.

This enables us to use the bound derived from Hastad's Switching Lemma in [LMN93] on the  $t$ -energy of a DNF formula:

**Theorem 5.8 [LMN93] (LMN energy bound)** *Let  $G$  be an  $m$ -clause DNF formula and  $t \geq 0$  be an integer, then  $\text{energy}(G; t) \leq 2m2^{-\sqrt{t}/20}$ .*

The proof of (an asymptotic version of) Theorem 1.1 follows by first substituting the bound of Theorem 5.8 in Theorem 8.1 and backtracking the bounds till Lemma 5.4. We summarize in Section 5.10; then we backtrack the bounds in Section 5.11.

## 5.10 Summary

The tables below summarize the main definitions and reductions.

Notion	Terminology	Definition
$k$ -bias of a function $g : \{0, 1\}^n \rightarrow \{0, 1\}$	$\text{bias}(g; k)$	Definition 5.1
Fourier truncation operator	$\text{Trn}_t$	Section 3
$t$ -energy of a function $g : \{0, 1\}^n \rightarrow \mathbb{R}$	$\text{energy}(g; t)$	Definition 5.2
$t$ -zero-energy of a function $g : \{0, 1\}^n \rightarrow \{0, 1\}$	$\text{zeroEnergy}(g; t)$	Definition 5.3
Boolean function of a DNF formula $G$	$G$ by notational abuse	Section 5.1
$u$ -skin function of a DNF formula $G$	$\text{skin}_{G,u}$	Definition 5.6
Cover function of a DNF formula $G$	$\text{cover}_G$	Definition 5.6

Reduction	Statement	Overview	Full presentation
bias $\preceq$ zero-energy	Lemma 5.5		Section 5.5
zero-energy $\preceq$ energy of cover	Theorem 7.3	Sections 5.6, 5.8	Section 7
energy of cover $\preceq$ energy of skin	Lemma 7.4	Section 5.8	Section 7
energy of skin $\preceq$ energy	Theorem 8.1	Section 5.9	Section 8

The full presentation in Sections 7 and 8 use the machinery in Section 6. In Sections 6, 7, and 8, we separate between the monotone case and the general case to introduce the arguments in a simple context. We recommend that the reader traverses first the monotone part of the full sections in the following order: Sections 6.1, 6.2, 6.3, 7.1, 7.2, 7.4, introduction of Section 8.

For future reference in Sections 5.11 and 9, we list below compact statements of the above reductions.

- **Lemma 5.4 (Focus on  $s$ -DNF):** Let  $k \geq s \geq 1$  be integers, and  $\epsilon \geq 0$ . If  $\text{bias}(F; k) \leq \epsilon$  for each  $m$ -clause  $s$ -DNF formula  $F$ , then  $\text{bias}(F; k) \leq \epsilon + m2^{-s}$  for each  $m$ -clause DNF formula  $F$ .

At the end, we will set  $s = \Theta(\sqrt{k})$ . Thus  $s = \Theta(\log m)$  in the typical case when  $k = \Theta(\log^2 m)$ .

- **Lemma 5.5 (bias  $\preceq$  zero-energy):** Let  $F$  be an  $m$ -clause  $s$ -DNF formula and let  $k \geq s$  be an integer. Then  $\text{bias}(F; k) \leq m \times \text{zeroEnergy}(F; t)$ , where  $t = \lfloor \frac{k-s}{2} \rfloor$ .

Note that  $t \approx k/2$  for  $s = \Theta(\sqrt{k})$ . Moreover,  $t = \Theta(\log^2 m)$  in the typical case when  $k = \Theta(\log^2 m)$ .

- **Theorem 7.3 (zero-energy  $\preceq$  energy of cover):** Let  $F$  be an  $s$ -DNF formula on the variables  $x_1, \dots, x_n$  and let  $t \geq s$  be an integer. Let  $A_1, \dots, A_m$  be the clauses of  $F$ . Let  $C = [m]$  be the set of indices of the clauses of  $F$ . If  $c \in C$ , let  $C_c = \{d \in C \setminus \{c\} : A_c \text{ and } A_d \text{ are consistent}\}$ . Let  $C_{\text{main}} = \{c \in C : C_c \neq \emptyset\}$ . For each  $c \in C_{\text{main}}$ , let  $F_c$  be the DNF formula on the variables  $x_1, \dots, x_n$  whose clauses are  $\{A_c \wedge A_d\}_{d \in C_c}$ . Then

$$\text{zeroEnergy}(F; t) \leq |C_{\text{main}}|^2 \max_{c \in C_{\text{main}}} \text{energy}(\text{cover}_{F_c}; t - s).$$

Note that for each  $c \in C_{main}$ ,  $F_c$  is a  $2s$ -DNF formula with at most  $m - 1$  clauses and least one clause. Moreover,  $|C_{main}| \leq |C| = m$ . Note also that  $t - s \approx t \approx k/2$  for  $t = \lfloor \frac{k-s}{2} \rfloor$  and  $s = \Theta(\sqrt{k})$ .

- **Lemma 7.4 (energy of cover  $\preceq$  energy of skin):** Let  $G$  be a DNF formula and let  $t \geq 0$  be an integer. Then

$$\text{energy}(\text{cover}_G; t) \leq \sup_{u \geq 0} \text{energy}(\text{skin}_{G,u}; t).$$

- **Theorem 8.1 (energy of skin  $\preceq$  energy):** Let  $G$  be a DNF formula whose clauses are  $A_1, \dots, A_m$ , and let  $t \geq 0$  be an integer. If  $d \in \mathbb{N}^m$ , construct from  $G$  a new DNF formula  $G_d$  by adding  $d_c$  auxiliary new nonnegated variables to each clause  $A_c$ . Let  $u \geq 0$  and  $v = u/\ln 2$ . Then

$$\text{energy}(\text{skin}_{G,u}; t) \leq \max_{d \in \{\lfloor v \rfloor, \lceil v \rceil\}^m} \text{energy}(G_d; t).$$

Note that for each  $d$ , the number of clauses of  $G_d$  is equal to that of  $G$ .

- **Theorem 5.8 (LMN energy bound):** Let  $G$  be an  $m$ -clause DNF formula and  $t \geq 0$  be an integer, then  $\text{energy}(G; t) \leq 2m2^{-\sqrt{t}/20}$ .

## 5.11 Backtracking

In this section, we derive an asymptotic version of Theorem 1.1 by substituting the bound of Theorem 5.8 in Theorem 8.1 and backtracking the bounds via Lemma 7.4, Theorem 7.3, Lemma 5.5, till Lemma 5.4. Namely, we show that if  $F$  is an  $m$ -clause DNF formula and  $k \geq 1$  is an integer, then  $\text{bias}(F; k) = O(m^{\Theta(1)}2^{-\Theta(\sqrt{k})})$ . Since the bound of Theorem 5.8 does not depend on the maximum number of literals in a clause, the needed calculations are minimal.

Let  $G$  be an  $m$ -clause DNF formula and let  $t \geq 0$ . Substituting the bound of Theorem 5.8 in Theorem 8.1, we get that  $\text{energy}(\text{skin}_{G,u}; t) \leq 2m2^{-\sqrt{t}/20}$ , for all  $u \geq 0$ . Substituting in Lemma 7.4, we obtain  $\text{energy}(\text{cover}_G; t) \leq 2m2^{-\sqrt{t}/20}$ . Thus, by Theorem 7.3, if  $F$  is an  $m$ -clause  $s$ -DNF formula and  $t \geq s$  is an integer, then  $\text{zeroEnergy}(F; t) \leq 2m^2(m - 1)2^{-\sqrt{t-s}/20}$ . It follows from Lemma 5.5 that if  $k \geq s$  is an integer, then  $\text{bias}(F; k) \leq 2m^3(m - 1)2^{-\sqrt{\lfloor (k-s)/2 \rfloor - s}/20}$ . Finally substituting in Lemma 5.4, we get that if  $F$  is an  $m$ -clause DNF formula and  $k \geq 1$  is an integer, then

$$\text{bias}(F; k) \leq m2^{-s} + 2m^3(m - 1)2^{-\sqrt{\lfloor (k-s)/2 \rfloor - s}/20},$$

for all integers  $s$  such that  $k \geq s \geq 1$ . Optimizing on  $s$ , we obtain  $\text{bias}(F; k) = O(m^4 2^{-\Theta(\sqrt{k})})$  for  $s = \Theta(\sqrt{k})$ .

The exact bound  $m^{2.2}2^{-\sqrt{k}/10}$  of Theorem 1.1 is derived in Section 9. It uses another form of the LMN energy bound which is tighter than Theorem 5.8 for  $s$ -DNF formulas when  $s$  is not relatively large (Theorem 9.1 in Section 9), and a sharper form of Lemma 7.4 (Part (a) of Lemma 7.4 in Section 7).

## 6 Möbius and Fourier analysis of DNF formulas and auxiliary functions

We develop in this section the proof machinery used in Sections 7 and 8.

We develop the monotone machinery in Sections 6.2 and 6.3. Monotone DNF formulas and their skin and cover auxiliary functions have natural expansions as linear combinations of monotone AND gates. We are interested in the coefficients of those expansions. The Fourier transforms of those functions can be extracted from those coefficients by a basis change. When expanding a real valued function  $f$  defined on the hypercube as a linear combinations of monotone AND gates, it is convenient to view the hypercube as the poset  $B_n$  of subsets of  $[n]$  ordered by inclusion. We note in Section 6.2 that this enables us to interpret the coefficients of the expansion of  $f$  as the Möbius transform of  $f$  with respect to the poset  $B_n$ . We also derive a simple change of basis formula to extract the Fourier transform of a function from its Möbius transform. In Section 6.3, we compute the Möbius and Fourier transforms of monotone DNF formulas and their auxiliary functions. The monotone machinery is used in Sections 7.1 and 8.

The poset language is essential to generalize to nonnecessarily monotone DNF formulas. The monotone machinery naturally generalizes to the nonnecessarily monotone case by essentially replacing the poset  $B_n$  with another poset  $B_n^{(2)}$ , defined in Section 6.4. We develop the general machinery in Sections 6.4 and 6.5. It is used in Section 7.3.

### 6.1 Posets preliminaries

For a general reference on posets, see [Sta97]. We only need few elementary definitions. A *poset* (partially ordered set)  $X$  is a set  $X$  with a reflexive, antisymmetric, and transitive binary relation  $\leq_X$ . We denote  $\leq_X$  by  $\leq$  when there is no confusion. We implicitly assume that  $X$  is finite. We denote the set of real-valued functions on  $X$  by  $L(X) = \{f : X \rightarrow \mathbb{R}\}$ . The *zeta function*  $\zeta_X$  of  $X$  is the linear transformation  $\zeta_X : L(X) \rightarrow L(X)$  given by

$$(\zeta_X f)(x) = \sum_{y \leq x} f(y) = \sum_{y \in X} \zeta_X(y, x) f(y).$$

That is, the matrix coefficients  $(\zeta_X(y, x))_{x,y}$  of  $\zeta_X$  are given by:

$$\zeta_X(y, x) = \begin{cases} 1 & \text{if } y \leq x \\ 0 & \text{otherwise.} \end{cases} \quad (6.1)$$

The zeta function  $\zeta_X$  is always nonsingular. The inverse  $\zeta_X^{-1}$  of  $\zeta_X$  is called the *Möbius function* of  $X$  and is denoted by  $\mu_X = \zeta_X^{-1}$ .

We are interested in two posets defined in Sections 6.2 and 6.4.

### 6.2 Poset $B_n$

Let  $[n] \stackrel{\text{def}}{=} \{1, \dots, n\}$ . Let  $B_n$  be the poset of subsets of  $[n]$  ordered by inclusion. We denote the subset inclusion  $a \subset b$  by  $a \leq b$  and  $b \geq a$ .

We identify the hypercube  $\{0, 1\}^n$  with the poset  $B_n$  by associating  $x \in \{0, 1\}^n$  with  $\text{support}(x) \stackrel{\text{def}}{=} \{i \in [n] : x_i = 1\} \in B_n$ . Thus  $x \in B_n$  means both a subset of  $[n]$  and a vector in  $\{0, 1\}^n$  depending on the context.

If  $z \in B_n$ , define the monotone AND function  $AND_z : B_n \rightarrow \{0, 1\}$  by

$$AND_z(x) \stackrel{\text{def}}{=} \bigwedge_{i \in z} x_i.$$

The functions  $\{AND_z\}_{z \in B_n}$  form a basis of  $L(B_n)$ . When working with the  $\{AND_z\}_{z \in B_n}$  basis, it is convenient to view the hypercube as the poset  $B_n$  since

$$AND_z(x) = \zeta_{B_n}(z, x) \tag{6.2}$$

by (6.1). That is, the functions  $\{AND_z\}_{z \in B_n}$  are the rows of the matrix of the zeta function  $\zeta_{B_n}$  of the poset  $B_n$ . Any function  $f \in L(B_n)$  can be expressed as

$$f = \sum_{z \in B_n} \tilde{f}(z) AND_z$$

for some  $\tilde{f} \in L(B_n)$ . By (6.2),

$$f(x) = \sum_{z \in B_n} \tilde{f}(z) \zeta_{B_n}(z, x) = (\zeta_{B_n} \tilde{f})(x).$$

That is,  $f = \zeta_{B_n} \tilde{f}$  and hence  $\tilde{f} = \mu_{B_n} f$ . We call  $\tilde{f} = \mu_{B_n} f$  the *Möbius transform* of  $f$ .

Our interest in the Möbius transform on  $B_n$  is motivated by the fact that monotone DNF formulas and auxiliary functions have natural expansions in the  $\{AND_z\}_z$  basis from which we can extract their Möbius transforms. We show that in Section 6.3.

Given the Möbius transform of a function, its Fourier transform can be extracted via a simple weighted summation, which we derive next. Recall the  $(\mathbb{Z}/2\mathbb{Z})^n$  group structure on  $\{0, 1\}^n$  from Section 3. In terms of the identification of  $\{0, 1\}^n$  with  $B_n$ ,  $B_n$  is an abelian group under the set exclusive union operation, which we denote by  $\oplus$ . In the  $B_n$ -terminology, the characters of this abelian group defined in Section 3 are  $\{\mathcal{X}_y(x) = (-1)^{|x \cap y|}\}_{y \in B_n}$ .

To extract the Fourier transform of a function  $f \in L(B_n)$  from its Möbius transform, we need a change of basis formula between the  $\{AND_z\}_z$  basis and the  $\{\mathcal{X}_y\}_y$  basis of  $L(B_n)$ . We have

$$AND_z(x) = \bigwedge_{i \in z} x_i = \prod_{i \in z} x_i = \prod_{i \in z} \frac{1 - (-1)^{x_i}}{2} = \frac{1}{2^{|z|}} \sum_{y \leq z} (-1)^{|y|} (-1)^{\sum_{i \in y} x_i}.$$

That is,

$$AND_z(x) = \frac{1}{2^{|z|}} \sum_{y \leq z} (-1)^{|y|} \mathcal{X}_y(x). \tag{6.3}$$

Note that if  $z = \emptyset$ , by convention  $\prod_{i \in z} x_i = \bigwedge_{i \in z} x_i = 1$ . Thus

$$\begin{aligned} f(x) &= \sum_z \tilde{f}(z) \text{AND}_z(x) \\ &= \sum_z \tilde{f}(z) 2^{-|z|} \sum_{y \leq z} (-1)^{|y|} \mathcal{X}_y(x) \\ &= \sum_y \mathcal{X}_y(x) (-1)^{|y|} \sum_{z \geq y} 2^{-|z|} \tilde{f}(z). \end{aligned}$$

Therefore, the desired change of basis formula is:

$$\hat{f}(y) = (-1)^{|y|} \sum_{z \geq y} 2^{-|z|} \tilde{f}(z) \quad \text{for all } y \in B_n \text{ and } f \in L(B_n). \quad (6.4)$$

Conversely, one can verify that

$$\tilde{f}(z) = (-1)^{|z|} 2^{|z|} \sum_{y \geq z} \hat{f}(y),$$

but we will not use that.

Finally, if  $f : \{0, 1\}^n \rightarrow \mathbb{R}$ , the degree of  $f$  is the smallest degree of a polynomial  $p \in \mathbb{R}[x_1, \dots, x_n]$  such that  $p(x) = f(x)$  for all  $x \in \{0, 1\}^n$ . In terms of the  $\{\mathcal{X}_y\}_y$  basis, the degree of  $f$  is equal to the maximal cardinality of  $y \in B_n$  such that  $\hat{f}(y) \neq 0$ . In terms of the  $\{\text{AND}_z\}_z$  basis, the degree of  $f$  is equal to the maximal cardinality of  $z \in B_n$  such that  $\tilde{f}(z) \neq 0$ .

### 6.3 Monotone DNF formulas and auxiliary functions

In this section we compute the Möbius and Fourier transforms of monotone DNF formulas and auxiliary functions.

A DNF formula is called *monotone* if none of its clauses contain a negated variable. We represent a monotone DNF formula  $F$  on  $n$  variables by a bipartite graph  $F = (C, [n], N)$  between the set  $C$  of clauses and the set  $[n]$  of variables indices. For each clause  $c \in C$ ,  $N(c)$  is the neighborhood of  $c$  consisting of the indices of the variables in  $c$ . To avoid degenerate cases, we assume that we have at least one clause, i.e.,  $|C| \geq 1$ , and that each clause contains at least one variable, i.e.,  $N(c) \neq \emptyset$  for each  $c \in C$ . If  $S \subset C$ ,  $N(S)$  denotes the neighborhood of  $S$ , i.e.,  $N(S) = \cup_{c \in S} N(c)$ . Finally, if  $s \geq 1$  is an integer, we call a monotone DNF formula  $F = (C, [n], N)$  an *s-DNF* formula if each clause contains at most  $s$  variables, i.e.,  $|N(c)| \leq s$  for each  $c \in C$ .

We chose the bipartite graph representation to allow for duplicate clauses; formulas possibly containing duplicate clauses will be derived in the proof of Theorem 1.1 (see Section 7.2.C).

If  $F = (C, [n], N)$  is a monotone DNF formula, we abuse notation and denote by  $F$  also the boolean function computed by  $F$ . That is, the boolean function  $F : B_n \rightarrow \{0, 1\}$  is given by

$$F(x) \stackrel{\text{def}}{=} \bigvee_{c \in C} \text{AND}_{N(c)}(x) \quad \text{for } x \in B_n.$$

Monotone DNF formulas and auxiliary functions have natural expansions in the  $\{AND_z\}_z$  basis from which we can extract their Möbius transforms. To get the Fourier transforms, we use the change of basis formula (6.4). To warm up, let us compute the Möbius and Fourier transform of  $F$ .

**Lemma 6.1** *Let  $F = (C, [n], N)$  be a monotone DNF formula. Then for all  $z, y \in B_n$ ,*

$$\widetilde{F}(z) = \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \quad (6.5)$$

$$\widehat{F}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|}. \quad (6.6)$$

**Proof:** By expanding  $F(x)$  as in (5.1) and grouping terms, we get

$$F(x) = \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} AND_{N(S)}(x) = \sum_{z \in B_n} AND_z(x) \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|}.$$

which verifies (6.5). The correctness of (6.6) follows from (6.5) via (6.4):

$$\widehat{F}(y) = (-1)^{|y|} \sum_{z \geq y} 2^{-|z|} \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|}. \quad \blacksquare$$

If  $F = (C, [n], N)$  is a monotone DNF formula and  $u \geq 0$ , recall from Definition 5.6 the auxiliary functions  $u$ -skin of  $F$  and cover of  $F$ :  $\text{skin}_{F,u}, \text{cover}_F \in L(B_n)$  are given by

$$\begin{aligned} \text{skin}_{F,u}(x) &\stackrel{\text{def}}{=} 1 - (1 - e^{-u}) \sum_{c \in C} AND_{N(c)}(x) \\ \text{cover}_F(x) &\stackrel{\text{def}}{=} \int_0^\infty \text{skin}_{F,u}(x) e^{-u} du, \end{aligned}$$

where  $\text{skin}_{F,u}$  is extended to  $u = 0$  by continuity, i.e.,  $\text{skin}_{F,0} = F$ .

We compute below their Möbius and Fourier transforms which play a critical role in the proof of Theorem 1.1 as shown in Sections 7 and 8.

**Lemma 6.2** *Let  $F = (C, [n], N)$  be a monotone DNF formula and  $u \geq 0$ . Then for all  $z, y \in B_n$ ,*

$$\widetilde{\text{skin}_{F,u}}(z) = \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} e^{-u|S|} \quad (6.7)$$

$$\widehat{\text{cover}_F}(z) = \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \frac{1}{|S| + 1} \quad (6.8)$$

$$\widehat{\text{skin}_{F,u}}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|} e^{-u|S|} \quad (6.9)$$

$$\widehat{\text{cover}_F}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S| + 1}. \quad (6.10)$$

**Remark 6.3** 1. Our interest in the cover and skin function originates from the last summation in (6.10). The analysis of the construction error term overviewed in Section 5.6 and fully presented in Section 7 leads to summations like the right side of (6.10) (see the proof of Lemma 7.1 and Section 7.2.A). To interpret this summation, we expressed  $\frac{1}{|S|+1}$  as  $\frac{1}{|S|+1} = \int_0^\infty e^{-u|S|} e^{-u} du$ , which lead us to the right side of (6.9). Using the Fourier-Möbius change of basis Formula (6.4), we obtained (6.7) which we identified as the Möbius transform of the skin function as shown in the proof below.

2. Comparing Lemmas 6.1 and 6.2, we see that the Möbius and Fourier transforms of the  $u$ -skin and cover of  $F$  are smoothed or weighted versions of those of  $F$ . The smoothing or weighting factor of the  $u$ -skin function is  $e^{-u|S|}$  and that of the cover function is  $\frac{1}{|S|+1}$ .

**Proof:** We start with (6.7). It is enough to verify it under the assumption that  $u > 0$ . The  $u = 0$  case follows from (6.5). We have

$$1 - (1 - e^{-u}) \sum_{c \in C} AND_{N(c)}(x) = 1 - \prod_{c \in C} (1 - e^{-u})^{AND_{N(c)}(x)}.$$

Since  $u > 0$ , we have  $(1 - e^{-u})^{AND_{N(c)}(x)} = 1 - e^{-u} AND_{N(c)}(x)$  for all  $c \in C$  and all  $x \in \{0, 1\}^n$  (if  $AND_{N(c)}(x) = 0$ , both terms are 1; if  $AND_{N(c)}(x) = 1$ , both terms are  $1 - e^{-u}$ ). Thus

$$\begin{aligned} 1 - (1 - e^{-u}) \sum_{c \in C} AND_{N(c)}(x) &= 1 - \prod_{c \in C} (1 - AND_{N(c)}(x) e^{-u}) \\ &= \sum_{S \neq \emptyset \subset C} -(-e^{-u})^{|S|} \prod_{c \in S} AND_{N(c)}(x) \\ &= \sum_{S \neq \emptyset \subset C} -(-1)^{|S|} e^{-u|S|} AND_{N(S)}(x) \\ &= \sum_{z \in B_n} \left( \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} e^{-u|S|} \right) AND_z(x), \quad (6.11) \end{aligned}$$

which verifies (6.7). Applying the linear operator  $\mu_{B_n}$  to  $\text{cover}_F = \int_0^\infty \text{skin}_{F,u} e^{-u} du$ , we get

$$\begin{aligned} \widetilde{\text{cover}}_F(z) &= \int_0^\infty \widetilde{\text{skin}}_{F,u}(z) e^{-u} du \\ &= \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \int_0^\infty e^{-u(|S|+1)} du \\ &= \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \frac{1}{|S|+1}, \end{aligned}$$

which verifies (6.10). Finally, as in Lemma 6.1, (6.10) and (6.9) follow immediately from (6.8) and (6.7) via (6.4). ■

## 6.4 The poset $B_n^{(2)}$

To generalize the monotone machinery in Sections 6.2 and 6.3 to the nonnecessarily monotone case, we basically only have to replace the poset  $B_n$  with another poset  $B_n^{(2)}$ , which we study in this section. Below, we define this poset and we give the corresponding analogs of (6.2) and (6.4).

When the DNF formula is not necessarily monotone, the AND gates are of the form

$$AND_{(z', z'')}(x) = \bigwedge_{i \in z'} x_i \wedge \bigwedge_{i \in z''} \neg x_i = AND_{z'}(x) \wedge AND_{z''}(x^c) \quad \text{for } x \in B_n,$$

where  $(z', z'') \in B_n \times B_n$  are such that  $z' \cap z'' = \emptyset$  and  $x^c \stackrel{\text{def}}{=} [n] \setminus x \in B_n$ .

This motivates looking at the poset  $B_n^{(2)}$  defined as follows. Consider the product poset  $B_n^2 \stackrel{\text{def}}{=} B_n \times B_n$ , i.e., the order relation on  $B_n^2$  is given by  $(x', x'') \leq (y', y'')$  if  $x' \leq y'$  and  $x'' \leq y''$ . Let  $B_n^{(2)}$  be the poset given by

$$B_n^{(2)} \stackrel{\text{def}}{=} \{(x', x'') \in B_n^2 : x' \cap x'' = \emptyset\}, \quad (6.12)$$

and ordered via the ordered relation of  $B_n^2$ . That is,  $B_n^{(2)}$  is the subset of  $B_n^2$  given by (6.12).

**If  $x \in B_n^{(2)}$ , we denote by  $x'$  and  $x''$  the elements of  $B_n$  such that  $x = (x', x'')$ .**

If  $z \in B_n^{(2)}$ , the corresponding AND gate  $AND_z : B_n \rightarrow \{0, 1\}$  is given by

$$AND_z(x) \stackrel{\text{def}}{=} AND_{z'}(x) \wedge AND_{z''}(x^c) \quad \text{for } x \in B_n.$$

To get a relation similar to (6.2), lift  $AND_z$  to  $B_n^{(2)}$  as follows. If  $z \in B_n^{(2)}$ , define  $\overline{AND}_z : B_n^{(2)} \rightarrow \{0, 1\}$  by

$$\overline{AND}_z(x) \stackrel{\text{def}}{=} AND_{z'}(x') \wedge AND_{z''}(x'') = \zeta_{B_n^{(2)}}(z, x).$$

Thus

$$AND_z(x) = \overline{AND}_z(x, x^c) \quad \text{for } x \in B_n,$$

and

$$\overline{AND}_z(x) = \zeta_{B_n^{(2)}}(z, x) \quad \text{for } x \in B_n^{(2)}, \quad (6.13)$$

which is the analog of (6.2) on  $B_n^{(2)}$ . We are using the bar-notation to indicate that  $\overline{AND}_z$  is the *lift* of  $AND_z$  from  $B_n$  to  $B_n^{(2)}$  (the bar-notation should not be confused with negation).

The functions  $\{\overline{AND}_z\}_{z \in B_n^{(2)}}$  form a basis for  $L(B_n^{(2)})$  since by (6.13) they are the rows of the matrix of the invertible linear transformation  $\zeta_{B_n^{(2)}}$ . Any function  $f \in L(B_n^{(2)})$  can be expressed as

$$f = \sum_{z \in B_n^{(2)}} \tilde{f}(z) \overline{AND}_z$$

for some  $\tilde{f} \in L(B_n^{(2)})$ . Indeed, by (6.13),

$$f(x) = \sum_z \tilde{f}(z) \zeta_{B_n^{(2)}}(z, x) = (\zeta_{B_n^{(2)}} f)(x).$$

That is,  $\tilde{f} = \mu_{B_n^{(2)}} f$  and  $f = \zeta_{B_n^{(2)}} \tilde{f}$ . We call  $\tilde{f} = \mu_{B_n^{(2)}} f$  the *Möbius transform* of  $f$ .

We need an analog of (6.4) on  $B_n^{(2)}$ . If  $f \in L(B_n^{(2)})$ , we relate next the Fourier transform of its projection to  $B_n$  to the Möbius transform of  $f$ .

Consider the injective embedding  $B_n \rightarrow B_n^{(2)}$ ,  $x \mapsto (x, x^c)$ . It induces the linear map  $\text{Proj} : L(B_n^{(2)}) \rightarrow L(B_n)$  given by  $(\text{Proj } f)(x) = f(x, x^c)$ . In terms of  $\text{Proj}$ , we have  $AND_z = \text{Proj } \widehat{AND}_z$ , thus

$$(\text{Proj } f)(x) = \sum_z \tilde{f}(z) AND_z(x).$$

If  $f \in L(B_n^{(2)})$ , we show below how to extract  $\widehat{\text{Proj } f}$  from  $\tilde{f}$ . Let  $z \in B_n^{(2)}$ . From (6.3), we have

$$AND_{z'}(x) = 2^{-|z'|} \sum_{y' \leq z'} (-1)^{|y'|} \mathcal{X}_{y'}(x),$$

and

$$AND_{z''}(x^c) = 2^{-|z''|} \sum_{y'' \leq z''} (-1)^{|y''|} \mathcal{X}_{y''}(x^c).$$

Now,

$$\mathcal{X}_{y''}(x^c) = (-1)^{|y'' \cap x^c|} = (-1)^{|y''| - |y'' \cap x|} = (-1)^{|y''|} \mathcal{X}_{y''}(x),$$

hence

$$AND_{z''}(x) = 2^{-|z''|} \sum_{y'' \leq z''} \mathcal{X}_{y''}(x).$$

It follows that

$$\begin{aligned} AND_z(x) &= AND_{z'}(x) AND_{z''}(x^c) \\ &= 2^{-|z'|} \sum_{y' \leq z'} (-1)^{|y'|} \mathcal{X}_{y'}(x) 2^{-|z''|} \sum_{y'' \leq z''} \mathcal{X}_{y''}(x) \\ &= 2^{-|z|} \sum_{y \leq z} (-1)^{|y'|} \mathcal{X}_{y' \oplus y''}(x) \\ &= 2^{-|z|} \sum_{y \leq z} (-1)^{|y'|} \mathcal{X}_{y' \cup y''}(x), \end{aligned} \tag{6.14}$$

where  $y' \oplus y'' = y' \cup y''$  since  $y' \cap y'' = \emptyset$ . Note that we are working in  $B_n^{(2)}$  and not in  $B_n^2$ , i.e., if  $z \in B_n^{(2)}$ , then summing over all  $y \leq z$  ( $y \geq z$ , respectively) means summing over all  $y \in B_n^{(2)}$  such that  $y \leq z$  ( $y \geq z$ , respectively). Thus

$$\begin{aligned} (\text{Proj } f)(x) &= \sum_z \tilde{f}(z) 2^{-|z|} \sum_{y \leq z} (-1)^{|y'|} \mathcal{X}_{y' \cup y''}(x) \\ &= \sum_{w \in B_n} \mathcal{X}_w(x) \sum_{a \leq w} (-1)^{|a|} \sum_{z \geq (a, w \setminus a)} 2^{-|z|} \tilde{f}(z). \end{aligned}$$

That is, the desired analog of the change of basis formula (6.4) on  $B_n^{(2)}$  is:

$$\widehat{(\text{Proj } f)}(w) = \sum_{a \leq w} (-1)^{|a|} \sum_{z \geq (a, w \setminus a)} 2^{-|z|} \tilde{f}(z) \quad \text{for all } w \in B_n \text{ and } f \in L(B_n^{(2)}). \tag{6.15}$$

Finally, we define some basic notions on  $B_n^{(2)}$  used in Section 7.3.

- **Size:** If  $x \in B_n^{(2)}$ , define the *size* or *rank* of  $x$  to be  $|x| \stackrel{\text{def}}{=} |x' \cup x''| = |x'| + |x''|$ , and note that  $0 \leq |x| \leq n$ .
- **Consistent elements and union:** We call two elements  $x, y \in B_n^{(2)}$  *consistent* if  $x'$  and  $y''$  are disjoint and  $x''$  and  $y'$  are disjoint. It is straight forward to verify that two elements  $x$  and  $y$  of  $B_n^{(2)}$  are consistent if and only if they have an *upper bound* (i.e., an element  $z$  of  $B_n^{(2)}$  such that  $z \geq x$  and  $z \geq y$ ), or equivalently, a *least upper bound* (i.e., an upper bound  $\leq$  all upper bounds of  $x$  and  $y$ ). If  $x, y \in B_n^{(2)}$  are consistent, their least upper bound, which we call also *union* and denote by  $x \cup y$ , is given by  $x \cup y \stackrel{\text{def}}{=} (x' \cup y', x'' \cup y'') \in B_n^{(2)}$ .

Let  $y, z \in B_n^{(2)}$ . If  $y$  and  $z$  are consistent, then  $\overline{AND}_y \overline{AND}_z = \overline{AND}_{y \cup z}$ . If  $y$  and  $z$  are not consistent, then  $\overline{AND}_y(x) \overline{AND}_z(x) = 0$  for all  $x \in B_n^{(2)}$ . The reason is that if  $\overline{AND}_y(x) \overline{AND}_z(x) = 1$ , then  $y \leq x$  and  $z \leq x$ , hence  $y$  and  $z$  have an upper bound, i.e., they are consistent.

- **Intersection and complement:** To avoid degenerate cases, we define the *intersection*  $x \cap y$  and *complement*  $x \setminus y$  only for consistent elements  $x, y \in B_n^{(2)}$  as follows. Let  $x \cap y \stackrel{\text{def}}{=} (x' \cap y', x'' \cap y'') \in B_n^{(2)}$  and  $y \setminus x \stackrel{\text{def}}{=} (y' \setminus x', y'' \setminus x'') \in B_n^{(2)}$ . Note that  $x \cap y \leq y$ ,  $y \setminus x \leq y$ , and  $(x \cap y) \cup (y \setminus x) = y$ . Note also that if  $x \leq y$ , then  $x$  and  $y$  are obviously consistent, hence  $y \setminus x$  is defined.
- **Separated elements:** We call two elements  $x, y \in B_n^{(2)}$  *separated* if  $x', x'', y', y''$  are mutually disjoint. Separated elements are consistent. If  $x$  and  $y$  are separated, then  $|y \cup x| = |y| + |x|$ . If  $x$  and  $y$  are consistent, but not necessarily separated, then  $y \setminus x$  and  $x$  are separated and  $(y \setminus x) \cup x = y \cup x$ , hence  $|y \cup x| = |y \setminus x| + |x|$ .

## 6.5 General DNF formulas and auxiliary functions

In this section we compute the Möbius and Fourier transforms of general DNF formulas and auxiliary functions.

We represent a DNF formula  $F$  on  $n$  variables by two bipartite graphs  $(C, [n], N')$  and  $(C, [n], N'')$  both between the set  $C$  of clauses and the set  $[n]$  of variable indices. For each clause  $c \in C$ ,  $N'(c)$  is the neighborhood of  $c$  consisting of the indices of the nonnegated variables of  $c$ , and  $N''(c)$  is the neighborhood of  $c$  consisting of the indices of the negated variables in  $c$ . We assume that  $N'(c) \cap N''(c) = \emptyset$  for each clause  $c \in C$ . To avoid degenerate cases, we assume, as in the monotone case, that we have at least one clause, i.e.,  $|C| \geq 1$ , and that each clause contains at least one literal, i.e.,  $N'(c) \cup N''(c) \neq \emptyset$  for each  $c \in C$ . Let  $N \stackrel{\text{def}}{=} (N', N'')$ , i.e.,  $N(c) \stackrel{\text{def}}{=} (N'(c), N''(c)) \in B_n^{(2)}$  for each clause  $c \in C$ , and  $N(S) \stackrel{\text{def}}{=} (N'(S), N''(S)) \in B_n^{(2)}$  for each set of clauses  $S \subset C$ . Note that  $N(S)$  is not necessarily in  $B_n^{(2)}$  since possibly  $N'(S) \cap N''(S) \neq \emptyset$ . We denote  $F$  by  $F = (C, [n], N)$ . Finally, if  $s \geq 1$  is an integer, we call a DNF formula  $F = (C, [n], N)$  an *s-DNF* formula if each clause contains at most  $s$  literals, i.e.,  $|N(c)| \leq s$  for each  $c \in C$ .

Let  $F = (C, [n], N)$  be a DNF formula. We have the boolean function computed by  $F$ , which by notational abuse we denote by  $F \in L(B_n)$ . We also have the  $u$ -skin and cover functions  $\text{cover}_F, \text{skin}_{F,u} \in L(B_n)$  associated with  $F$  (for  $u \geq 0$ ). They are given by Definition 5.6 as:  $F \stackrel{\text{def}}{=} \bigvee_{c \in C} \text{AND}_{N(c)}$ ,  $\text{skin}_{F,u} \stackrel{\text{def}}{=} 1 - (1 - e^{-u}) \sum_{c \in C} \text{AND}_{N(c)}$ ,  $\text{cover}_F \stackrel{\text{def}}{=} \int_0^\infty \text{skin}_{F,u} e^{-u} du$ , where  $\text{skin}_{F,0} = F$  by taking the limit. By lifting each of the AND gates of the DNF from  $B_n$  to  $B_n^{(2)}$ , we get the following natural lifts of  $F$  as a boolean function,  $\text{cover}_F$ , and  $\text{skin}_{F,u}$ :

**Definition 6.4 (Lifted DNF boolean function, skin, and cover)** *Let  $F = (C, [n], N)$  be a DNF formula and  $u \geq 0$ , define  $\overline{F}, \overline{\text{cover}}_F, \overline{\text{skin}}_{F,u} \in L(B_n^{(2)})$  as:*

$$\begin{aligned} \overline{F}(x) &\stackrel{\text{def}}{=} \bigvee_{c \in C} \overline{\text{AND}}_{N(c)}(x) \\ \overline{\text{skin}}_{F,u}(x) &\stackrel{\text{def}}{=} 1 - (1 - e^{-u}) \sum_{c \in C} \overline{\text{AND}}_{N(c)}(x) \\ \overline{\text{cover}}_F(x) &\stackrel{\text{def}}{=} \int_0^\infty \overline{\text{skin}}_{F,u}(x) e^{-u} du \end{aligned}$$

for all  $x \in B_n^{(2)}$ . Thus  $F = \text{Proj } \overline{F}$ ,  $\text{cover}_F = \text{Proj } \overline{\text{cover}}_F$ , and  $\text{skin}_{F,u} = \text{Proj } \overline{\text{skin}}_{F,u}$ .

Those lifted functions have natural expansions in the  $\{\overline{\text{AND}}_z\}_z$  basis from which we can extract their Möbius transforms. To get the Fourier transforms of the original functions, we use (6.15).

We have the following analog of Lemmas 6.1 and 6.2.

**Lemma 6.5** *Let  $F = (C, [n], N)$  be a DNF formula and  $u \geq 0$ . Then for all  $z \in B_n^{(2)}$  and  $w \in B_n$ ,*

$$\widetilde{\overline{F}}(z) = \sum_{S \neq \emptyset \subset C : N(S)=z} -(-1)^{|S|} \quad (6.16)$$

$$\widetilde{\overline{\text{skin}}}_{F,u}(z) = \sum_{S \neq \emptyset \subset C : N(S)=z} -(-1)^{|S|} e^{-u|S|} \quad (6.17)$$

$$\widetilde{\overline{\text{cover}}}_F(z) = \sum_{S \neq \emptyset \subset C : N(S)=z} -(-1)^{|S|} \frac{1}{|S| + 1} \quad (6.18)$$

$$\widehat{F}(w) = \sum_{a \leq w} (-1)^{|a|} \sum_{S \neq \emptyset \subset C : N'(S) \cap N''(S) = \emptyset \ \& \ N(S) \geq (a, w \setminus a)} -(-1)^{|S|} 2^{-|N(S)|} \quad (6.19)$$

$$\widehat{\text{skin}}_{F,u}(w) = \sum_{a \leq w} (-1)^{|a|} \sum_{S \neq \emptyset \subset C : N'(S) \cap N''(S) = \emptyset \ \& \ N(S) \geq (a, w \setminus a)} -(-1)^{|S|} 2^{-|N(S)|} e^{-u|S|} \quad (6.20)$$

$$\widehat{\text{cover}}_F(w) = \sum_{a \leq w} (-1)^{|a|} \sum_{S \neq \emptyset \subset C : N'(S) \cap N''(S) = \emptyset \ \& \ N(S) \geq (a, w \setminus a)} -(-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S| + 1}. \quad (6.21)$$

**Proof:** For  $u \geq 0$ , we have

$$\begin{aligned} 1 - \prod_{c \in C} (1 - \overline{\text{AND}}_{N(c)}(x) e^{-u}) &= \sum_{S \neq \emptyset \subset C} -(-e^{-u})^{|S|} \prod_{c \in S} \overline{\text{AND}}_{N(c)}(x) \\ &= \sum_{S \neq \emptyset \subset C} -(-e^{-u})^{|S|} \text{AND}_{N'(S)}(x') \text{AND}_{N''(S)}(x''). \end{aligned}$$

If  $AND_{N'(S)}(x')AND_{N''(S)}(x'') = 1$ , then  $N'(S) \leq x'$  and  $N''(S) \leq x''$ , hence  $N'(S)$  and  $N''(S)$  must be disjoint since  $x'$  and  $x''$  are disjoint, i.e., we can exclude from the sum the sets  $S$  such that  $N'(S) \cap N''(S) \neq \emptyset$ . Thus

$$\begin{aligned} 1 - \prod_{c \in C} (1 - \overline{AND}_{N(c)}(x)e^{-u}) &= \sum_{S \neq \emptyset \subset C: N'(S) \cap N''(S) = \emptyset} -(-1)^{|S|} e^{-u|S|} \overline{AND}_{N(S)}(x) \\ &= \sum_{z \in B_n^{(2)}} \left( \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} e^{-u|S|} \right) \overline{AND}_z(x). \end{aligned}$$

Setting  $u = 0$ , we get (6.16). To establish (6.17), it is enough to note that, for  $u > 0$ , we have

$$1 - (1 - e^{-u}) \sum_{c \in C} \overline{AND}_{N(c)}(x) = 1 - \prod_{c \in C} (1 - e^{-u}) \overline{AND}_{N(c)}(x) = 1 - \prod_{c \in C} (1 - \overline{AND}_{N(c)}(x)e^{-u}).$$

This verifies (6.17) for  $u > 0$ . The  $u = 0$  case follows by taking the limit.

Applying the linear operator  $\mu_{B_n^{(2)}}$  to  $\overline{\text{cover}}_F = \int_0^\infty \overline{\text{skin}}_{F,u} e^{-u} du$ , we get

$$\begin{aligned} \overline{\overline{\text{cover}}}_F(z) &= \int_0^\infty \overline{\overline{\text{skin}}}_{F,u}(z) e^{-u} du \\ &= \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \int_0^\infty e^{-u(|S|+1)} du \\ &= \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \frac{1}{|S| + 1}, \end{aligned}$$

which establishes (6.18).

The correctness of (6.19) follows from (6.16) via (6.15):

$$\begin{aligned} \widehat{F}(w) &= \sum_{a \leq w} (-1)^{|a|} \sum_{z \geq (a, w \setminus a)} 2^{-|z|} \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|} \\ &= \sum_{a \leq w} (-1)^{|a|} \sum_{S \neq \emptyset \subset C: N'(S) \cap N''(S) = \emptyset \ \& \ N(S) \geq (a, w \setminus a)} -(-1)^{|S|} 2^{-|N(S)|}. \end{aligned}$$

Similarly, (6.20) and (6.21) follow from (6.17) and (6.18) via (6.15). ■

## 6.6 Miscellaneous remarks

This section can be skipped without loss of continuity. In poset terminology<sup>5</sup>,  $B_n^{(2)}$  can be alternatively defined as the order ideal of  $B_n^2$  generated by the antichain  $A_n \stackrel{\text{def}}{=} \{(x, x^c) : x \in B_n\} \subset B_n^2$ .

---

<sup>5</sup>An *antichain* of a poset  $X$  is a subset  $A$  of  $X$  such that any two distinct elements of  $A$  are incomparable. A subset  $I$  of  $X$  is called an *order ideal* if  $x \in I$  and  $y \leq x$ , then  $y \in I$ . We say that an order ideal is *generated by* a subset  $A$  of  $X$  if  $I = \{x \in X : x \leq y \text{ for some } y \in A\}$ . Any order ideal has a generating antichain.

If  $X$  is a poset, the matrix coefficients of  $\mu_X$  are denoted by  $(\mu_X(y, x))_{x, y}$ , i.e.,

$$(\mu_X f)(x) = \sum_{y \in X} \mu_X(y, x) f(y). \quad (6.22)$$

Since  $\zeta_X$  is lower triangular,  $\mu_X$  is also lower triangular, i.e.,  $\mu_X(y, x) = 0$  if  $y \not\leq x$ .

It is worth mentioning that the coefficients of the Möbius functions of the posets  $B_n$  and  $B_n^{(2)}$  have the following simple expressions:

- i)  $\mu_{B_n}(y, x) = (-1)^{|x \setminus y|}$  if  $y \leq x \in B_n$  (see [Sta97]).
- ii)  $\mu_{B_n^{(2)}}(y, x) = (-1)^{|x \setminus y|}$  if  $y \leq x \in B_n^{(2)}$ . This can be derived from (i) via the fact that  $B_n^{(2)}$  is an order ideal of  $B_n^2 = B_n \times B_n$  (In general, if  $I$  is an order ideal of a poset  $X$  regarded as a subposet of  $X$ , then  $\mu_I(y, x) = \mu_X(y, x)$  for all  $x, y \in I$ ).

Those expressions are not used in the poof since, instead of using (6.22) to compute the Möbius transforms of DNF formulas and their auxiliary functions, we extracted the Möbius transforms from natural expansions of the functions in the  $\{AND_z\}_z$  basis.

## 7 From zero-energy to energies of auxiliary functions

In this section, we reduce the problem of bounding the  $t$ -zero-energy of an  $s$ -DNF formula  $F$  to that of bounding the  $(t - s)$ -energies of auxiliary functions derived from  $F$ .

Let  $F = (C, [n], N)$  be an  $s$ -DNF formula, and let  $t \geq s$  be an integer. We want to bound the  $t$ -zero-energy of  $F$ . That is, we want to construct  $f \in L(B_n)$  of degree at most  $t$  such that the mean square error  $E(F - f)^2$  is small, and  $f$  satisfies the zeros-constraint:  $f = 0$  whenever  $F = 0$  (i.e.,  $f(x) = 0$  for each  $x \in \{0, 1\}^n$  such that  $F(x) = 0$ ). For any such  $f$ , we have the bound  $\text{zeroEnergy}(F; t) \leq E(F - f)^2 = \|\widehat{F - f}\|_2^2$ .

We presented the construction of  $f$  in the monotone case in Section 5.6 without analyzing its mean square error. In Sections 7.1 and 7.2, we analyze the Fourier transform of the construction error term  $\widehat{F - f}$  in the monotone case. Then we generalize to arbitrary DNF formulas in Section 7.3. The analysis of  $\widehat{F - f}$  naturally leads us to the cover and  $u$ -skin functions. In Lemmas 7.1 and 7.2, we express  $\widehat{F - f}$  in terms of the Fourier transforms of cover functions of DNF derived from  $F$ . The analysis in Sections 7.1, 7.2, and 7.3 uses the machinery developed in Section 6.

In Section 7.4, we apply two simple bounds to the obtained expression of  $\widehat{F - f}$ . The first bound reduces the problem of bounding the  $t$ -zero-energy of  $F$  to that of bounding the  $(t - s)$ -energies of the cover functions of the derived DNF formulas. The second bound reduces the latter problem to bounding the  $(t - s)$ -energies of the  $u$ -skin functions of the derived formulas, for all  $u \geq 0$ .

## 7.1 Monotone case error analysis

This section uses the monotone machinery developed in Sections 6.2 and 6.3.

In this section we analyze the mean square error of the monotone construction defined in Section 5.6. Assume that  $F = (C, [n], N)$  is monotone and construct  $f$  as in Section 5.6. That is, define  $f \in L(B_n)$  as

$$f \stackrel{\text{def}}{=} \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} E_{c \in S} \text{AND}_{N(c)} \text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)}. \quad (7.1)$$

We know from Section 5.6 that  $\deg(f) \leq t$  and  $f$  satisfies the zeros-constraint.

The difficult part is estimating the mean square error  $E(F - f)^2 = \|\widehat{F - f}\|_2^2$  as  $t$  grows. Toward this end, we start analyzing  $\widehat{F - f}$  by first computing the Fourier transform of  $f$ , which gives some intuition as to why one would speculate that the mean square error of this construction decays as  $t$  grows. Then we interpret  $\widehat{F - f}$  in Lemma 7.1 in terms of the Fourier transforms of cover functions of DNF formulas derived from  $F$ .

Recall from (6.6) that

$$\widehat{F}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|} \quad (7.2)$$

We show below that the Fourier transform of  $f$  is given by

$$\widehat{f}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|} \text{Pr}_{c \in S}[|y \cup N(c)| \leq t]. \quad (7.3)$$

**Proof of (7.3):** Recall from (6.3) that

$$\text{AND}_z = \frac{1}{2^{|z|}} \sum_{y \leq z} (-1)^{|y|} \mathcal{X}_y.$$

Let  $S \neq \emptyset \subset C$ , and let  $c \in S$ . Then

$$\text{AND}_{N(c)} = \frac{1}{2^{|N(c)|}} \sum_{y_1 \leq N(c)} (-1)^{|y_1|} \mathcal{X}_{y_1}$$

and

$$\text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)} = \frac{1}{2^{|N(S) \setminus N(c)|}} \sum_{y_2 \leq N(S) \setminus N(c): |y_2| \leq t - |N(c)|} (-1)^{|y_2|} \mathcal{X}_{y_2}.$$

Thus

$$\begin{aligned} \text{AND}_{N(c)} \text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)} &= \frac{1}{2^{|N(S)|}} \sum_{\substack{y_1 \leq N(c), \\ y_2 \leq N(S) \setminus N(c): \\ |y_2| \leq t - |N(c)|}} (-1)^{|y_1| + |y_2|} \mathcal{X}_{y_1 \oplus y_2} \\ &= \frac{1}{2^{|N(S)|}} \sum_{y \leq N(S): |y \cup N(c)| \leq t} (-1)^{|y|} \mathcal{X}_y, \end{aligned} \quad (7.4)$$

where we used the fact that  $y_1 \oplus y_2 = y_1 \cup y_2$  since  $y_1$  and  $y_2$  are disjoint because  $N(c)$  and  $N(S) \setminus N(c)$  are disjoint.

Substituting (7.4) in (7.1) and writing the expectation operator in (7.1) as  $E_{c \in S} = \frac{1}{|S|} \sum_{c \in S}$ , we get

$$\begin{aligned} f &= \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} \frac{1}{|S|} \sum_{c \in S} 2^{-|N(S)|} \sum_{y \leq N(S): |y \cup N(c)| \leq t} (-1)^{|y|} \mathcal{X}_y \\ &= \sum_y \mathcal{X}_y (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S|} \sum_{c \in S: |y \cup N(c)| \leq t} 1, \end{aligned}$$

after rearranging the summations. Noting that  $\frac{1}{|S|} \sum_{c \in S: |y \cup N(c)| \leq t} 1 = \Pr_{c \in S}[|y \cup N(c)| \leq t]$ , we obtain (7.3).  $\blacksquare$

Comparing the expressions of  $\widehat{F}$  and  $\widehat{f}$  in (7.2) and (7.3), and noting that  $|N(c)| \leq s \leq t$  for each  $c \in C$  since  $F$  is an  $s$ -DNF formula, we get

$$\widehat{f}(y) = \begin{cases} \widehat{F}(y) & \text{if } |y| \leq t - s \\ 0 & \text{if } |y| > t. \end{cases}$$

This relation gives some intuition why the mean square error  $\|\widehat{F} - \widehat{f}\|_2^2$  of this construction decays as  $t$  grows. It says that  $\widehat{f}(y) = \widehat{\text{Trn}_t F}(y)$  if  $|y| \leq t - s$  or  $|y| > t$ . We know from LMN energy bound (Theorem 5.8) that  $E(F - \text{Trn}_t F)^2 = \|\widehat{F} - \widehat{\text{Trn}_t F}\|_2^2$  decays quickly as  $t$  grows. Thus, we can hope that  $\widehat{f}(y)$  is not too bad in the region  $t - s < |y| \leq t$  and accordingly speculate that the mean square error  $\|\widehat{F} - \widehat{f}\|_2^2$  also decays as  $t$  grows.

Unfortunately, we could not turn this intuition into a bound on  $\|\widehat{F} - \widehat{f}\|_2^2$  since the frequencies in the region  $t - s < |y| \leq t$  are too many to handle separately by trivial bounds. We can think of other equally intuitive constructions, which we briefly mention in Section 7.2.D below. The reason behind favoring this choice of  $f$  is analytical. Using analytical means, we managed to bound  $\|\widehat{F} - \widehat{f}\|_2^2$  by interpreting the Fourier transform of the error term  $f - F$  in terms of the Fourier transforms of cover functions of DNF formulas derived from  $F$  as shown in Lemma 7.1 below. We consider  $f - F$  instead of  $F - f$  for technical convenience.

**Lemma 7.1 (Error interpretation)** *Let  $F = (C, [n], N)$  be a monotone  $s$ -DNF formula, and let  $t \geq s$  be an integer. Assume that  $|C| \geq 2$ . Let  $f \in L(B_n)$  be given by*

$$f = \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} E_{c \in S} \text{AND}_{N(c)} \text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)}. \quad (7.5)$$

*For each clause  $c \in C$ , define the new DNF formula  $F_c = (C_c, [n], N_c)$  resulting from removing the clause  $c$  from  $F$  and adding its variables to all the other clauses, i.e.,  $C_c = C \setminus \{c\}$  and  $N_c(d) = N(d) \cup N(c)$  for each  $d \in C_c$ .*

*Then: i)  $\deg(f) \leq t$ ; ii)  $f$  satisfies the zeros-constraint:  $f = 0$  whenever  $F = 0$ ; and iii) for each  $y \in B_n$*

$$\widehat{(f - F)}(y) = \sum_{c \in C: |y \cup N(c)| > t} \widehat{\text{cover}_{F_c}}(y). \quad (7.6)$$

We show in Section 7.4 that Equation (7.6) immediately leads to a bound on  $\|\widehat{f - F}\|_2^2$  in terms of the  $(t - s)$ -energies of the covers of the derived DNF formulas.

It is important to note that the following error analysis is the origin of the cover and skin functions. We highlight in Section 7.2.A how the error analysis in the proof below naturally leads to the definitions of skin and cover.

**Proof of Lemma 7.1:** We have (i) and (ii) from Section 5.6. We need to establish (7.6). Let

$$\Delta \stackrel{\text{def}}{=} f - F.$$

Subtracting the summation (7.2) from (7.3), we get

$$\widehat{\Delta}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} (-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S|} \sum_{c \in S: |y \cup N(c)| > t} 1 \quad (7.7)$$

for all  $y \in B_n$ .

First, for technical convenience, we note that the summation  $\sum_{S \subset C: S \neq \emptyset}$  in (7.7) can be replaced with  $\sum_{S \subset C: |S| > 1}$ . The condition  $|S| > 1$  is nonrestrictive since the size-1 subsets  $S$  of  $C$  do not contribute to the summation. Indeed, assume that  $|S| = 1$ , hence  $S = \{c_0\}$  for some  $c_0 \in C$ . Thus the expression in (7.7) is nonzero only if  $N(c_0) \geq y$  and  $|y \cup N(c_0)| > t$ . But then we get  $|N(c_0)| > t$  since  $y \cup N(c_0) = N(c_0)$  because  $N(c_0) \geq y$ . This is not possible since  $|N(c_0)| \leq s$  because  $F$  is an  $s$ -DNF and  $s \leq t$  by the lemma hypothesis.

If we replace  $\sum_{S \subset C: S \neq \emptyset}$  with  $\sum_{S \subset C: |S| > 1}$  and reverse the order of the summations in (7.7), we get

$$\begin{aligned} \widehat{\Delta}(y) &= \sum_{\substack{c \in C: \\ |y \cup N(c)| > t}} (-1)^{|y|} \sum_{\substack{S \subset C: |S| > 1 \\ N(S) \geq y \\ c \in S}} (-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S|} \\ &= \sum_{\substack{c \in C: \\ |y \cup N(c)| > t}} (-1)^{|y|} \sum_{z \geq y} \left( \sum_{\substack{S \subset C: |S| > 1 \\ N(S) = z \ \& \ c \in S}} (-1)^{|S|} \frac{1}{|S|} \right) 2^{-|z|}. \end{aligned}$$

Using the change of basis formula (6.4), we obtain

$$\widehat{\Delta}(y) = \sum_{c \in C: |y \cup N(c)| > t} \widehat{X}_c(y),$$

where  $X_c \in L(B_n)$  is a function given by its Möbius transform

$$\widetilde{X}_c(z) = \sum_{\substack{S \subset C: |S| > 1 \\ N(S) = z \\ c \in S}} (-1)^{|S|} \frac{1}{|S|}.$$

By a change of variables from  $S$  to  $T = S \setminus \{c\}$ , we can write this as:

$$\widetilde{X}_c(z) = \sum_{\substack{T \neq \emptyset \subset C \setminus \{c\} : \\ N(T \cup \{c\}) = z}} (-1)^{|T|+1} \frac{1}{|T|+1}.$$

By the definition of the formula  $F_c$ , we have  $C_c = C \setminus \{c\}$  and  $N_c(d) = N(d) \cup N(c)$  for each  $d \in C_c$ . Hence for each  $T \subset C_c$ ,  $N_c(T) = \cup_{d \in T} N_c(d) = \cup_{d \in T \cup \{c\}} N(d) = N(T \cup \{c\})$ . Thus

$$\widetilde{X}_c(z) = \sum_{T \neq \emptyset \subset C_c : N_c(T) = z} -(-1)^{|T|} \frac{1}{|T|+1}. \quad (7.8)$$

Using (6.8), we identify this as the Möbius transform of the cover of  $F_c$ , i.e.,  $X_c = \text{cover}_{F_c}$ , which proves (7.6).  $\blacksquare$

## 7.2 Discussion

In this section, we make some remarks related to the above construction.

**A. Origin of the cover and skin functions.** Equation (7.6) is the reason behind our interest in the cover and skin functions. We explain below how the analysis of  $\widehat{f - F}$  lead us to the cover and skin functions. As shown in the above proof of Lemma 7.1,  $\widehat{f - F}$  can be expressed as

$$\widehat{(f - F)}(y) = \sum_{c \in C : |y \cup N(c)| > t} \widehat{X}_c(y),$$

for some function  $X_c \in L(B_n)$  whose Möbius transform is given by (7.8). By expressing  $\frac{1}{|T|+1}$  as  $\frac{1}{|T|+1} = \int_0^\infty e^{-u|T|} e^{-u} du$ , we concluded that  $X_c = \int_0^\infty Y_{c,u} e^{-u} du$ , for some family of functions  $Y_{c,u} \in L(B_n)$  whose Möbius transforms are given by:

$$\widetilde{Y}_{c,u}(z) = \sum_{T \neq \emptyset \subset C_c : N_c(T) = z} -(-1)^{|T|} e^{-u|T|}.$$

The  $u$ -skin function was identified from its Möbius transform via (6.11) with respect to  $F_c$ :

$$\sum_{z \in B_n} \left( \sum_{T \neq \emptyset \subset C_c : N_c(T) = z} -(-1)^{|T|} e^{-u|T|} \right) \text{AND}_z(x) = 1 - (1 - e^{-u}) \sum_{d \in C_c} \text{AND}_{N_c(d)}(x).$$

That is, we first did the computations in Lemma 6.2 backward on  $F_c$  to first conclude that  $Y_{c,u} = \text{skin}_{F_c,u}$  and hence  $X_c = \text{cover}_{F_c}$  by evaluating the integral  $\int_0^\infty \text{skin}_{F_c,u} e^{-u} du$ .

The right way to understand (7.6) is in the Fourier domain. It does not have a simple analogue outside this domain due to the condition  $|y \cup N(c)| > t$  on  $y$  in the summation.

We do not have an intuitive nonanalytical explanation of (7.6). Recall that we constructed  $f$  so that it satisfies the zeros-constraint and the low degree condition. As explained in the discussion preceding the theorem statement, the construction intuition is “hopefully  $\widehat{f}(y)$ ”

is not too bad in the region  $t - s < |y| \leq t''$ . The same intuition applies to other similar constructions of  $f$  which we failed to analyze (see below). The issue is that we could not turn this intuition into an argument. We managed instead in (7.6) to interpret the construction error term using analytical means which lead us to the cover and skin functions.

**B. Identifying the components of (7.6).** From a big perspective, the components of (7.6) can be identified with the definition of  $f$  in (7.5) as follows. Write the expectation operator in (7.5) as  $E_{c \in S} = \frac{1}{|S|} \sum_{c \in S}$ . The summation on  $c$  in  $E_{c \in S}$  corresponds to the summation on  $c$  in (7.6). The latter summation is subject to the condition  $|y \cap N(c)| > t$  which comes from the truncation operator in (7.5). The  $\frac{1}{|S|}$  term of the expectation operator  $E_{c \in S}$  corresponds to the  $\frac{1}{|T|+1}$  weighting factor in (7.8) via the change of variables done in the proof of Lemma 7.1 from  $S$  to  $T = S \setminus \{c\}$ . Note that this  $\frac{1}{|T|+1}$  weighting factor is what distinguishes the cover of  $F_c$  from  $F_c$  in the Möbius and Fourier domains (see Remark 6.3.2).

**C. Duplicate clauses issue.** It is possible that there exist  $c, d_1, d_2 \in C$ , such that  $N(d_1) \neq N(d_2)$ , but  $N(c) \cup N(d_1) = N(c) \cup N(d_2)$ , i.e.,  $N_c(d_1) = N_c(d_2)$ . Thus it is possible that the DNF formula  $F_c$  has duplicate clauses even if  $F$  does not have. Duplicate clauses in  $F_c$  affect the function  $\text{cover}_{F_c}$ . This is the reason why we chose to represent a DNF formula by a bipartite graph and not a set of clauses, i.e., a subset of  $B_n$  (or  $B_n^{(2)}$  in the general case).

**D. Alternative constructions.** In what follows, we briefly mention two intuitive alternative constructions of  $f$  which we failed to analyze. Instead of starting from (5.1), it is natural to try grouping terms first, i.e., express

$$F(x) = \sum_{z \in B_n} \tilde{F}(z) \text{AND}_z(x),$$

where the Möbius transform of  $F$  is given in (6.5) by  $\tilde{F}(z) = \sum_{S \neq \emptyset \subset C: N(S)=z} -(-1)^{|S|}$ . Starting from this expression, we can define

$$f' = \sum_z \tilde{F}(z) E_{c \in N^{-1}(z)} \text{AND}_{N(c)} \text{Trn}_{t-|N(c)|} \text{AND}_{z \setminus N(c)}.$$

where  $N^{-1}(z) = \{c \in C : c \leq z\}$ . Here again the degree of  $f'$  is at most  $t$ ,  $f'$  satisfies the zeros-constraint, and the same intuition behind the above construction of  $f$  applies to  $f'$ . We can also express  $\widehat{(f' - F)}(y)$  similarly to (7.6) as a sum over  $c \in C$  of the Fourier coefficients of some functions. The issue however is that, unlike the covers of the derived formulas, those function are hard to analyze and they have no clear interpretation.

The second construction is the following. Rather than averaging over all the  $c \in S$ , we could have fixed a arbitrary map  $\alpha : 2^C \rightarrow C$  which attaches to each  $S \subset C$  a fixed element  $c_S \stackrel{\text{def}}{=} \alpha(S) \in S$  (e.g., the smallest clause in  $S$ ). Then we can define

$$f_\alpha = \sum_{S \subset C: S \neq \emptyset} -(-1)^{|S|} \text{AND}_{N(c_S)} \text{Trn}_{t-|N(c_S)|} \text{AND}_{N(S) \setminus N(c_S)}.$$

Here again, the degree of  $f_\alpha$  is at most  $t$ ,  $f_\alpha$  satisfies the zeros-constraint, and the same intuition behind the above construction of  $f$  applies to  $f_\alpha$ . The issue is again in the difficulty

of interpreting and analyzing the functions in the resulting analog of (7.6). We can view  $f$  however as the average of  $f_\alpha$  over all the maps  $\alpha$ , i.e.,  $f = E_\alpha f_\alpha$ .

### 7.3 General case construction

This section uses the general machinery in Sections 6.4 and 6.5.

If  $F$  is not necessarily monotone, the monotone case construction generalizes naturally as follows. Following the derivations in the proof of Lemma 6.5 (for  $u = 0$ ), expand  $F$  as

$$\begin{aligned}
F(x) &= 1 - \prod_{c \in C} (1 - \text{AND}_{N(c)}(x)) \\
&= \sum_{S \neq \emptyset \subset C} -(-1)^{|S|} \prod_{c \in S} \text{AND}_{N(c)}(x) \\
&= \sum_{S \neq \emptyset \subset C} -(-1)^{|S|} \text{AND}_{N'(S)}(x') \text{AND}_{N''(S)}(x'') \\
&= \sum_{S \neq \emptyset \subset C: N'(S) \cap N''(S) = \emptyset} -(-1)^{|S|} \text{AND}_{N'(S)}(x') \text{AND}_{N''(S)}(x'') \\
&= \sum_{S \neq \emptyset \subset C: N'(S) \cap N''(S) = \emptyset} -(-1)^{|S|} \text{AND}_{N(S)}(x).
\end{aligned}$$

Let  $S \subset C$  such that  $S \neq \emptyset$  and  $N'(S) \cap N''(S) = \emptyset$ , i.e.,  $N(S) \in B_n^{(2)} \setminus \{(\emptyset, \emptyset)\}$ . For each  $c \in S$ , we have

$$\text{AND}_{N(S)} = \text{AND}_{N(c)} \text{AND}_{N(S) \setminus N(c)}$$

Recall from Section 6.4 that we defined  $y \setminus x \in B_n^{(2)}$  for consistent elements  $x, y \in B_n^{(2)}$ , and note that  $N(c)$  and  $N(S)$  are consistent since  $N(c) \leq N(S)$ . Averaging over all  $c \in S$ , we trivially get

$$\text{AND}_{N(S)} = E_{c \in S} \text{AND}_{N(c)} \text{AND}_{N(S) \setminus N(c)},$$

hence

$$F = \sum_{S \neq \emptyset \subset C: N'(S) \cap N''(S) = \emptyset} -(-1)^{|S|} E_{c \in S} \text{AND}_{N(c)} \text{AND}_{N(S) \setminus N(c)}.$$

We construct  $f$  as in the monotone case by truncating each  $\text{AND}_{N(S) \setminus N(c)}$  to  $\text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)}$ .

**Lemma 7.2 (Error interpretation)** *Let  $F = (C, [n], N)$  be an  $s$ -DNF formula and let  $t \geq s$  be an integer. Let  $f \in L(B_n)$  be given by*

$$f = \sum_{S \subset C: S \neq \emptyset \ \& \ N'(S) \cap N''(S) = \emptyset} -(-1)^{|S|} E_{c \in S} \text{AND}_{N(c)} \text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)}.$$

*If  $c \in C$ , let  $C_c$  be the set of clauses other than  $c$  which are consistent with  $c$ , i.e.,  $C_c = \{d \in C \setminus \{c\} : N(d) \text{ and } N(c) \text{ are consistent}\}$ . Let  $C_{\text{main}}$  be the set of clauses which are consistent with at least one clause of  $F$  other than themselves, i.e.,  $C_{\text{main}} = \{c \in C : C_c \neq \emptyset\}$ .*

For each clause  $c \in C_{\text{main}}$ , define the new DNF formula  $F_c = (C_c, [n], N_c)$ , where  $N_c(d) = N(d) \cup N(c)$  for each  $d \in C_c$ . That is,  $F_c$  is the formula resulting from removing from  $F$  the clause  $c$  and all the clauses not consistent with  $c$ , and adding the literals of  $c$  to each of the remaining clauses.

Then: i)  $\deg(f) \leq t$ ; ii)  $f$  satisfies the zeros-constraint:  $f = 0$  whenever  $F = 0$ ; and iii) for each  $w \in B_n$

$$\widehat{(f - F)}(w) = \sum_{c \in C_{\text{main}}: |w \cup N'(c) \cup N''(c)| > t} \widehat{\text{cover}}_{F_c}(w). \quad (7.9)$$

**Proof:** We have  $\deg(f) \leq t$  since the degree of each  $\text{Trn}_{t-|N(c)|} \text{AND}_{N(S) \setminus N(c)}$  is at most  $t - |N(c)|$  and the degree of  $\text{AND}_{N(c)}$  is  $|N(c)|$ . Moreover, if  $F(x) = 0$ , then  $\text{AND}_{N(c)}(x) = 0$  for each  $c \in C$ , thus  $f(x) = 0$ , and hence (ii). We have to establish (7.9). Let

$$\Delta \stackrel{\text{def}}{=} f - F.$$

We have

$$\begin{aligned} \Delta &= \sum_{S \neq \emptyset \subset C: N'(S) \cap N''(S) = \emptyset} (-1)^{|S|} E_{c \in S} \text{AND}_{N(c)} (1 - \text{Trn}_{t-|N(c)|}) \text{AND}_{N(S) \setminus N(c)}. \\ &= \sum_{S \subset C: |S| > 1 \ \& \ N'(S) \cap N''(S) = \emptyset} (-1)^{|S|} E_{c \in S} \text{AND}_{N(c)} (1 - \text{Trn}_{t-|N(c)|}) \text{AND}_{N(S) \setminus N(c)}. \end{aligned} \quad (7.10)$$

As in the monotone case, we impose the nonrestrictive condition  $|S| > 1$  for technical convenience. This is nonrestrictive since if  $|S| = 1$ , then  $S = \{c_0\}$  for some  $c_0 \in C$ , hence  $\text{AND}_{N(S) \setminus N(c_0)} = 1$ . Therefore  $(1 - \text{Trn}_{t-|N(c_0)|}) \text{AND}_{N(S) \setminus N(c_0)} = 0$  since  $t - |N(c_0)| \geq 0$  because  $|N(c_0)| \leq s$  as  $F$  is an  $s$ -DNF and  $t \geq s$  by the lemma hypothesis.

Recall from (6.14) that

$$\text{AND}_z = \sum_{y \leq z} (-1)^{|y'|} \mathcal{X}_{y' \cup y''},$$

and recall also that we are working in  $B_n^{(2)}$  and not in  $B_n^2$ , i.e., if  $z \in B_n^{(2)}$ , then summing over all  $y \leq z$  ( $y \geq z$ , respectively) means summing over all  $y \in B_n^{(2)}$  such that  $y \leq z$  ( $y \geq z$ , respectively).

Let  $S \neq \emptyset \subset C$  such that  $N'(S) \cap N''(S) = \emptyset$ , i.e., such that  $N(S) \in B_n^{(2)}$ , and let  $c \in S$ . Then

$$\text{AND}_{N(c)} = \frac{1}{2^{|N(c)|}} \sum_{y_1 \leq N(c)} (-1)^{|y_1'|} \mathcal{X}_{y_1' \cup y_1''}$$

and

$$(1 - \text{Trn}_{t-|N(c)|}) \text{AND}_{N(S) \setminus N(c)} = \frac{1}{2^{|N(S) \setminus N(c)|}} \sum_{y_2 \leq N(S) \setminus N(c): |y_2| > t - |N(c)|} (-1)^{|y_2'|} \mathcal{X}_{y_2' \cup y_2''}.$$

Thus

$$\begin{aligned}
AND_{N(c)}(1 - \text{Trn}_{t-|N(c)|})AND_{N(S)\setminus N(c)} &= \frac{1}{2^{|N(S)|}} \sum_{\substack{y_1 \leq N(c), \\ y_2 \leq N(S)\setminus N(c) : \\ |y_2| > t - |N(c)|}} (-1)^{|y'_1|+|y'_2|} \mathcal{X}_{y'_1 \cup y'_1 \oplus y'_2 \cup y'_2} \\
&= \frac{1}{2^{|N(S)|}} \sum_{y \leq N(S) : |y \setminus N(c)| > t - |N(c)|} (-1)^{|y'|} \mathcal{X}_{y' \cup y''}. \quad (7.11)
\end{aligned}$$

To verify (7.11), recall first from Section 6.4 the definitions of basic operations in  $B_n^{(2)}$ . (7.11) follows from the fact that  $N(c)$  and  $N(S)\setminus N(c)$  are separated and hence  $y_1$  and  $y_2$  are separated. Thus summing over  $y_1 \leq N(c)$  and  $y_2 \leq N(S)\setminus N(c)$  is equivalent to summing over  $y = y_1 \cup y_2 \leq N(c) \cup (N(S)\setminus N(c)) = N(S)$ . Since  $y_1$  and  $y_2$  are separated, i.e.,  $y'_1, y''_1, y'_2, y''_2$  are mutually disjoint, we have  $y'_1 \cup y''_1 \oplus y'_2 \cup y''_2 = y'_1 \cup y''_1 \cup y'_2 \cup y''_2 = y' \cup y''$  and  $|y'_1| + |y'_2| = |y'_1 \cup y'_2| = |y'|$ .

Now, note that

$$|y \setminus N(c)| + |N(c)| = |y \cup N(c)| = |(y \cup N(c))' \cup (y \cup N(c))''| = |y' \cup N'(c) \cup y'' \cup N''(c)|. \quad (7.12)$$

The first equality holds because, in general, if  $x$  and  $y$  are consistent, then  $|y \cup x| = |y \setminus x| + |x|$  as noted at the end of Section 6.4 ( $y$  and  $N(c)$  are consistent as they have the upper bound  $N(S)$  in  $B_n^{(2)}$ ). The second (third, respectively) equality follows from the definition of size (union, respectively) in  $B_n^{(2)}$ .

Substituting (7.12) in (7.11), and then (7.11) in (7.10), we get

$$\begin{aligned}
\Delta &= \sum_{\substack{S \subset C : |S| > 1 \\ N'(S) \cap N''(S) = \emptyset}} (-1)^{|S|} \frac{1}{|S|} \sum_{c \in S} \frac{1}{2^{|N(S)|}} \sum_{\substack{y \leq N(S) : \\ |(y' \cup y'') \cup N'(c) \cup N''(c)| > t}} (-1)^{|y'|} \mathcal{X}_{y' \cup y''}
\end{aligned}$$

By rearranging the summations, we extract the Fourier coefficients of  $\Delta$ :  $\Delta = \sum_{w \in B_n} \widehat{\Delta}(w) \mathcal{X}_w$ , where

$$\begin{aligned}
\widehat{\Delta}(w) &= \sum_{\substack{S \subset C : |S| > 1 \\ N'(S) \cap N''(S) = \emptyset}} (-1)^{|S|} \frac{1}{|S|} \sum_{\substack{c \in S : \\ |w \cup N'(c) \cup N''(c)| > t}} \frac{1}{2^{|N(S)|}} \sum_{\substack{a \leq w : \\ (a, w \setminus a) \leq N(S)}} (-1)^{|a|} \\
&= \sum_{\substack{c \in C : \\ |w \cup N'(c) \cup N''(c)| > t}} \sum_{a \leq w} (-1)^{|a|} \sum_{\substack{S \subset C : |S| > 1 \\ N'(S) \cap N''(S) = \emptyset \\ N(S) \geq (a, w \setminus a) \\ c \in S}} (-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S|} \\
&= \sum_{\substack{c \in C : \\ |w \cup N'(c) \cup N''(c)| > t}} \sum_{a \leq w} (-1)^{|a|} \sum_{z \geq (a, w \setminus a)} \left( \sum_{\substack{S \subset C : |S| > 1 \\ N(S) = z \text{ \& } c \in S}} (-1)^{|S|} \frac{1}{|S|} \right) 2^{-|z|}.
\end{aligned}$$

Using (6.15), we obtain

$$\widehat{\Delta}(w) = \sum_{\substack{c \in C : \\ |w \cup N'(c) \cup N''(c)| > t}} \widehat{\text{Proj } X_c}(w), \quad (7.13)$$

where  $X_c \in L(B_n^{(2)})$  is given by its Möbius transform

$$\begin{aligned} \widetilde{X}_c(z) &= \sum_{\substack{S \subset C : |S| > 1 \\ N(S) = z \ \& \ c \in S}} (-1)^{|S|} \frac{1}{|S|} \\ &= \sum_{T \neq \emptyset \subset C \setminus \{c\} : N(T \cup \{c\}) = z} (-1)^{|T|+1} \frac{1}{|T| + 1} \end{aligned}$$

after a change of variables from  $S$  to  $T = S \setminus \{c\}$ .

We will show that  $X_c = \overline{\text{cover}}_{F_c}$  if  $c \in C_{main}$ , and  $X_c = 0$  if  $c \in C \setminus C_{main}$ .

First we handle the degenerate case when  $c \in C \setminus C_{main}$ . Assume  $c \in C \setminus C_{main}$ , thus there is no  $d \neq c \in C$  such that  $N(d)$  and  $N(c)$  are consistent. Hence  $T = \emptyset$  is the only  $T \subset C \setminus \{c\}$  such that  $N(T \cup \{c\}) \in B_n^{(2)}$ . But  $T = \emptyset$  is not allowed in the summation. It follows that  $X_c = 0$ .

Now, assume that  $c \in C_{main}$ . By the definition of the formula  $F_c$ ,  $C_c = \{d \in C \setminus \{c\} : N(d) \text{ and } N(c) \text{ are consistent}\} \neq \emptyset$ , and  $N_c(d) = N(d) \cup N(c)$  for each  $d \in C_c$ . Hence for each  $T \subset C_c$ ,

$$N_c(T) = (N'_c(T), N''_c(T)) = (N'(T \cup \{c\}), N''(T \cup \{c\})) = N(T \cup \{c\}).$$

Moreover, if  $T \subset C \setminus \{c\}$  but  $T \not\subset C_c$ , then  $T$  contains an element  $d$  of  $C$  such that  $N(c)$  and  $N(d)$  are not consistent. Consequently,  $N'(\{c, d\}) \cap N''(\{c, d\}) \neq \emptyset$ , and hence  $N'(T \cup \{c\}) \cap N''(T \cup \{c\}) \neq \emptyset$ , i.e.,  $N(T \cup \{c\}) \notin B_n^{(2)}$ . Therefore  $T$  does not contribute to the summation. It follows that

$$\widetilde{X}_c(z) = \sum_{T \neq \emptyset \subset C_c : N_c(S) = z} -(-1)^{|T|} \frac{1}{|T| + 1}.$$

Using (6.18), we identify this as the Möbius transform of the lifted cover of  $F_c$ , i.e.,  $X_c = \overline{\text{cover}}_{F_c}$ . Thus  $\text{Proj } X_c = \text{cover}_{F_c}$ , and hence (7.9) follows from (7.13).  $\blacksquare$

## 7.4 Bounds

The analysis in Sections 7.1 and 7.3 consists of exact derivations not because we are interested in exact evaluations, but because we could not use approximations since the involved summations have exponentially many terms of alternating signs. The expression of  $\widehat{f - \bar{F}}$  in Lemma 7.1 (or Lemma 7.2) can be regarded as a way to hide those huge summations in the Fourier transforms of the cover functions.

In this section, we apply two simple bounds to this expression. The first bound (Theorem 7.3) reduces the problem of bounding the  $t$ -zero-energy of  $F$  to that of bounding the  $(t - s)$ -energies of the cover functions of DNF formulas derived from  $F$ . The second bound (Lemma 7.4) reduces the latter problem to bounding the  $(t - s)$ -energies of the  $u$ -skin functions of the derived formulas, for all  $u \geq 0$ .

**Theorem 7.3 (zero-energy  $\prec$  energy of cover)** *Let  $F = (C, [n], N)$  be an  $s$ -DNF formula and let  $t \geq s$  be an integer.*

- a) **(Monotone case)** *Assume that  $F$  is monotone and  $|C| \geq 2$ . For each clause  $c \in C$ , define the new DNF formula  $F_c = (C_c, [n], N_c)$  resulting from removing the clause  $c$  from  $F$  and adding its variables to all the other clauses, i.e.,  $C_c = C \setminus \{c\}$  and  $N_c(d) = N(d) \cup N(c)$  for each  $d \in C_c$ . Then*

$$\text{zeroEnergy}(F; t) \leq |C|^2 \max_{c \in C} \text{energy}(\text{cover}_{F_c}; t - s).$$

- b) **(General case)** *In general, if  $c \in C$ , let  $C_c$  be the set of clauses other than  $c$  which are consistent with  $c$ , i.e.,  $C_c = \{d \in C \setminus \{c\} : N(d) \text{ and } N(c) \text{ are consistent}\}$ . Let  $C_{\text{main}}$  be the set of clauses which are consistent with at least one clause of  $F$  other than themselves, i.e.,  $C_{\text{main}} = \{c \in C : C_c \neq \emptyset\}$ .*

*For each clause  $c \in C_{\text{main}}$ , define the new DNF formula  $F_c = (C_c, [n], N_c)$ , where  $N_c(d) = N(d) \cup N(c)$  for each  $d \in C_c$ . That is,  $F_c$  is the formula resulting from removing from  $F$  the clause  $c$  and all the clauses not consistent with  $c$ , and adding the literals of  $c$  to each of the remaining clauses. Then*

$$\text{zeroEnergy}(F; t) \leq |C_{\text{main}}|^2 \max_{c \in C_{\text{main}}} \text{energy}(\text{cover}_{F_c}; t - s).$$

**Proof:** (a) Let  $f \in L(B_n)$  be as defined in the statement of Lemma 7.1. Thus

$$\text{zeroEnergy}(F; t) \leq E(F - f)^2 = \|\widehat{f - F}\|_2^2$$

and

$$\widehat{(f - F)}(y) = \sum_{c \in C: |y \cup N(c)| > t} \widehat{\text{cover}_{F_c}}(y),$$

for each  $y \in B_n$ . To hide the dependency of the summation on  $y$ , for each  $c \in C$ , define  $a_c \in L(B_n)$  by

$$a_c(y) = \begin{cases} \widehat{\text{cover}_{F_c}}(y) & \text{if } |y \cup N(c)| > t \\ 0 & \text{otherwise.} \end{cases}$$

Thus  $\widehat{f - F} = \sum_{c \in C} a_c$ . Applying a triangular inequality, we get

$$\begin{aligned}
\|\widehat{f - F}\|_2 &\leq \sum_{c \in C} \|a_c\|_2 \\
&= \sum_{c \in C} \left( \sum_{y \in B_n: |y \cup N(c)| > t} \widehat{\text{cover}}_{F_c}(y)^2 \right)^{1/2} \\
&\leq \sum_{c \in C} \left( \sum_{y \in B_n: |y| > t-s} \widehat{\text{cover}}_{F_c}(y)^2 \right)^{1/2} \\
&= \sum_{c \in C} \sqrt{\text{energy}(\text{cover}_{F_c}; t-s)},
\end{aligned}$$

where the second inequality follows from the fact that  $|y \cup N(c)| \leq |y| + |N(c)| \leq |y| + s$  since  $F$  is an  $s$ -DNF. It follows that

$$\begin{aligned}
\text{zeroEnergy}(F; t) &\leq \left( \sum_{c \in C} \sqrt{\text{energy}(\text{cover}_{F_c}; t-s)} \right)^2 \\
&\leq |C|^2 \max_{c \in C} \text{energy}(\text{cover}_{F_c}; t-s).
\end{aligned}$$

(b) The general case follows from exactly the same argument. Just replace Lemma 7.1 with Lemma 7.2,  $y$  with  $w$ ,  $C$  with  $C_{\text{main}}$ , and  $N(c)$  with  $N'(c) \cup N''(c)$ . ■

It is not clear how to estimate the  $t$ -energy of the cover function without resorting to the  $u$ -skin function, for all  $u \geq 0$ .

**Lemma 7.4 (energy of cover  $\preceq$  energy of skin)** *Let  $G$  be a DNF formula and let  $t \geq 0$  be an integer. Then :*

a)

$$\text{energy}(\text{cover}_G; t) \leq \left( \int_0^\infty \sqrt{\text{energy}(\text{skin}_{G,u}; t)} e^{-u} du \right)^2.$$

b)

$$\text{energy}(\text{cover}_G; t) \leq \sup_{u \geq 0} \text{energy}(\text{skin}_{G,u}; t)$$

**Proof:** Part (b) follows immediately from Part (a) since  $(\int_0^\infty e^{-u} du)^2 = 1$ .

Part (a) follows from the fact that  $\text{cover}_G = \int_0^\infty \text{skin}_{G,u} e^{-u} du$  via Cauchy-Schwarz inequality as we explain next. Let  $a_u = \text{skin}_{G,u}$  and  $b = \int_0^\infty a_u e^u du$ . Thus  $\widehat{b} = \int_0^\infty \widehat{a}_u e^u du$  by the linearity of the Fourier transform operator. Hence

$$\widehat{b}(y)^2 = \int_0^\infty \int_0^\infty \widehat{a}_{u_1}(y) \widehat{a}_{u_2}(y) e^{-u_1 - u_2} du_1 du_2.$$

for all  $y \in \{0, 1\}^n$ . It follows from Cauchy-Schwarz inequality that

$$\begin{aligned}
\sum_{y:|y|>t} \widehat{b}(y)^2 &= \int_0^\infty \int_0^\infty \left( \sum_{y:|y|>t} \widehat{a}_{u_1}(y) \widehat{a}_{u_2}(y) \right) e^{-u_1-u_2} du_1 du_2 \\
&\leq \int_0^\infty \int_0^\infty \sqrt{\sum_{y:|y|>t} \widehat{a}_{u_1}(y)^2} \sqrt{\sum_{y:|y|>t} \widehat{a}_{u_2}(y)^2} e^{-u_1-u_2} du_1 du_2 \\
&= \left( \int_0^\infty \sqrt{\sum_{y:|y|>t} \widehat{a}_u(y)^2} e^{-u} du \right)^2 \\
&= \left( \int_0^\infty \sqrt{\text{energy}(a_u; t)} e^{-u} du \right)^2.
\end{aligned}$$

■

**Problem 7.5** Let  $G = (C, [n], N)$  be a DNF formula,  $t \geq 0$  an integer, and  $u \geq 0$ . If  $G$  is monotone, it follows from (6.10) and (6.9) that

$$\begin{aligned}
\text{energy}(\text{cover}_G; t) &= \sum_{y \in B_n: |y| > t} \left( \sum_{S \subset C: N(S) \geq y} (-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S| + 1} \right)^2 \\
\text{energy}(\text{skin}_{G,u}; t) &= \sum_{y \in B_n: |y| > t} \left( \sum_{S \subset C: N(S) \geq y} (-1)^{|S|} 2^{-|N(S)|} e^{-u|S|} \right)^2.
\end{aligned}$$

In general, it follows from (6.21) and (6.20) that

$$\begin{aligned}
\text{energy}(\text{cover}_G; t) &= \sum_{w \in B_n: |w| > t} \left( \sum_{a \leq w} (-1)^{|a|} \sum_{S \subset C: N'(S) \cap N''(S) = \emptyset \ \& \ N(S) \geq (a, w \setminus a)} (-1)^{|S|} 2^{-|N(S)|} \frac{1}{|S| + 1} \right)^2 \\
\text{energy}(\text{skin}_{G,u}; t) &= \sum_{w \in B_n: |w| > t} \left( \sum_{a \leq w} (-1)^{|a|} \sum_{S \subset C: N'(S) \cap N''(S) = \emptyset \ \& \ N(S) \geq (a, w \setminus a)} (-1)^{|S|} 2^{-|N(S)|} e^{-u|S|} \right)^2.
\end{aligned}$$

Analyze and estimate those sums without using the technique of Section 8 below. In the monotone case the sums are expressions associated with bipartite graphs. Note also that, experimentally, it is evident that  $\text{energy}(\text{skin}_{G,u}; t)$  exponentially decreases with  $u$  for fixed  $G$  and  $t$ .

## 8 Back to DNF formulas

Let  $G$  be a DNF formula,  $u \geq 0$ , and  $t \geq 0$  be an integer. We want to estimate the  $t$ -energy of the  $u$ -skin of  $G$ . In this section, we bound the  $t$ -energy of the  $u$ -skin of  $G$  by the  $t$ -energies of DNF formulas derived from  $G$  by adding auxiliary new variables.

To motivate the technique, assume for simplicity that  $G$  is monotone. Let  $G = (C, [n], N)$ . The key idea behind the reduction can be easily pointed out using the Fourier transform expression of the  $u$ -skin function derived in Section 6.3.

We have

$$\text{energy}(\text{skin}_{G,u}; t) = \sum_{y:|y|>t} \widehat{\text{skin}}_{G,u}(y)^2$$

with

$$\widehat{\text{skin}}_{G,u}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|} e^{-u|S|}$$

by (6.9). Recall also from (6.6) that

$$\widehat{G}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-|N(S)|}.$$

We have a bound on  $\text{energy}(G; t)$  from LMN energy bound (Theorem 5.8). Since  $G = \text{skin}_{G,0}$ , this is a bound on  $\text{energy}(\text{skin}_{G,u}; t)$  for  $u = 0$ . We need a bound for all  $u \geq 0$ .

The key observation is the following. Consider the special case when  $e^{-u} = 2^{-v}$ , where  $v$  is a nonnegative integer. Then

$$\widehat{\text{skin}}_{G,u}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N(S) \geq y} -(-1)^{|S|} 2^{-(|N(S)|+v|S|)}.$$

Construct from  $G$  a new monotone DNF formula  $G_v$  by adding  $v$  auxiliary new variables to each clause  $c \in C$ . That is, for each  $c \in C$ , let  $\ddot{N}(c)$  be a size- $v$  set of variable indices such that:  $\ddot{N}(c_1) \cap \ddot{N}(c_2) = \emptyset$  and  $\ddot{N}(c_1) \cap [n] = \emptyset$  for all  $c_1 \neq c_2 \in C$ . Let  $\ddot{I} = \cup_{c \in C} \ddot{N}(c)$ . Then  $G_v = (C, [n] \cup \ddot{I}, N_v)$ , where  $N_v(c) = N(c) \cup \ddot{N}(c)$  for each  $c \in C$ .

If  $S \subset C$ , then  $N_v(S) = N(S) \cup \ddot{N}(S)$  and hence  $|N_v(S)| = |N(S)| + v|S|$ . Moreover if  $y \subset [n]$ , then  $N(S) \geq y$  if and only if  $N_v(S) \geq y$  because  $\ddot{I}$  and  $[n]$  are disjoint. Thus

$$\widehat{\text{skin}}_{G,u}(y) = (-1)^{|y|} \sum_{S \neq \emptyset \subset C: N_v(S) \geq y} -(-1)^{|S|} 2^{-|N_v(S)|}, \quad \text{for all } y \subset [n].$$

This leads us to the *key Fourier relation*:

$$\widehat{\text{skin}}_{G,u}(y) = \widehat{G}_v(y), \quad \text{for all } y \subset [n], \quad (8.1)$$

which is the key point behind adding auxiliary new variables. It immediately implies that

$$\begin{aligned} \text{energy}(\text{skin}_{G,u}; t) &= \sum_{y \subset [n]: |y| > t} \widehat{\text{skin}}_{G,u}(y)^2 \\ &= \sum_{y \subset [n]: |y| > t} \widehat{G}_v(y)^2 \\ &\leq \sum_{y \subset [n] \cup \ddot{I}: |y| > t} \widehat{G}_v(y)^2 \\ &= \text{energy}(G_v; t), \end{aligned} \quad (8.2)$$

which enables us to use LMN energy bound. The same argument works if  $G$  is not necessarily monotone. The above argument assumes that  $v$  is an integer. If  $v$  is not necessarily an integer and  $G$  is not necessarily monotone, we prove the following:

**Theorem 8.1 (energy of skin  $\preceq$  energy)** *Let  $G = (C, [n], N)$  be a DNF formula and let  $t \geq 0$  be an integer.*

*If  $d \in \mathbb{N}^C$ , construct from  $G$  a new DNF formula  $G_d$  by adding  $d_c$  auxiliary new non-negated variables to each clause  $c \in C$ . That is, for each  $c \in C$ , let  $\dot{N}(c)$  be a size- $d_c$  set of variables indices such that:  $\dot{N}(c_1) \cap \dot{N}(c_2) = \emptyset$  and  $\dot{N}(c_1) \cap [n] = \emptyset$  for all  $c_1 \neq c_2 \in C$ . Let  $\dot{I} = \cup_{c \in C} \dot{N}(c)$ . Then  $G_d = (C, [n] \cup \dot{I}, N_d)$ , where  $N_d(c) = (N'(c) \cup \dot{N}(c), N''(c))$  for each  $c \in C$ .*

*Let  $u \geq 0$  and let  $v \geq 0$  such that  $e^{-u} = 2^{-v}$ , i.e.,  $v = u / \ln 2$ .*

*Then*

$$\text{energy}(\text{skin}_{G,u}; t) \leq \max_{d \in \{\lfloor v \rfloor, \lceil v \rceil\}^C} \text{energy}(G_d; t).$$

The underlying analogue of the key Fourier relation in (8.1) is the following. If  $v$  is not necessarily an integer, let  $0 \leq p \leq 1$  such that  $p2^{-\lceil v \rceil} + (1-p)2^{-\lfloor v \rfloor} = 2^{-v} = e^{-u}$ . We note in the poof below that  $\widehat{\text{skin}}_{G,u}(y) = E_D \widehat{G}_D(y)$ , for each  $y \subset [n]$ , where  $D$  is a random vector chosen from  $\{\lfloor v \rfloor, \lceil v \rceil\}^C$  by independently setting each of its entries to  $\lceil v \rceil$  with probability  $p$  and to  $\lfloor v \rfloor$  with probability  $1-p$ .

As noted above, the key point behind adding auxiliary new variables is (8.1), which can be easily seen by examining the summation in the expression of the Fourier transform of the  $u$ -skin function. We can directly verify (8.1) and its analog without going into this summation, but with little insight into what is going on. We do that below to avoid the messy summations in the nonnecessarily monotone case.

**Remark 8.2** 1. It is not clear how tight the bound is, i.e., it is not clear how much we are losing in (8.2).

2. Experimentally, it is evident that  $\text{energy}(\text{skin}_{G,u}; t)$  exponentially decreases with  $u$  for fixed  $G$  and  $t$ . We conjecture that there is a bound on  $\text{energy}(\text{skin}_{G,u}; t)$  in terms of  $u$ , the number  $m$  of clauses, and  $t$  which exponentially decreases with  $u$ .

Note that  $\text{skin}_{G,u}(x)$  is a strictly decreasing function in  $u$  for fixed  $G$  and  $x$ . But this fact alone is not enough to conclude anything about the variation of its energy  $\text{energy}(\text{skin}_{G,u}; t)$  with  $u$  for fixed  $G$  and  $t$ .

3. It is not clear whether the bound of Theorem 8.1 decays exponentially with  $u$  for fixed  $m$  and  $t$ . It is not hard to derive an exponentially decaying bound for  $t = 1$ . We were not able to do that for larger values of  $t$ .

4. It is worth mentioning that the  $u$ -skin of  $G$  does not simplify to a low degree polynomial under random restrictions, which excludes the possibility of directly adapting the argument in [LMN93] to the  $u$ -skin function without going into the the process of deriving the formulas  $G_d$  from  $G$ . The same holds for the cover function.

## 8.1 Proof of Theorem 8.1

We view  $G_d$  and  $\widehat{G}_d$  as functions defined on  $\{0, 1\}^n \times \{0, 1\}^{\dot{I}}$ . We denote the elements of  $\{0, 1\}^n \times \{0, 1\}^{\dot{I}}$  by  $(x, \ddot{x})$  or  $(y, \ddot{y})$ .

Let  $0 \leq p \leq 1$  such that  $p2^{-\lceil v \rceil} + (1-p)2^{-\lfloor v \rfloor} = 2^{-v} = e^{-u}$ . Such  $p$  exists since  $2^{-\lceil v \rceil} \leq 2^{-v} \leq 2^{-\lfloor v \rfloor}$ . Consider the random vector  $D = (D_c)_{c \in C} \in \{\lfloor v \rfloor, \lceil v \rceil\}^C$  whose entries are chosen independently by setting each to  $\lceil v \rceil$  with probability  $p$  and to  $\lfloor v \rfloor$  with probability  $1-p$ . Thus  $E_{D_c} 2^{-D_c} = 2^{-v} = e^{-u}$ , for each  $c \in C$ , by the definition of  $p$ .

We show below that

$$\widehat{\text{skin}}_{G,u}(y) = E_D \widehat{G}_D(y, 0) \tag{8.3}$$

for each  $y \in \{0, 1\}^n$ .

Theorem 8.1 follows from (8.3) as follows. We have

$$\widehat{\text{skin}}_{G,u}(y)^2 = (E_D \widehat{G}_D(y, 0))^2 \leq E_D \widehat{G}_D(y, 0)^2,$$

since  $0 \leq E(X - EX)^2 = EX^2 - (EX)^2$  for any random variable  $X$ . Thus

$$\begin{aligned} \text{energy}(\widehat{\text{skin}}_{G,u}; t) &= \sum_{y \in \{0,1\}^n: |y| > t} \widehat{\text{skin}}_{G,u}(y)^2 \\ &\leq E_D \sum_{y \in \{0,1\}^n: |y| > t} \widehat{G}_D(y, 0)^2 \\ &\leq E_D \sum_{(y, \ddot{y}) \in \{0,1\}^n \times \{0,1\}^{\dot{I}}: |(y, \ddot{y})| > t} \widehat{G}_D(y, \ddot{y})^2 \\ &= E_D \text{energy}(G_D; t) \\ &\leq \max_{d \in \{\lfloor v \rfloor, \lceil v \rceil\}^C} \text{energy}(G_d; t). \end{aligned}$$

To establish (8.3), we use an intermediate function. If  $d \in \mathbb{N}^C$ , define  $\text{shell}_{G,d} : \{0, 1\}^n \rightarrow \mathbb{R}$  as

$$\text{shell}_{G,d}(x) \stackrel{\text{def}}{=} 1 - \prod_{c \in C} (1 - 2^{-d_c} \text{AND}_{N(c)}(x)).$$

This is a nonuniform variation of the skin function where the clauses are weighted differently.

We show that

**Lemma 8.3** *If  $d \in \mathbb{N}^C$ , then*

$$\widehat{G}_d(y, 0) = \widehat{\text{shell}}_{G,d}(y)$$

*for each  $y \in \{0, 1\}^n$ .*

**Lemma 8.4** *For all  $u \geq 0$ ,*

$$E_D \text{shell}_{G,D} = \text{skin}_{G,u}.$$

Thus, by the linearity of the Fourier transform operator,

$$\begin{aligned}
\widehat{\text{skin}}_{G,u}(y) &= E_x \text{skin}_{G,u}(x) \mathcal{X}_y(x) \\
&= E_x E_D \text{shell}_{G,D}(x) \mathcal{X}_y(x) \\
&= E_D E_x \text{shell}_{G,D}(x) \mathcal{X}_y(x) \\
&= E_D \widehat{\text{shell}}_{G,D}(y) \\
&= E_D \widehat{G}_D(y, 0),
\end{aligned}$$

which verifies (8.3).

**Proof of Lemma 8.3:** We have

$$\widehat{G}_d(y, \ddot{y}) = E_{(x, \ddot{x})} G_d(x, \ddot{x}) \mathcal{X}_{(y, \ddot{y})}(x, \ddot{x}).$$

Thus

$$\begin{aligned}
\widehat{G}_d(y, 0) &= E_{(x, \ddot{x})} G_d(x, \ddot{x}) \mathcal{X}_{(y, 0)}(x, \ddot{x}) = E_{(x, \ddot{x})} G_d(x, \ddot{x}) \mathcal{X}_y(x) \\
&= E_{x \in \{0, 1\}^n} \left( E_{\ddot{x} \in \{0, 1\}^i} G_d(x, \ddot{x}) \right) \mathcal{X}_y(x).
\end{aligned} \tag{8.4}$$

First note that if  $c \in C$  and  $(x, \ddot{x}) \in \{0, 1\}^n \times \{0, 1\}^i$ , then

$$\text{AND}_{N_d(c)}(x, \ddot{x}) = \text{AND}_{N(c)}(x) \text{AND}_{\check{N}(c)}(\ddot{x}).$$

Thus

$$G_d(x, \ddot{x}) = \bigvee_{c \in C} \text{AND}_{N_d(c)}(x, \ddot{x}) = 1 - \prod_{c \in C} (1 - \text{AND}_{N(c)}(x) \text{AND}_{\check{N}(c)}(\ddot{x})).$$

Since each of the auxiliary new variables belongs to one and only one clause, by decomposing  $\ddot{x} \in \{0, 1\}^i$  as  $\ddot{x} = (\ddot{x}_c)_{c \in C} \in \prod_{c \in C} \{0, 1\}^{\check{N}(c)}$ , we get

$$\begin{aligned}
E_{\ddot{x} \in \{0, 1\}^i} G_d(x, \ddot{x}) &= 1 - \prod_{c \in C} \left( 1 - \text{AND}_{N(c)}(x) \left( E_{\ddot{x}_c \in \{0, 1\}^{\check{N}(c)}} \text{AND}_{\check{N}(c)}(\ddot{x}_c) \right) \right) \\
&= 1 - \prod_{c \in C} (1 - \text{AND}_{N(c)}(x) 2^{-d_c}) \\
&= \text{shell}_{G,d}(x).
\end{aligned}$$

Substituting in (8.4), we get

$$\widehat{G}_d(y, 0) = E_{x \in \{0, 1\}^n} \text{shell}_{G,d}(x) \mathcal{X}_y(x) = \widehat{\text{shell}}_{G,d}(y).$$

■

**Proof of Lemma 8.4:** Since, by construction, the entries of the random vector  $D$  are independent and the expected value of each is  $e^{-u}$ , we have

$$\begin{aligned} E_D \text{ shell}_{G,D}(x) &= 1 - E_D \prod_{c \in C} (1 - 2^{-D_c} \text{AND}_{N(c)}(x)) \\ &= 1 - \prod_{c \in C} (1 - (E_{D_c} 2^{-D_c}) \text{AND}_{N(c)}(x)) \\ &= 1 - \prod_{c \in C} (1 - e^{-u} \text{AND}_{N(c)}(x)). \end{aligned}$$

If  $u = 0$ , we get  $E_D \text{ shell}_{G,D}(x) = 1 - \prod_{c \in C} (1 - \text{AND}_{N(c)}(x)) = G(x) = \text{skin}_{G,0}(x)$  by the definition of the extension of  $\text{skin}_{G,u}$  to  $u = 0$ .

If  $u > 0$ , we have  $1 - e^{-u} \text{AND}_{N(c)}(x) = (1 - e^{-u})^{\text{AND}_{N(c)}(x)}$  for all  $c \in C$  and all  $x \in \{0, 1\}^n$  (if  $\text{AND}_{N(c)}(x) = 0$ , both terms are 1; if  $\text{AND}_{N(c)}(x) = 1$ , both terms are  $1 - e^{-u}$ ). Thus

$$\begin{aligned} E_D \text{ shell}_{G,D}(x) &= 1 - \prod_{c \in C} (1 - e^{-u})^{\text{AND}_{N(c)}(x)} \\ &= 1 - (1 - e^{-u})^{\sum_{c \in C} \text{AND}_{N(c)}(x)} \\ &= \text{skin}_{G,u}(x). \end{aligned}$$

It follows that  $E_D \text{ shell}_{G,D}(x) = \text{skin}_{G,u}(x)$ , for all  $u \geq 0$ . ■

## 9 A sharper bound

We derived in Section 5.11 an asymptotic version of Theorem 1.1 based on the LMN energy bound stated in Theorem 5.8. In this section, we drive the exact bound  $16m^{2.2}2^{-\sqrt{k}/10}$  of Theorem 1.1 using: 1) another form of the LMN energy bound (Theorem 9.1 below), and 2) Part (a) instead of Part (b) of Lemma 7.4.

If we know that  $F$  is an  $s$ -DNF formula, we can extract from [LMN93] the following bound which is tighter than that of Theorem 5.8 when  $s$  is not relatively large.

**Theorem 9.1** [LMN93] **(LMN energy bound for  $s$ -DNF)** *Let  $G$  be an  $m$ -clause  $s$ -DNF formula and  $t \geq 0$  be an integer, then  $\text{energy}(G; t) \leq 2e^{-t/(10es)}$  if  $t > 40es$ .*

**Proof:** It follows from Lemmas 5 and 6 in [LMN93] that if  $f : \{0, 1\}^n \rightarrow \{0, 1\}$  and  $pt > 8$ , then

$$\text{energy}(f; t) \leq 2Pr_\rho[\text{deg}(f_\rho) > pt/2],$$

where  $\rho$  is a random  $p$ -restriction. Corollary 1 in [LMN93] on Hastad's Switching Lemma asserts that if  $G$  is an  $s$ -DNF formula, then

$$Pr_\rho[\text{deg}(G_\rho) > k] < (5ps)^k,$$

where  $\rho$  is a random  $p$ -restriction. It follows that  $\text{energy}(G; t) \leq 2(5ps)^{pt/2}$  if  $pt > 8$ . Setting  $p = 1/(5es)$  to minimize  $(5ps)^{pt/2}$ , we get  $\text{energy}(G; t) \leq 2e^{-t/(10es)}$  if  $t > 40es$ . ■

First, we replace the bound of Theorem 9.1 in Theorem 8.1.

**Corollary 9.2 (energy of skin)** *Let  $G$  be an  $m$ -clause  $s$ -DNF formula,  $u \geq 0$ , and let  $t \geq 0$  be an integer. Then*

$$\text{energy}(\text{skin}_{G,u}; t) \leq 2 \exp\left(-\frac{t}{10e(s + \lceil u/\ln 2 \rceil)}\right)$$

if  $t > 40e(s + \lceil u/\ln 2 \rceil)$ .

**Proof:** It is enough to note that in the setting of Theorem 8.1, for each  $d \in \{\lceil v \rceil, \lceil v \rceil\}^m$ ,  $G_d$  is by construction an  $(s + \lceil v \rceil)$ -DNF whose number of clauses equals that of  $G$ . ■

**Remark 9.3** In Section 5.11, we obtained from Theorem 8.1 via Theorem 5.8 the bound  $\text{energy}(\text{skin}_{G,u}; t) \leq 2m2^{-\sqrt{t}/20}$ . This bound does not depend on  $u$ . In the above corollary, we used Theorem 9.1 to derive a bound on  $\text{energy}(\text{skin}_{G,u}; t)$  which depends on  $u$ . If  $u$  is less than some value, this bound is better than  $2m2^{-\sqrt{t}/20}$ . However, the bound increases with  $u$  for  $s$  and  $t$  fixed, contradicting the experimental behavior of  $\text{energy}(\text{skin}_{G,u}; t)$ . This stems from the fact that the special structure of  $G_d$  (the large number of auxiliary new variables) was not exploited when bounding  $\text{energy}(G_d; t)$ . Is it possible to exploit this structure to get a better bound? See also Remark 8.2 and Problem 7.5 for related open problems and improvement directions.

Now we substitute the bound of Corollary 9.2 in Part (a) of Lemma 7.4, which says that

$$\text{energy}(\text{cover}_G; t) \leq \left(\int_0^\infty \sqrt{\text{energy}(\text{skin}_{G,u}; t)} e^{-u} du\right)^2,$$

for each DNF formula  $G$  and each integer  $t \geq 0$ .

**Corollary 9.4 (energy of cover)** *Let  $G$  be an  $m$ -clause  $s$ -DNF formula and  $t \geq 0$  be an integer. Then*

$$\text{energy}(\text{cover}_G; t) \leq 6 \times 2^{-\left(\sqrt{t/(5e \ln 2) + (s+1)^2} - (s+1)\right)}$$

if  $\sqrt{t/(5e \ln 2) + (s+1)^2} - (s+1) > 4/\ln 2$ .

**Proof:** To bound  $\text{energy}(\text{cover}_G; t)$  in terms of  $s$ , we divide the integral into two parts  $\int_0^{u_0}$  and  $\int_{u_0}^\infty$ , where  $u_0 > 0$  is a parameter we optimize on. In the range  $0 < u \leq u_0$ , we use the bound of Corollary 9.2. For  $u > u_0$ , we use the trivial bound  $\text{energy}(\text{skin}_{G,u}; t) \leq \sum_y \widehat{\text{skin}}_{G,u}^2(y) = E[\text{skin}_{G,u}^2] \leq 1$  (we can do better than that for  $u > u_0$  but that will not significantly help). That is,

$$\begin{aligned} \sqrt{\text{energy}(\text{cover}_G; t)} &\leq \int_0^{u_0} \sqrt{\text{energy}(\text{skin}_{G,u}; t)} e^{-u} du + \int_{u_0}^\infty \sqrt{\text{energy}(\text{skin}_{G,u}; t)} e^{-u} du \\ &\leq \left(2 \exp\left(-\frac{t}{10e(s + \lceil u_0/\ln 2 \rceil)}\right)\right)^{1/2} \int_0^{u_0} e^{-u} du + \int_{u_0}^\infty e^{-u} du \\ &\leq \sqrt{2} \exp\left(-\frac{t}{20e(s + u_0/\ln 2 + 1)}\right) + e^{-u_0}, \end{aligned}$$

since  $\int_0^{u_0} e^{-u} du \leq 1$  and  $\int_{u_0}^{\infty} e^{-u} du = e^{-u_0}$ . This holds assuming that  $t > 40e(s + \lceil u_0/\ln 2 \rceil)$ , which is satisfied if  $t > 40e(s + u_0/\ln 2 + 1)$ . We use a suboptimal value of  $u_0$  to simplify the bound. Set  $u_0 \geq 0$  so that the two exponents are equal, i.e.,  $t = 20e(s + u_0/\ln 2 + 1)u_0$ . Solving the quadratic equation, we get  $2u_0/\ln 2 = \sqrt{t/(5e \ln 2) + (s+1)^2} - (s+1)$ . Since  $t = 20e(s + u_0/\ln 2 + 1)u_0$ , the condition on  $t$  is equivalent to  $u_0 > 2$ , i.e.,  $2u_0/\ln 2 > 4/\ln 2$ . Therefore  $\text{energy}(\text{cover}_G; t) \leq ((1 + \sqrt{2})e^{-u_0})^2 < 6e^{-2u_0} = 6 \times 2^{-2u_0/\ln 2}$ . ■

Then we substitute the bound of Corollary 9.4 in Theorem 7.3.

**Corollary 9.5 (zero-energy of  $s$ -DNF)** *Let  $F$  be an  $m$ -clause  $s$ -DNF formula and  $t \geq s$  be an integer. Then*

$$\text{zeroEnergy}(F; t) \leq 6m^2 2^{-\left(\sqrt{(t-s)/(5e \ln 2) + (2s+1)^2} - (2s+1)\right)} \quad (9.1)$$

if  $m \geq 4$ .

**Proof:** It is enough to note that, in the language of Part (b) of Theorem 7.3, for each  $c \in C_{\text{main}}$ ,  $F_c$  is a  $2s$ -DNF formula with at most  $m - 1$  clauses and least one clause (by the definition of  $C_{\text{main}}$ ). Moreover,  $|C_{\text{main}}| \leq |C| = m$ . Theorem 7.3 implies (9.1) subject the condition  $\sqrt{(t-s)/(5e \ln 2) + (2s+1)^2} - (2s+1) > 4/\ln 2$ . If this condition is not satisfied then the upper bound in (9.1) is  $\geq 6m^2 2^{-4/\ln 2} = 6m^2 e^{-4} > 1$  if  $m \geq 4$ . That is, under the assumption  $m \geq 4$ , the upper bound in (9.1) is trivial when the condition is not satisfied. ■

Substituting the bound of Corollary 9.5 in Lemma 5.5 for  $t = \lfloor \frac{k-s}{2} \rfloor$ , we obtain:

**Corollary 9.6 (bias of  $s$ -DNF)** *Let  $F$  be an  $m$ -clause  $s$ -DNF formula and  $k \geq 3s$  be an integer, then*

$$\text{bias}(F; k) \leq 6m^3 2^{-\left(\sqrt{(k-3s-1)/(10e \ln 2) + (2s+1)^2} - (2s+1)\right)}.$$

**Proof:** The condition  $t = \lfloor \frac{k-s}{2} \rfloor \geq s$  is equivalent to  $k \geq 3s$ . To simplify the exponent, we used the bound  $t = \lfloor \frac{k-s}{2} \rfloor \geq \frac{k-s}{2} - \frac{1}{2}$ , hence  $t-s \geq \frac{k-3s-1}{2}$ . Finally, we dropped the condition  $m \geq 4$  since if  $m < 4$ , then  $F$  has at most  $ms \leq 3s$  variables, in which case the condition  $k \geq 3s$  implies that  $\text{bias}(F; k) = 0$ . ■

Finally, substituting the bound in Corollary 9.6 in Lemma 5.4 and optimizing on  $s$ , we conclude the the proof of Theorem 1.1. We set below  $s = \Theta(\sqrt{k})$  if  $k = \Omega(\log^2 m)$ .

**Corollary 9.7 (bias of DNF)** *Let  $F$  be an  $m$ -clause DNF formula and  $k \geq 0$  be an integer, then*

$$\text{bias}(F; k) \leq 16m^{2.2} 2^{-\sqrt{k}/10}.$$

**Proof:** Let  $\epsilon = \text{bias}(F; k)$ . Substituting the bound of Corollary 9.6 in Lemma 5.4, we get that for each integer  $s \geq 1$  such that  $k \geq 3s$ , we have

$$\begin{aligned} \epsilon &\leq 6m^3 2^{-\left(\sqrt{(k-3s-1)/(10e \ln 2) + (2s+1)^2} - (2s+1)\right)} + m2^{-s} \\ &= m \left( 2^{-\left(\sqrt{(k-3s-1)/(10e \ln 2) + (2s+1)^2} - (2s+1) - \log(6m^2)\right)} + 2^{-s} \right). \end{aligned}$$

Allowing  $s$  to take noninteger values, we obtain

$$\begin{aligned}\epsilon &\leq m \left( 2^{-\left(\sqrt{(k-3\lfloor s \rfloor - 1)/(10e \ln 2) + (2\lfloor s \rfloor + 1)^2 - (2\lfloor s \rfloor + 1) - \log(6m^2)}\right)} + 2^{-\lfloor s \rfloor} \right) \\ &\leq m \left( 2^{-\left(\sqrt{(k-3s-1)/(10e \ln 2) + (2s-1)^2 - (2s+1) - \log(6m^2)}\right)} + 2^{-(s-1)} \right),\end{aligned}\quad (9.2)$$

for each real number  $s \geq 1$  such that  $k \geq 3s$ . We will equate the two exponents of (9.2) and solve for  $s$ . If  $s$  is a real number such that the two exponents are equal, i.e.,

$$\sqrt{(k-3s-1)/(10e \ln 2) + (2s-1)^2 - (2s+1) - \log(6m^2)} = s-1, \quad (9.3)$$

then  $\epsilon \leq 2m2^{-(s-1)}$ . Note that we ignored the conditions on  $s$ :  $s \geq 1$  and  $k \geq 3s$ . We can do that since if  $s < 1$ , then  $m2^{-(s-1)} > m \geq 1$ , and hence the RHS of (9.2) a trivial bound on  $\epsilon$  since  $\epsilon \leq 1$ . Similarly, if  $k < 3s$  and  $s \geq 1$ , then  $m2^{-\left(\sqrt{(k-3s-1)/(10e \ln 2) + (2s-1)^2 - (2s+1) - \log(6m^2)}\right)} > m \geq 1$ , which again makes the RHS of (9.2) is a trivial bound on  $\epsilon$ . We verify below that

$$s = \sqrt{\frac{k}{M} + \alpha \log^2(6m^2) + \beta \log(6m) + \gamma - a \log(6m^2) - b}, \quad (9.4)$$

is a solution of (9.3), where  $M \approx 94.208$ ,  $\alpha = 0.16$ ,  $\beta \approx 0.499$ ,  $\gamma \approx 0.362$ ,  $a = 2.2$ , and  $b \approx 0.416$ . It follows that

$$\begin{aligned}\epsilon &\leq 2m2^{-(s-1)} = 2^2 m 2^{-\sqrt{k/M + \alpha \log^2(6m^2) + \beta \log(6m^2) + \gamma + a \log(6m^2) + b}} \\ &< 2^2 m 2^{-\sqrt{k/M + a \log(6m^2) + b}} = (2^{2+b} 6^a) m^{1+2a} 2^{-\sqrt{k/M}} \\ &< 16m^{2.2} 2^{-\sqrt{k}/10}.\end{aligned}$$

To verify that that (9.4) is a solution of of (9.3), write (9.3) as

$$\begin{aligned}0 &= (3s + \log(6m^2))^2 - (2s-1)^2 - \frac{k-3s-1}{10e \ln 2} \\ &= 5s^2 + \left(6 \log(6m^2) + 4 + \frac{3}{10e \ln 2}\right) s + \log^2(6m^2) - 1 + \frac{1}{10e \ln 2} - \frac{k}{10e \ln 2} \\ &= 5 \left( s^2 + 2(a \log(6m^2) + b)s + c \log^2(6m^2) - d - \frac{k}{M} \right),\end{aligned}$$

where  $a = 0.6$ ,  $b = 0.4 + \frac{3}{100e \ln 2} \approx 0.416$ ,  $c = 0.2$ ,  $d = 0.2 - \frac{1}{50e \ln 2} \approx 0.189$ , and  $M = 50e \ln 2 \approx 94.208$ . The larger solution is

$$\begin{aligned}s &= \sqrt{\frac{k}{M} + (a \log(6m^2) + b)^2 - c \log^2(6m^2) + d - a \log(6m^2) - b} \\ &= \sqrt{\frac{k}{M} + \alpha \log^2(6m^2) + \beta \log(6m^2) + \gamma - a \log(6m^2) - b},\end{aligned}$$

where  $\alpha = a^2 - c = 0.16$ ,  $\beta = 2ab \approx 0.499$ , and  $\gamma = b^2 + d \approx 0.362$ . ■

## 10 Optimal solution

The proof of Theorem 1.1 does not depend on this section since the former is based on the suboptimal solution constructed in Section 7.

Let  $F = (C, [n], N)$  be a DNF formula and  $t \geq 0$ . Recall that  $\text{zeroEnergy}(F; t)$  is the minimum value of  $E(F - f)^2$  over the choice of  $f \in L(B_n)$  such that:  $\text{deg}(f) \leq t$ , and  $f$  satisfies the  $F$ -zeros-constraint:  $f(x) = 0$  for each  $x \in B_n$  such that  $F(x) = 0$ .

In this section, we derive a compact form of the optimal solution of the least square problem underlying the definition of  $\text{zeroEnergy}(F; t)$ . For simplicity, we restrict our attention to the case when  $F$  is a monotone DNF formula. The optimal solution can be characterized in terms of the zeta function of the dual order ideal  $P_F$  of  $B_n$  consisting of satisfying assignments of  $F$ . Unable to estimate the optimal solution, we leave the problem open.

Recall first the posets terminology in Section 6.1. We need the following additional elementary poset notions. An *antichain* of a poset  $X$  is a subset  $A$  of  $X$  such that any two distinct elements of  $A$  are incomparable. A subset  $I$  of  $X$  is called a *dual order ideal* if  $x \in I$  and  $y \geq x$ , then  $y \in I$ . We say that a dual order ideal is *generated by* a subset  $A$  of  $X$  if  $I = \{x \in X : x \geq y \text{ for some } y \in A\}$ . Any dual order ideal has a generating antichain.

Let  $F = (C, [n], N)$  be a monotone DNF. We can associate with  $F$  the subposet  $P_F$  of  $B_n$  consisting of the satisfying assignments of  $F$ , i.e.,  $P_F = \{x \in B_n : F(x) = 1\}$ . Let  $A_F$  be the set of clauses of  $F$  regarded as subsets of  $[n]$ , i.e.,  $A_F = \{N(c) : c \in C\} \subset B_n$ . Equivalently,  $P_F$  is the dual order ideal of  $B_n$  generated by  $A_F$ . We call a dual order ideal of  $B_n$  *nontrivial* if it is not the empty ideal or  $B_n$  itself. Recall that we assumed in the definition of a DNF formula that it contains at least one clause and no empty clauses to avoid degenerate cases. Thus  $P_F$  is a nontrivial dual order ideal of  $B_n$ . Conversely, to each nontrivial dual order ideal  $P$  of  $B_n$  and to each set of generator  $A$  of  $P$ , we can associate a monotone DNF formula  $F$  such that  $A_F = A$  and  $P_F = P$ . The formula  $F$  is unique up to duplicate clauses. Note also that  $A_F$  is an antichain if and only if no clause of  $F$  can be removed without changing the boolean function computed by  $F$ .

A key remark is the following.

**Lemma 10.1** *Let  $F$  be a monotone DNF formula on  $n$  variables. If  $f \in L(B_n)$ , then  $f$  satisfies the  $F$ -zeros-constraint if and only if  $f$  is a linear combination of  $\{AND_z\}_{z \in P_F}$ .*

**Proof:** Let  $Z_F = B_n \setminus P_F = \{x \in B_n : F(x) = 0\}$ . Thus the  $F$ -zeros-constraint on  $f$  is:  $f|_{Z_F} = 0$ . The if part follows from the fact that, by the definitions of  $Z_F$  and  $P_F$ ,  $AND_z|_{Z_F} = 0$  for all  $z \in P_F$ . One way to demonstrate the only if part is to note that, since  $\{AND_z\}_{z \in B_n}$  are linearly independent,  $\dim \text{span}\{AND_z\}_{z \in P_F} = |P_F| = \dim\{f \in L(P_F) : f|_{Z_F} = 0\}$ . ■

We cast the zero-energy problem in the language of zeta functions of dual order ideals.

**Definition 10.2** *Say that  $P$  is a nontrivial dual order ideal of  $B_n$ , and let  $t \geq 0$  be an integer. Let  $P_t = \{z \in P : |z| \leq t\}$  and define the projection map  $\pi_t : L(P) \rightarrow L(P_t)$ ,  $f \mapsto f|_{P_t}$ , and its transpose  $\pi_t^T : L(P_t) \rightarrow L(P)$ , the extension by zeros map. Consider the zeta function  $\zeta_P$  of  $P$  as a linear transformation  $L(P) \rightarrow L(P)$ , and consider the linear*

transformation  $\zeta_P \pi_t^T : L(P_t) \rightarrow L(P)$ . Define

$$\Delta_t(P) \stackrel{\text{def}}{=} \min_{g \in L(P_t)} \|1_P - \zeta_P \pi_t^T g\|_2^2,$$

where  $1_P \in L(P)$  is the all ones function and  $\|\cdot\|_2$  the  $L_2$ -norm on  $L(P)$ . Note that if  $P_t = \emptyset$ , by convention,  $L(P_t)$  consists of the zero function.

That is,  $\Delta_t(P)$  is the least square  $L_2$ -approximation error resulting from approximating the all ones function on  $P$  by  $\zeta_P \pi_t^T g$  over the choice of  $g \in L(P_t)$ .

**Lemma 10.3** *Let  $F$  be a monotone DNF formula on  $n$  variables and  $t \geq 0$  be an integer, then  $2^n \text{zeroEnergy}(F; t) = \Delta_t(P_F)$ .*

**Proof:** Let  $P = P_F$ ,  $Z = B_n \setminus P = \{x \in B_n : F(x) = 0\}$ , and  $V_t = \{f \in L(B_n) : f|_Z = 0 \text{ and } \deg(f) \leq t\}$ . Thus

$$2^n \text{zeroEnergy}(F; t) = \min_{f \in V_t} 2^n E(F - f)^2 = \min_{f \in V_t} \|1_P - f|_P\|_2^2,$$

since  $F|_Z = f|_Z = 0$  and  $F|_P = 1_P$ . The lemma then follows from the key remark in Lemma 10.1, which says that if  $f \in L(B_n)$ , then  $f|_Z = 0$  if and only if there exists  $g \in L(P)$  such that  $f = \sum_{z \in P} g(z) \text{AND}_z$ . Note that: 1)  $\deg(f) \leq t$  if and only if  $g \in \pi_t^T L(P_t)$ , and 2)  $f = \sum_{z \in P} g(z) \text{AND}_z$  can be expressed as  $f = \pi_P^T \zeta_P g$ , where  $\pi_P^T : L(P) \rightarrow L(B_n)$  is the extension by zeros map (the transpose of the projection map  $\pi_P : L(B_n) \rightarrow L(P)$ ,  $f \mapsto f|_P$ ).  $\blacksquare$

**Lemma 10.4** *Let  $P$  be a nontrivial dual order of ideal of  $B_n$ , and let  $t \geq 0$  be an integer such that  $P_t \neq \emptyset$ . Let  $v = \pi_t \zeta_P^T 1_P$ , and let  $M = \pi_t \zeta_P^T \zeta_P \pi_t^T$ , i.e.,  $M$  is the  $P_t$ -truncation of the matrix  $\zeta_P^T \zeta_P$ . Then  $M$  is invertible and*

$$\Delta_t(P) = |P| - v^T M^{-1} v. \quad (10.1)$$

Moreover,

$$v = 2^n (2^{-|x|})_{x \in P_t}, \quad (10.2)$$

$$M = 2^n (2^{-|x \cup y|})_{x, y \in P_t}. \quad (10.3)$$

We can also express  $\Delta_t(P)$  as follows. Let

$$M^* = \begin{bmatrix} 1 & v^T \\ v & M \end{bmatrix},$$

and let  $D$  be the value which, when added to the  $(\emptyset, \emptyset)$ -entry of the matrix  $M^*$ , makes it singular, then

$$\Delta_t(P) = |P| - D - 1 \quad (10.4)$$

$$= |P| + \frac{\det(M^*)}{\det(M)} - 1. \quad (10.5)$$

**Proof:** We have a least square problem of the form  $\min_g \|b - Ag\|_2^2$ , where  $b = 1_P$  and  $A = \zeta_P \pi_t^T$ . The matrix  $A$  has full column rank since  $\zeta_P$  is nonsingular. The optimal solution is  $\|b - Ag^*\|_2^2 = b^T b - (A^T b)^T g^*$  where  $A^T A g^* = A^T b$ . Since  $A$  has full columns rank, the matrix  $A^T A$  is invertible. In our case, we have  $b^T b = 1_P^T 1_P = |P|$ ,  $A^T b = \pi_t \zeta_P^T 1_P = v$ , and  $A^T A = \pi_t \zeta_P^T \zeta_P \pi_t^T = M$ . This proves (10.1).

To verify (10.2), let  $x \in P$ . We have  $(\zeta_P^T 1_P)(x) = \sum_{y \in P: x \leq y} 1 = \sum_{y \in B_n: x \leq y} 1$  since  $x \in P$  and  $P$  is a dual order ideal of  $B_n$ . Thus  $(\zeta_P^T 1_P)(x) = 2^{[n] \setminus |x|} = 2^n 2^{-|x|}$ . Then (10.2) follows from restricting  $x$  to  $P_t$ . To verify (10.3), let  $f \in L(P)$  and  $x \in P$ . We have

$$(\zeta_P^T \zeta_P f)(x) = \sum_{z \in P: z \geq x} \sum_{y \in P: y \leq z} f(y) = \sum_{y \in P} f(y) \sum_{z \in P: z \geq x, y} 1.$$

Since  $x \in P$  and  $P$  is a dual order ideal, we have

$$\sum_{z \in P: z \geq x, y} 1 = \sum_{z \in B_n: z \geq x, y} 1 = \sum_{z \in B_n: z \geq x \cup y} 1 = 2^{[n] \setminus (x \cup y)} = 2^n 2^{-|x \cup y|}.$$

Hence  $(\zeta_P^T \zeta_P f)(x) = 2^n \sum_{y \in P} f(y) 2^{-|x \cup y|}$ . Then (10.3) follows by restricting  $f$  to  $L(P_t)$  and  $x$  to  $P_t$ .

To verify (10.4), write (10.1) as  $\Delta_t(P) = |P| - v^T g^*$ , where  $M g^* = v$ . In matrix form, we can express this system as

$$\begin{bmatrix} 1 + D & v^T \\ v & M \end{bmatrix} \begin{bmatrix} 1 \\ -g^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where  $D = |P| - 1 - \Delta_t(P)$ . Then (10.4) follows from the fact that the  $\emptyset$ -entry of any vector in the null space of the perturbation of  $M^*$  by  $D$  must be nonzero since  $M$  is nonsingular.

Finally, (10.5) follows from (10.4). In general if  $M$  is a  $p \times p$  matrix,  $M^* = \begin{bmatrix} a & * \\ * & M \end{bmatrix}$  is a  $(p+1) \times (p+1)$  augmentation of  $M$ , and  $M^{*'} = \begin{bmatrix} a + D & * \\ * & M \end{bmatrix}$  is a perturbation of  $M^{*'}$ , then  $\det(M^{*'}) = \det(M^*) + D \det(M)$ . Thus when  $M^{*'}$  is singular and  $M$  is nonsingular, we get  $D = -\det(M^*)/\det(M)$ . ■

**Problem 10.5** *Let  $P$  be a nontrivial dual order ideal of  $B_n$  generated by  $m$  elements of  $B_n$  and let  $t \geq 0$ . Study and bound  $\Delta_t(P)$  in terms of  $m$  and  $t$  starting from the characterization in Lemma 10.4.*

We leave this problem open. We can conclude the following bounds from Lemma 10.3 and Corollary 9.5.

**Corollary 10.6** *Let  $P$  be a nontrivial dual order of ideal of  $B_n$  generated by  $m$  elements of  $B_n$  each of size at most  $s$ , and let  $t \geq s$  be an integer. Then we have the following bound:*

$$2^{-n} \Delta_t(P) \leq 6m^2 2^{-\left(\sqrt{(t-s)/(5e \ln 2) + (2s+1)^2} - (2s+1)\right)}$$

if  $m \geq 4$ .

# 11 Concluding remarks

After the results of this paper were described in a preliminary form [Baz07], Razborov [Raz08] has obtained a simpler construction of a function satisfying the zeros-constraints leading to a simpler proof of an asymptotic version of our main result in Theorem 1.1. The construction of Razborov is randomized and it simplifies the second step of the proof, which reduces the  $s$ -DNF constrained  $L_2$ -approximation problem to the  $s$ -DNF  $L_2$ -approximation problem.

We conclude with some problems.

The bound in Theorem 1.1 probably can be improved by studying the problems in Remarks 8.2 and 9.3 and Problems 7.5 and 10.5.

Is it possible to somehow generalize the argument of Theorem 1.1 from depth-2 circuits to  $AC_0$  circuits, i.e., to show that  $\log^{O(d)} n$ -wise independence  $o(1)$ -fools polynomial-size depth- $d$  circuits? A different approach toward proving this is the low degree polynomial predictors approach in [Baz03] (Section 5.7).

One of the basic questions motivating the work reported in this paper is the quadratic residues PRG introduced in [AGHP92]. Let  $p$  be an odd prime and denote by  $\mathbb{F}_p$  the finite field of size  $p$ . Fix a subset <sup>6</sup>  $I \subset \mathbb{F}_p$  of size  $n \geq 1$ . The *quadratic residues PRG (QR-PRG)* is given by  $G_p^I : \mathbb{F}_p \rightarrow \{0, 1\}^I$ , where for each  $t \in I$ ,  $G_p^I(a)_t = 1$  if  $a + t$  is a quadratic residue and 0 otherwise.

The irregularity of the quadratic residues distribution promises great derandomization capabilities and has intrigued mathematicians long before complexity theory existed. The following conjecture was the motivation behind the work reported in this paper.

**Conjecture 11.1** *For all positive integers  $m, n$  and every  $\epsilon > 0$ , there is an integer  $p_0 = \text{poly}(m, n, \frac{1}{\epsilon})$  such that if  $p \geq p_0$  is a prime, and  $I \subset \mathbb{F}_p$  is of size  $n$ , then the QR-PRG  $G_p^I$   $\epsilon$ -fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables.*

The QR-PRG was introduced in [AGHP92] as a  $\frac{n}{\sqrt{p}}$ -biased probability distribution. This follows from Weil's theorem on the analog of the Riemann Hypothesis for curves over finite fields. Using the  $\frac{n}{\sqrt{p}}$ -bias property of the QR-PRG, we obtain from Corollary 2.3 the following quasi-polynomial version.

**Corollary 11.2** *For all positive integers  $m, n$  and every  $\epsilon > 0$ , there is an integer  $p_0 = 2^{O(\log^2 \frac{m}{\epsilon} \log n)}$  such that if  $p \geq p_0$  is a prime, and  $I \subset \mathbb{F}_p$  is of size  $n$ , then the QR-PRG  $G_p^I$   $\epsilon$ -fools any boolean function computable by an  $m$ -clause DNF (or CNF) formula on  $n$  variables.*

The conjecture would imply the first (unconditional) polynomial complexity PRG for depth-2 circuits. Note that there are no reasons not to believe that the derandomization capabilities of the QR-PRG are far beyond the small bias property. Conjecture 11.1 is a natural starting point. On the other extreme, can one construct an infinite family of

---

<sup>6</sup>In [AGHP92],  $I = \{0, 1, \dots, n-1\}$  but the authors's analysis does not use this restriction.

(unrestricted) circuits  $\{C_n\}_n$ , where  $C_n$  is a polynomial-size circuit on  $n$  variables, such that the prime cannot be made polynomially large enough in  $n$  and  $\frac{1}{\epsilon}$  in order for the QR-PRG to  $\epsilon$ -fool  $C_n$ ?

## Acknowledgments

The author would like to thank Sanjoy Mitter, Daniel Spielman, and Madhu Sudan for very helpful discussions on this material, Widad Machmouchi for valuable comments on the first and second drafts of the paper, and the anonymous referees for valuable comments which significantly improved the presentation of the paper.

## References

- [ABFR94] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The Expressive Power of Voting Polynomials. *Combinatorica*, 14(2): 135-148, 1994.
- [AGHP92] N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple Constructions of Almost  $k$ -wise Independent Random Variables. *Random Structures and Algorithms*, 3(3):289-304, 1992.
- [AGM02] N. Alon, O. Goldreich, Y. Mansour. Almost  $k$ -wise independence versus  $k$ -wise independence. In *Electronic Colloquium on Computational Complexity*, Report No. 48, 2002.
- [AW85] M. Ajtai and A. Wigderson. Deterministic Simulation of Probabilistic Constant Depth Circuits. In *Proc. 26th IEEE Symposium on Foundations of Computer Science*, pages 11-19, 1985.
- [Baz03] Louay Bazzi. Minimum Distance of Error Correcting Codes versus Encoding Complexity, Symmetry, and Pseudorandomness. Ph.D. dissertation, MIT, Cambridge, Mass., 2003.
- [Baz07] Louay Bazzi. Polylogarithmic independence can fool DNF formulas. In *Proc. of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 63-73, 2007.
- [BRS91] R. Beigel, N. Reingold, and D. Spielman, The Perceptron Strikes Back. In *Proc. 6th Annual IEEE Conference on Structure in Complexity Theory*, pages 286-291, 1991.
- [BM82] M. Blum and S. Micali. How to generate Cryptographically Strong Sequences of Pseudo-Random Bits. *SIAM journal on Computing*, 13(4):850-864, 1984.
- [Has86] Johan Hastad. Computational Limitations for Small Depth Circuits. Ph.D. dissertation, MIT, Cambridge, Mass., 1986.

- [IW97] R. Impagliazzo and A. Wigderson. P = BPP if E Requires Exponential Circuits: Derandomizing the XOR Lemma. In *Proc. 29th Annual ACM Symposium on the Theory of Computing*, pages 220-229, 1997.
- [KKL88] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proc. of the 29th Annual Symposium on Foundations of Computer Science*, pages 68-80, 1988.
- [Lec71] Robert J. Lechner. Harmonic Analysis of Switching Functions. In *Recent Development in Switching Theory*, pages 122-229. Academic Press, 1971.
- [Lub85] Michael Luby. A simple parallel algorithm for the maximal independent set problem. In *Proc. 17th Annual ACM Symposium on the Theory of Computing*, pages 1-10, 1985.
- [LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the Association for Computing Machinery*, 40(3):607-620, 1993.
- [LN90] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349-365, 1990.
- [LV96] M. Luby and B. Velickovic, On Deterministic Approximation of DNF. *Algorithmica*, 16(4/5):415-433, 1996.
- [LVW93] M. Luby, B. Velickovic, and A. Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd ISTCS*, pages 18-24, 1993.
- [Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 12(4):63-70, 1991.
- [NN93] J. Naor and M. Naor. Small bias probability spaces: efficient constructions and applications. *SIAM J. on Computing*, 22(4):838-856, 1993.
- [NS94] N. Nisan and M. Szegey. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301-313, 1994,
- [NW88] N. Nisan and A. Wigderson. Hardness vs. Randomness. In *Proc. 29th IEEE Symposium on Foundations of Computer Science*, pages 2-11, 1988.
- [Raz87] Alexander Razborov. Lower bounds on the size of bounded depth networks over a complete basis with logical addition. *Mathematicheskije Zametki*, 41(4):598-607, 1987.
- [Raz08] Alexander Razborov. A simple proof of Bazzi's theorem. *Electronic Colloquium on Computational Complexity*. Report TR08-081, 2008.
- [Sta97] Richard P. Stanley. *Enumerative Combinatorics, Volume I*. Cambridge University Press, 1997.

- [Tre04] Luca Trevisan. A Note on Deterministic Approximate Counting for  $k$ -DNF. In *Proc. of APPROX-RANDOM*, pages 417-426, 2004.
- [Vaz86] Umesh Vazirani. Randomness, adversaries, and computation. Ph.D. dissertation, University of California, Berkeley, 1986.
- [Yao82] Andrew C. Yao. Theory and application of Trapdoor functions. In *Proc. 23rd IEEE Annual Symposium on Foundations of Computer Science*, pages 80-91, 1982.

## APPENDIX

### A LP duality calculations appendix

In this appendix we show the LP duality calculations needed to characterize the class of functions that are fooled by the  $(\delta, k)$ -bias property. The characterization is in Theorem A.1 below and is in terms of  $L_1$ -approximability by sandwiching polynomials of degree at most  $k$  and small  $L_1$ -norm in the Fourier domain.

Recall that we stated in Theorem 4.2 of Section 4 the special case of Theorem A.1 corresponding to the  $k$ -wise independence property, i.e., when  $\delta = 0$ .

Let  $\mu$  be a probability distribution on  $\{0, 1\}^n$ ,  $k \geq 0$  an integer, and  $\delta \geq 0$ . By definition  $\mu$  has the  $(\delta, k)$ -bias property if  $\mu$   $\delta$ -fools all parity functions on  $k$  or fewer of the  $n$  binary variables. In terms of the characters  $\{\mathcal{X}_y\}_y$ , this is equivalent to saying that  $|E_\mu \mathcal{X}_y| \leq 2\delta$  for each nonzero  $y$  in  $\{0, 1\}^n$  whose weight is less than or equal to  $k$ .

**Theorem A.1** *Let  $g : \{0, 1\}^n \rightarrow \{0, 1\}$ ,  $k \geq 0$  an integer, and  $\delta, \epsilon \geq 0$ . Then the  $(\delta, k)$ -bias property  $\epsilon$ -fools  $g$  if and only if there exist  $g_l, g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  such that:*

- i)  $\deg(g_l) \leq k$  and  $\deg(g_u) \leq k$*
- ii)  $g_l \leq g \leq g_u$*
- iii)  $2\delta \sum_{y \neq 0} |\widehat{g}_l(y)| + E(g - g_l) \leq \epsilon$  and  $2\delta \sum_{y \neq 0} |\widehat{g}_u(y)| + E(g_u - g) \leq \epsilon$ , where the expectation is over the uniform probability distribution.*

*Therefore, asymptotically and for  $\delta > 0$ , the  $(\delta, k)$ -bias property  $o(\epsilon)$ -fools a boolean function  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  if and only if there exist  $g_l, g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  such that:*

- **(low degree)**  $\deg(g_l) \leq k$  and  $\deg(g_u) \leq k$
- **(sandwiching polynomials)**  $g_l \leq g \leq g_u$
- **(small  $L_1$ -norm in the Fourier domain)**  $\|\widehat{g}_l\|_1 = o\left(\frac{\epsilon}{\delta}\right)$  and  $\|\widehat{g}_u\|_1 = o\left(\frac{\epsilon}{\delta}\right)$
- **(small  $L_1$ -approximation error)**  $E(g_u - g_l) = o(\epsilon)$ .

**Proof:** The proof is by linear-programming duality. Let  $M_k \subset \mathbb{R}^{\{0,1\}^n}$  be the convex polytope of  $(\delta, k)$ -biased probability distributions  $\mu$  on  $\{0, 1\}^n$ .

If  $\mu$  is a probability distribution  $\mu$  on  $\{0, 1\}^n$ , then by definition  $\mu$  is  $(\delta, k)$ -biased if  $|E_\mu \mathcal{X}_y| \leq 2\delta$  for each nonzero  $y$  in  $\{0, 1\}^n$  whose weight is less than or equal to  $k$ .

Thus  $M_k$  consists of all  $\mu : \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $\mu \geq 0$ ,  $\sum_x \mu(x) = 1$ , and  $-2\delta \leq \sum_x \mu(x) \mathcal{X}_y(x) \leq 2\delta$  for each  $y \in N_k^*$ , where  $N_k^* = \{y \in \{0, 1\}^n : y \neq 0 \text{ and } |y| \leq k\}$ .

Fix  $g : \{0, 1\}^n \rightarrow \{0, 1\}$  and note that if  $\mu$  is a probability distribution on  $\{0, 1\}^n$ , then  $Pr_{x \sim \mu}[g(x) = 1] = E_\mu g$  since  $g$  takes binary values.

We have two feasible linear programs:

$$P_u = \max_{\mu \in M_k} E_\mu g - Eg \text{ and } P_l = \max_{\mu \in M_k} Eg - E_\mu g.$$

It is enough to show that the dual linear programs are:

- I)  $P_u = \min_{g_u} E(g_u - g) + 2\delta \sum_{y \neq 0} |\widehat{g}_u(y)|$ , where we are minimizing over all  $g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $deg(g_u) \leq k$  and  $g_u(x) \geq g(x)$  for all  $x \in \{0, 1\}^n$ .
- II)  $P_l = \min_{g_l} E(g - g_l) + 2\delta \sum_{y \neq 0} |\widehat{g}_l(y)|$ , where we are minimizing over all  $g_l : \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $deg(g_l) \leq k$  and  $g_l(x) \leq g(x)$  for all  $x \in \{0, 1\}^n$ .

Actually, we have to establish only (I) since (II) follows from (I) by replacing  $g$  with  $1 - g$  and a performing a change of variable from  $g_u$  to  $1 - g_u$ .

Explicitly,  $P_u = \max_\mu \sum_x \mu(x) g(x) - Eg$ , where  $\mu : \{0, 1\}^n \rightarrow \mathbb{R}$  is subject to the constraints:

$$\begin{aligned} \sum_x \mu(x) &= 1 \\ \sum_x \mu(x) \mathcal{X}_y(x) &\leq 2\delta \quad \text{for all } y \in N_k^* \\ -\sum_x \mu(x) \mathcal{X}_y(x) &\leq 2\delta \quad \text{for all } y \in N_k^* \\ \mu(x) &\geq 0 \quad \text{for all } x \in \{0, 1\}^n. \end{aligned}$$

Its dual is thus  $P_u = \min \alpha_0 + 2\delta \sum_{y \in N_k^*} (\alpha'_y + \alpha''_y) - Eg$ , where  $\alpha_0$ ,  $\{\alpha'_y\}_{y \in N_k^*}$ , and  $\{\alpha''_y\}_{y \in N_k^*}$  are real coefficients subject to the constraints:

$$\begin{aligned} \alpha_0 + \sum_{y \in N_k^*} (\alpha'_y - \alpha''_y) \mathcal{X}_y(x) &\geq g(x) \quad \text{for all } x \in \{0, 1\}^n \\ \alpha'_y, \alpha''_y &\geq 0 \quad \text{for all } y \in N_k^*. \end{aligned}$$

In general, if  $a$  is real number, then  $\min\{a' + a'' : a', a'' \geq 0 \text{ s.t. } a' - a'' = a\} = |a|$ . Applying this to  $a' = \alpha'_y$ ,  $a'' = \alpha''_y$ , and  $a = \alpha_y = \alpha'_y - \alpha''_y$ , we get  $P_u = \min \alpha_0 + 2\delta \sum_{y \in N_k^*} |\alpha_y| - Eg$ , where  $\alpha_0$  and  $\{\alpha_y\}_y$  are real coefficients subject to the constraints:

$$\alpha_0 + \sum_{y \in N_k^*} \alpha_y \mathcal{X}_y(x) \geq g(x) \quad \text{for all } x \in \{0, 1\}^n.$$

Let  $g_u = \alpha_0 + \sum_{y \in N_k^*} \alpha_y \mathcal{X}_y$ . Noting that  $\alpha_0 = Eg_u$  and  $\alpha_y = \widehat{g}_u(y)$  for all  $y \in N_k^*$ , we get  $P_u = \min E(g_u - g) + 2\delta \sum_{y \neq 0} |\widehat{g}_u(y)|$ , where we are minimizing over all  $g_u : \{0, 1\}^n \rightarrow \mathbb{R}$  such that  $deg(g_u) \leq k$  and  $g_u(x) \geq g(x)$  for all  $x \in \{0, 1\}^n$ . ■

## B What will not work appendix

To justify the move from  $L_1$  to  $L_2$  in Section 5.5, it is appropriate to briefly mention two natural  $L_1$ -approaches which fall short of bounding the  $k$ -bias of  $s$ -DNF formulas.

**Inclusion-Exclusion:** It is natural to try to construct the sandwiching polynomials of an  $s$ -DNF formula by inclusion-exclusion as explained in [Baz03] (Section 5.5). This approach can be used to resolve the case of read-once DNF formulas (i.e., distinct clauses do not share variables), but we were unable to push it beyond the read-once case.

**Lift and reduce to an LP:** Let  $F$  be an  $s$ -DNF formula on  $n$  variables and let  $A_1, \dots, A_m$  be the clauses of  $F$ . Let  $\mu$  be a  $k$ -wise independent probability distribution on  $\{0, 1\}^n$  such that  $k > s$ . Let  $\mu_{unif}$  be the uniform probability distribution on  $\{0, 1\}^n$ . Consider the map  $L : \{0, 1\}^n \rightarrow \{0, 1\}^m$ ,  $x \mapsto (A_c(x))_{c=1}^m$ . Let  $\mu^*$  ( $\mu_{unif}^*$ , respectively) be the probability distribution induced via  $L$  on  $\{0, 1\}^m$  by  $\mu$  ( $\mu_{unif}$ , respectively). Thus  $Pr_\mu[F(x) = 0] = \mu^*(0)$ ,  $Pr_{\mu_{unif}}[F(x) = 0] = \mu_{unif}^*(0)$ , and  $E_{\mu^*} \mathcal{X}_y = E_{\mu_{unif}^*} \mathcal{X}_y$  for each  $y \in \{0, 1\}^m$  such that  $|y| \leq \lfloor k/s \rfloor$ .

This suggests relaxing the problem to the following LP:  $\max_{\mu_1, \mu_2} |\mu_1(0) - \mu_2(0)|$ , where  $\mu_1, \mu_2$  are probability distributions on  $\{0, 1\}^m$  such that  $E_{\mu_1} \mathcal{X}_y = E_{\mu_2} \mathcal{X}_y$ , for each  $y \in \{0, 1\}^m$  such that  $|y| \leq t$ , where  $t = \lfloor k/s \rfloor$ .

Unfortunately, that will not work. This follows from the approximate inclusion-exclusion lower bound of [LN90], which implies that the maximum of the above LP cannot be made arbitrarily small unless  $t = \Omega(\sqrt{m})$ . One of the issues of this relaxation is that it ignores the actual values of the  $t$ -moments<sup>7</sup> of  $\mu_1$  and  $\mu_2$ . It only uses the fact that the  $t$ -moments of  $\mu_1$  and  $\mu_2$  are equal. The values of those moments are simple and easy to derive from  $F$ , but taking them into consideration gives us an intriguing LP, which is not clear how to bound.

---

<sup>7</sup>If  $\mu$  is a probability distribution on  $\{0, 1\}^m$  and  $t \geq 0$  is an integer, define the *moments* vector of  $\mu$  to be  $c_\mu = (c_\mu(A) \stackrel{\text{def}}{=} E_{x \sim \mu} \wedge_{i \in A} x_i)_{A \subset [m]}$ , and define the  $t$ -*moments* of  $\mu$  to be the vector  $(c_\mu(A))_{A \subset [m]; |A| \leq t}$ . Note that  $c_\mu = \zeta_{B_m}^T \mu$ , where  $\zeta_{B_m}$  is the zeta function of the poset  $B_m$  consisting of the set of subsets of  $[m]$  ordered by inclusion.