

Polynomial Representations of Threshold Functions and Algorithmic Applications

Josh Alman*

Timothy M. Chan[†]Ryan Williams[‡]

Abstract

We design new polynomials for representing threshold functions in three different regimes: *probabilistic polynomials* of low degree, which need far less randomness than previous constructions, *polynomial threshold functions* (PTFs) with “nice” threshold behavior and degree almost as low as the probabilistic polynomials, and a new notion of *probabilistic PTFs* where we combine the above techniques to achieve even lower degree with similar “nice” threshold behavior. Utilizing these polynomial constructions, we design faster algorithms for a variety of problems:

- **Offline Hamming Nearest (and Furthest) Neighbors:** Given n red and n blue points in d -dimensional Hamming space for $d = c \log n$, we can find an (exact) nearest (or furthest) blue neighbor for every red point in randomized time $n^{2-1/O(\sqrt{c} \log^{2/3} c)}$ or deterministic time $n^{2-1/O(c \log^2 c)}$. These improve on a randomized $n^{2-1/O(c \log^2 c)}$ bound by Alman and Williams (FOCS’15), and also lead to faster MAX-SAT algorithms for sparse CNFs.
- **Offline Approximate Nearest (and Furthest) Neighbors:** Given n red and n blue points in d -dimensional ℓ_1 or Euclidean space, we can find a $(1 + \epsilon)$ -approximate nearest (or furthest) blue neighbor for each red point in randomized time near $dn + n^{2-\Omega(\epsilon^{1/3}/\log(1/\epsilon))}$. This improves on an algorithm by Valiant (FOCS’12) with randomized time near $dn + n^{2-\Omega(\sqrt{\epsilon})}$, which in turn improves previous methods based on locality-sensitive hashing.
- **SAT Algorithms and Lower Bounds for Circuits With Linear Threshold Functions:** We give a satisfiability algorithm for $AC^0[m] \circ LTF \circ LTF$ circuits with a subquadratic number of linear threshold gates on the bottom layer, and a subexponential number of gates on the other layers, that runs in deterministic 2^{n-n^ϵ} time. This strictly generalizes a SAT algorithm for $ACC^0 \circ LTF$ circuits of subexponential size by Williams (STOC’14) and also implies new circuit lower bounds for threshold circuits, improving a recent gate lower bound of Kane and Williams (STOC’16). We also give a randomized 2^{n-n^ϵ} -time SAT algorithm for subexponential-size $MAJ \circ AC^0 \circ LTF \circ AC^0 \circ LTF$ circuits, where the top MAJ gate and middle LTF gates have $O(n^{6/5-\delta})$ fan-in.

*Computer Science Department, Stanford University, jalman@cs.stanford.edu. Supported by NSF CCF-1212372 and NSF DGE-114747.

[†]Cheriton School of Computer Science, University of Waterloo, tmchan@uwaterloo.ca. Supported by an NSERC grant.

[‡]Computer Science Department, Stanford University, rrw@cs.stanford.edu. Supported in part by NSF CCF-1212372 and CCF-1552651 (CAREER). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

1 Introduction

The polynomial method is a powerful tool in circuit complexity. The idea of the method is to transform all circuits of some class into “nice” polynomials which represent the circuit in some way. If the polynomial is always sufficiently nice (e.g. has low degree), and one can prove that a certain Boolean function f cannot be represented so nicely, one concludes that the circuit class is unable to compute f .

Recently, these tools have found surprising uses in algorithm design. If a subproblem of an algorithmic problem can be modeled by a simple circuit, and that circuit can be transformed into a “nice” polynomial (or “nice” distribution of polynomials), then fast algebraic algorithms can be applied to evaluate or manipulate the polynomial quickly. This approach has led to advances on problems such as All-Pairs Shortest Paths [Wil14a], Orthogonal Vectors and Constraint Satisfaction [WY14, AWY15, Wil14d], All-Nearest Neighbor problems [AW15], and Stable Matching [MPS16].

In most applications, the key step is to randomly convert simple circuits into so-called *probabilistic* polynomials. If f is a Boolean function on n variables, and R is a ring, a *probabilistic polynomial over R for f with error $1/s$ and degree d* is a distribution \mathcal{D} of degree- d polynomials over R such that for all $x \in \{0, 1\}^n$, $\Pr_{p \sim \mathcal{D}}[p(x) = f(x)] \geq 1 - \frac{1}{s}$. Razborov [Raz87] and Smolensky [Smo87] introduced the notion of a probabilistic polynomial, and showed that any low-depth circuit consisting of AND, OR, and PARITY gates can be transformed into a low degree probabilistic polynomial by constructing constant degree probabilistic polynomials for those three gates. Many polynomial method algorithms use this transformation.

In this work, we are interested in polynomial representations of threshold functions. The threshold function TH_θ determines whether at least a θ fraction of its input bits are 1s. Threshold functions are among the simplest Boolean functions that do not have constant degree probabilistic polynomials: Razborov and Smolensky showed that the MAJORITY function (a special case of a threshold function) requires degree $\Omega(\sqrt{n \log s})$. Nonetheless, as we will see throughout this paper, there are many important problems which can be reduced to evaluating circuits involving threshold gates on many inputs, and so further study of polynomial representations of threshold functions is warranted.

Threshold functions have been extensively studied in theoretical computer science for many years; there are numerous applications of linear and polynomial threshold functions to complexity and learning theory (a sample includes [BRS91, BS92, ABR94, Bei95, KS01, OS10, She14]).

1.1 Our Results

We consider three different notions of polynomials representing TH_θ . Each achieves different tradeoffs between polynomial degree, the randomness required, and how accurately the polynomial represents TH_θ . Each leads to improved algorithms in our applications.

Less Randomness. First, we revisit probabilistic polynomials. Alman and Williams [AW15] designed a probabilistic polynomial for TH_θ which already achieves a tight degree bound of $\Theta(\sqrt{n \log s})$. However, their construction uses $\Omega(n)$ random bits, which makes it difficult to apply in deterministic algorithms. We show how their low-degree probabilistic polynomials for threshold functions can use substantially fewer random bits:

Theorem 1.1. *For any $0 \leq \theta \leq 1$, there is a probabilistic polynomial for the function TH_θ of degree $O(\sqrt{n \log s})$ on n bits with error $1/s$ that can be randomly sampled using only $O(\log n \log(ns))$ random bits.*

Polynomial Threshold Function Representations. Second, we consider deterministic Polynomial Threshold Functions (PTFs). A PTF for a Boolean function f is a polynomial (not a distribution on polynomials) $p : \{0, 1\}^n \rightarrow \mathbb{R}$ such that $p(x)$ is smaller than a fixed value when $f(x) = 0$, and $p(x)$ is larger

than the value when $f(x) = 1$. In our applications, we seek PTFs with “good threshold behavior”, such that $|p(x)| \leq 1$ when $f(x) = 0$, and $p(x)$ is very large otherwise. We can achieve almost the same degree for a PTF as for a probabilistic polynomial, and even better degree for an approximate threshold function:

Theorem 1.2. *We can construct a polynomial $P_{s,t,\varepsilon} : \mathbb{R} \rightarrow \mathbb{R}$ of degree $O(\sqrt{1/\varepsilon} \log s)$, such that*

- if $x \in \{0, 1, \dots, t\}$, then $|P_{s,t,\varepsilon}(x)| \leq 1$;
- if $x \in (t, (1 + \varepsilon)t)$, then $P_{s,t,\varepsilon}(x) > 1$;
- if $x \geq (1 + \varepsilon)t$, then $P_{s,t,\varepsilon}(x) \geq s$.

For the “exact” setting with $\varepsilon = 1/t$, we can alternatively bound the degree by $O(\sqrt{t \log(st)})$.

By summing multiple copies of the polynomial from Theorem 1.2, we immediately obtain a PTF with the same degree for the OR of $O(s)$ threshold functions (needed in our applications). This theorem follows directly from known extremal properties of Chebyshev polynomials, as well as the lesser known *discrete* Chebyshev polynomials. Because Theorem 1.2 gives a single polynomial instead of a distribution on polynomials, it is especially helpful for designing deterministic algorithms. Chebyshev polynomials are well-known to yield good approximate polynomials for computing certain Boolean functions over the reals [NS94, Pat92, KS01, She13, Val12] (please see the Preliminaries for more background).

Probabilistic PTFs. Third, we introduce a new (natural) notion of a *probabilistic PTF* for a Boolean function f . This is a distribution on PTFs, where for each input x , a PTF drawn from the distribution is highly likely to agree with f on x . Combining the techniques from probabilistic polynomials for TH_θ and the deterministic PTFs in a simple way, we construct a probabilistic PTF with good threshold behavior whose degree is *lower* than both the deterministic PTF and the degree bounds attainable by probabilistic polynomials (surprisingly breaking the “square-root barrier”):

Theorem 1.3. *We can construct a probabilistic polynomial $\tilde{P}_{n,s,t,\varepsilon} : \{0, 1\}^n \rightarrow \mathbb{R}$ of degree $O((1/\varepsilon)^{1/3} \log s)$, such that*

- if $\sum_{i=1}^n x_i \leq t$, then $|\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n)| \leq 1$ with probability at least $1 - 1/s$;
- if $\sum_{i=1}^n x_i \in (t, t + \varepsilon n)$, then $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) > 1$ with probability at least $1 - 1/s$;
- if $\sum_{i=1}^n x_i \geq t + \varepsilon n$, then $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) \geq s$ with probability at least $1 - 1/s$.

For the “exact” setting with $\varepsilon = 1/n$, we can alternatively bound the degree by $O(n^{1/3} \log^{2/3}(ns))$.

The PTFs of Theorem 1.3 can be sampled using only $O(\log(n) \cdot \log(ns))$ random bits as well; their lower degree will allow us to design faster randomized algorithms for a variety of problems. For emphasis, we will sometimes refer to PTFs as *deterministic PTFs* to distinguish them from probabilistic PTFs.

These polynomials for TH_θ can be applied to many different problems:

Offline Hamming Nearest Neighbor Search. In the Hamming Nearest Neighbor problem, we wish to preprocess a set D of n points in $\{0, 1\}^d$ such that, for a query $q \in \{0, 1\}^d$, we can quickly find the $p \in D$ with smallest Hamming distance to q . This problem is central to many problems throughout Computer Science, especially in search and error correction [Ind04]. However, it suffers from the *curse of dimensionality* phenomenon, where known algorithms achieve the nearly trivial runtimes of either $2^{\Omega(d)}$ or $\Omega(n/\text{poly}(\log n))$, with matching lower bounds in many data structure models (see e.g. [BR02]). Using our PTFs, we instead design a new algorithm for the natural offline version of this problem:

Theorem 1.4. *Given n red and n blue points in $\{0,1\}^d$ for $d = c \log n \ll \log^3 n / \log^5 \log n$, we can find an (exact) Hamming nearest/farthest blue neighbor for every red point in randomized time $n^{2-1/O(\sqrt{c} \log^{3/2} c)}$.*

Using the same ideas, we are also able to derandomize our algorithm, to achieve *deterministic* time $n^{2-1/O(c \log^2 c)}$ (see Remark 3 in Section 5). When $d = c \log n$ for constant c , these algorithms both have “truly subquadratic” runtimes. These both improve on Alman and Williams’ algorithm [AW15] which runs in randomized time $n^{2-1/O(c \log^2 c)}$, and only gives a nontrivial algorithm for $d \ll \log^2 n / \log^3 \log n$. Applying reductions from [AW15], we can achieve similar runtimes for finding closest pairs in ℓ_1 for vectors with small integer entries, and pairs with maximum inner product or Jaccard coefficient.

It is worth noting that there may be a serious limit to solving this problem much faster. Theorem 1.4 (and [AW15]) shows for all c there is a $\delta > 0$ such that Offline Hamming Nearest Neighbor search in dimension $d = c \log n$ takes $O(n^{2-\delta})$ time. Showing that there is a universal $\delta > 0$ that works for all c would disprove the Strong Exponential Time Hypothesis [AW15, Theorem 1.4].

Offline Approximate Nearest Neighbor Search. The problem of finding high-dimensional *approximate* nearest neighbors has received even more attention. Locality-sensitive hashing yields data structures that can find $(1 + \varepsilon)$ -factor approximate nearest neighbors to any query point in $\tilde{O}(dn^{1-\Omega(\varepsilon)})$ (randomized) time after preprocessing in $\tilde{O}(dn + n^{2-\Omega(\varepsilon)})$ time and space,¹ for not only Hamming space but also ℓ_1 and ℓ_2 space [HIM12, AI06]. Thus, a batch of n queries can be answered in $\tilde{O}(dn^{2-\Omega(\varepsilon)})$ randomized time. Exciting recent work on locality-sensitive hashing [AINR14, AR15] has improved the constant factor in the $\Omega(\varepsilon)$ bound, but not the growth rate in ε . In 2012, G. Valiant [Val12] reported a surprising algorithm running in $\tilde{O}(dn + n^{2-\Omega(\sqrt{\varepsilon})})$ randomized time for the offline version of the problem in ℓ_2 . We obtain a still faster algorithm for the offline problem, with $\sqrt{\varepsilon}$ improved to about $\varepsilon^{1/3}$:

Theorem 1.5. *Given n red and n blue points in $[U]^d$ and $\varepsilon \gg \frac{\log^6 \log n}{\log^3 n}$, we can find a $(1 + \varepsilon)$ -approximate ℓ_1 or ℓ_2 nearest/farthest blue neighbor for each red point in $(dn + n^{2-\Omega(\varepsilon^{1/3}/\log(1/\varepsilon))}) \cdot \text{poly}(\log(nU))$ randomized time.*

Valiant’s algorithm, like Alman and Williams’ [AW15], relied on fast matrix multiplication, and it also used Chebyshev polynomials but in a seemingly more complicated way. Our new probabilistic PTF construction is inspired by our attempt to unify Valiant’s approach with Alman and Williams’, which leads to not only a simplification but also an improvement of Valiant’s algorithm. (We also almost succeed in derandomizing Valiant’s $n^{2-\tilde{\Omega}(\sqrt{\varepsilon})}$ result in the Hamming case, except for an initial dimension reduction step; see Remark 3 in Section 5.)

Numerous applications to high-dimensional computational geometry follow; for example, we can approximate the diameter or Euclidean minimum spanning tree in roughly the same running time.

MAX-SAT. Another application is MAX-SAT: finding an assignment that satisfies the maximum number of clauses in a given CNF formula with n variables. In the sparse case when the number of clauses is cn , a series of papers have given faster exact algorithms, for example, achieving $2^{n-n/O(c \log c)}$ time by Dantsin and Wolpert [DW06], $2^{n-n/O(c \log c)^{2/3}}$ time by Sakai et al. [SSTT15a], and $2^{n-n/O(\sqrt{c})}$ time by Chen and Santhanam [CS15]. Using the polynomial method and our new probabilistic PTF construction, we obtain the following improved result:

Theorem 1.6. *Given a CNF formula with n variables and $cn \ll n^4 / \log^{10} n$ clauses, we can find an assignment that satisfies the maximum number of clauses in randomized $2^{n-n/O(c^{1/3} \log^{7/3} c)}$ time.*

¹Throughout the paper, the \tilde{O} notation hides polylogarithmic factors, $[U]$ denotes $\{0, 1, \dots, U-1\}$, and $\text{poly}(n)$ denotes a fixed polynomial in n .

For general dense instances, the problem becomes tougher. Williams [Wil04] gave an $O(2^{0.792n})$ -time algorithm for MAX-2-SAT, but an $O(2^{(1-\delta)n})$ -time algorithm for MAX-3-SAT (for a universal $\delta > 0$) has remained open; currently the best reported time bound [SSTT15b] is $2^{n-\Omega(n/\log n)^{1/3}}$, which can be slightly improved to $2^{n-\Omega(\sqrt{n/\log n})}$ with more care. We make new progress on not only MAX-3-SAT but also MAX-4-SAT:

Theorem 1.7. *Given a weighted 4-CNF formula F with n variables with positive integer weights bounded by $\text{poly}(n)$, we can find an assignment that maximizes the total weight of clauses satisfied in F , in randomized $2^{n-n/O(\log^2 n \log^2 \log n)}$ time. In the sparse case when the clauses have total weight cn , the time bound improves to $2^{n-n/O(\log^2 c \log^2 \log c)}$.*

LTF-LTF Circuit SAT Algorithms and Lower Bounds. Using our small sample space for probabilistic MAJORITY polynomials (Theorem 1.1), we construct a new circuit satisfiability algorithm for circuits with linear threshold functions (LTFs) which improves over several prior results. Let $\text{AC}^0[d, m] \circ \text{LTF} \circ \text{LTF}[S_1, S_2, S_3]$ be the class of circuits with a layer of S_3 LTFs at the bottom layer (nearest the inputs), a layer of S_2 LTFs above the bottom layer, and a size- S_1 $\text{AC}^0[m]$ circuit of depth d above the two LTF layers.²

Theorem 1.8. *For every integer $d > 0$, $m > 1$, and $\delta > 0$, there is an $\varepsilon > 0$ and an algorithm for satisfiability of $\text{AC}^0[d, m] \circ \text{LTF} \circ \text{LTF}[2^{n^\varepsilon}, 2^{n^\varepsilon}, n^{2-\delta}]$ circuits that runs in deterministic 2^{n-n^ε} time.*

Williams [Wil14b] gave a comparable SAT algorithm for $\text{ACC}^0 \circ \text{LTF}$ circuits of 2^{n^ε} size, where $\varepsilon > 0$ is sufficiently small.³ Theorem 1.8 strictly generalizes the previous algorithm, allowing another layer of $n^{2-\varepsilon}$ linear threshold functions below the existing LTF layer. Theorem 1.8 also trivially implies deterministic SAT algorithms for $\text{LTF} \circ \text{LTF}$ circuits of up to $n^{2-o(1)}$ gates, improving over the recent SAT algorithms of Chen, Santhanam, and Srinivasan [CSS16] which only work for $n^{1+\varepsilon}$ -wire circuits for $\varepsilon \ll 1$, and the SAT algorithms of Impagliazzo, Paturi, and Schneider [IPS13].

Here we sketch the ideas in the SAT algorithm for $\text{ACC}^0 \circ \text{LTF} \circ \text{LTF}$. Similar to the SAT algorithm for $\text{ACC}^0 \circ \text{LTF}$ circuits [Wil14b], the bottom layer of LTFs can be replaced by a layer of DNFs, via a weight reduction trick. We replace LTFs in the middle layer with $\text{AC}^0 \circ \text{MAJ}$ circuits (modifying a construction of Maciel and Thérien [MT98] to keep the fan-in of MAJ gates low), then replace these MAJ gates of $n^{2-\Theta(\delta)}$ fan-in with probabilistic \mathbb{F}_2 -polynomials of degree $n^{1-\Theta(\delta)+\Theta(\varepsilon)}$ over a small sample space, provided by Theorem 1.1. Taking a majority vote over all samples, and observing that an \mathbb{F}_2 -polynomial is a $\text{MOD}_2 \circ \text{AND}$ circuit, we obtain a $\text{MAJ} \circ \text{ACC}^0$ circuit, but with $2^{n^{1-o(\delta)}}$ size in some of its layers. By carefully applying known depth reduction techniques, we can convert the circuit into a depth-two circuit of size $2^{n^{1-\Omega(\varepsilon)}}$ which can then be evaluated efficiently on many inputs. (This is not obvious: applying the Beigel-Tarui depth reduction to a $2^{O(n^{1-\varepsilon})}$ -size circuit would make its new size *quasi-polynomial* in $2^{O(n^{1-\varepsilon})}$, yielding an intractable bound of $2^{n^{O(1)}}$.)

Applying the known connection between circuit satisfiability algorithms and circuit lower bounds for E^{NP} problems [Wil10, Wil14c, JMV15], the following is immediate:

Corollary 1.1. *For every $d > 0$, $m > 1$, and $\delta \in (0, 1)$, there is an $\varepsilon > 0$ such that the class E^{NP} does not have non-uniform circuits in $\text{AC}^0[d, m] \circ \text{LTF} \circ \text{LTF}[2^{n^\varepsilon}, 2^{n^\varepsilon}, n^{2-\delta}]$. In particular, for every $\varepsilon > 0$, E^{NP} does not have $\text{ACC}^0 \circ \text{LTF} \circ \text{LTF}$ circuits where the $\text{ACC}^0 \circ \text{LTF}$ subcircuit has $2^{n^{o(1)}}$ size and the bottom LTF layer has $n^{2-\varepsilon}$ gates.*

²Recall that for an integer $m \geq 2$, $\text{AC}^0[m]$ refers to constant-depth unbounded fan-in circuits over the basis $\{\text{AND}, \text{OR}, \text{MOD}_m\}$, where MOD_m outputs 1 iff the sum of its input bits is divisible by m .

³Recall ACC^0 is the infinite union of $\text{AC}^0[m]$ for all integers $m \geq 2$.

Most notably, Corollary 1.1 proves lower bounds with $n^{2-\varepsilon}$ LTFs on the bottom layer and *subexponentially many* LTFs on the second layer. This improves upon recent LTF \circ LTF gate lower bounds of Kane and Williams [KW16], at the cost of raising the complexity of the hard function from TC_3^0 to E^{NP} . Suguru Tamaki [Tam16] has recently reported similar results for depth-two circuits with both symmetric and threshold gates.

A Powerful Randomized SAT Algorithm. Finally, combining the probabilistic PTF for MAJORITY (Theorem 1.3) with the probabilistic polynomial of [AW15], we give a randomized SAT algorithm for a rather powerful class of circuits. The class $\text{MAJ} \circ \text{AC}^0 \circ \text{LTF} \circ \text{AC}^0 \circ \text{LTF}$ denotes the class of circuits with a majority gate at the top, along with two layers of linear threshold gates, and arbitrary $O(1)$ -depth AC^0 circuitry between these three layers. This circuit class is arguably much more powerful than TC_3^0 ($\text{MAJ} \circ \text{MAJ} \circ \text{MAJ}$), based on known low-depth circuit constructions for arithmetic functions (e.g. [CSV84, MT98, MT99]).

Theorem 1.9. *For all $\varepsilon > 0$ and integers $d \geq 1$, there is a $\delta > 0$ and a randomized satisfiability algorithm for $\text{MAJ} \circ \text{AC}^0 \circ \text{LTF} \circ \text{AC}^0 \circ \text{LTF}$ circuits of depth d running in $2^{n-\Omega(n^\delta)}$ time, on circuits with the following properties:*

- the top MAJ gate, along with every LTF on the middle layer, has $O(n^{6/5-\varepsilon})$ fan-in, and
- there are $O(2^{n^\delta})$ many AND/OR gates (anywhere) and LTF gates at the bottom layer.

Theorem 1.9 applies the probabilistic PTF of degree about $n^{1/3}$ (Theorem 1.3) to the top MAJ gate, probabilistic polynomials over \mathbb{Z} of degree about $n^{1/2}$ (Theorem 1.1) to the middle LTFs, and weight reduction to the bottom LTFs; the rest can be represented with $\text{poly}(n^\delta)$ degree.

It would not be surprising (to at least one author) if the above circuit class contained strong pseudo-random function candidates; that is, it seems likely that the Natural Proofs barrier applies to this circuit class. Hence from the circuit lower bounds perspective, the problem of derandomizing the SAT algorithm of Theorem 1.9 is extremely interesting.

2 Preliminaries

Notation. In what follows, for $(x_1, \dots, x_n) \in \{0, 1\}^n$ define $|x| := \sum_{i=1}^n x_i$. For a logical predicate P , we use the notation $[P]$ to denote the function which outputs 1 when P is true, and 0 when P is false.

For $\theta \in [0, 1]$, define $\text{TH}_\theta : \{0, 1\}^n \rightarrow \{0, 1\}$ to be the *threshold function* $\text{TH}_\theta(x_1, \dots, x_n) := [|x|/n \geq \theta]$. In particular, $\text{TH}_{1/2} = \text{MAJORITY}$.

For classes of circuits \mathcal{C} and \mathcal{D} , $\mathcal{C} \circ \mathcal{D}$ denotes the class of circuits consisting of a single circuit $C \in \mathcal{C}$ whose inputs are the outputs of some circuits from \mathcal{D} . That is, $\mathcal{C} \circ \mathcal{D}$ is simply the composition of circuits from \mathcal{C} and \mathcal{D} .

Rectangular Matrix Multiplication. One of our key tools is fast rectangular matrix multiplication:

Lemma 2.1 (Coppersmith [Cop82]). *For all sufficiently large N , multiplication of an $N \times N^{.172}$ matrix with an $N^{.172} \times N$ matrix can be done in $O(N^2 \log^2 N)$ arithmetic operations over any field.*

A proof can be found in the appendix of [Wil14b].

Chebyshev Polynomials in TCS. Another key to our work is that we find new applications of Chebyshev polynomials to algorithm design. This is certainly not a new phenomenon in itself; here we briefly survey some prior related usages of Chebyshev polynomials. First, Nisan and Szegedy [NS94] used Chebyshev polynomials to compute the OR function on n Boolean variables with an “approximating” polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$, such that for all $x \in \{0, 1\}^n$ we have $|\text{OR}(x) - p(x)| \leq 1/3$, yet $\deg(p) = O(\sqrt{n})$. They also

proved the degree bound is tight up to constants in the big-O; Paturi [Pat92] generalized the upper and lower bound to all symmetric functions.

This work has led to several advances in learning theory. Building on the polynomials of Nisan and Szegedy, Klivans and Servedio [KS01] showed how to compute an OR of t ANDs of w variables with a PTF of degree $O(\sqrt{w}\log t)$, similar to our degree bound for computing an OR of t MAJORITYs of w variables of Theorem 1.2 (however, note our bound in the “exact” setting is a bit better, due to our use of discrete Chebyshev polynomials). They also show how to compute an OR of s ANDs on n variables with a *deterministic* PTF of $O(n^{1/3}\log s)$ degree, similar to our cube-root-degree probabilistic PTF for the OR of MAJORITY of Theorem 1.3 in the “exact” setting. However, it looks difficult to generalize Klivans-Servedio’s $O(n^{1/3}\log s)$ degree bound to compute an OR of MAJORITY: part of their construction uses a reduction to decision lists which works for conjunctions but not for MAJORITY functions. Klivans, O’Donnell and Servedio [KOS04] show how to compute an AND of k MAJORITY on n variables with a PTF of degree $O(\sqrt{w}\log k)$. By a simple transformation via De Morgan’s law, there is a polynomial for OR of MAJORITY with the same degree. Their degree is only slightly worse than ours in terms of k (because we use discrete Chebyshev polynomials).

In streaming algorithms, Harvey, Nelson, and Onak [HNO08] use Chebyshev polynomials to design efficient algorithms for computing various notions of entropy in a stream. As a consequence of a query upper bound in quantum computing, Ambainis et al. [ACR⁺10] show how to approximate any Boolean formula of size s with a polynomial of degree $\sqrt{s}^{1+o(1)}$, improving on earlier bounds of O’Donnell and Servedio [OS10] that use Chebyshev polynomials. Sachdeva and Vishnoi [SV13] give applications of Chebyshev polynomials to graph algorithms and matrix algebra. Linial and Nisan [LN90] use Chebyshev polynomials to approximate inclusion-exclusion formulas, and Sherstov [She08] extends this to arbitrary symmetric functions.

3 Derandomizing Probabilistic Polynomials for Threshold Functions

In this section, we revisit the previous probabilistic polynomial for the majority function on n bits, and show it can be implemented using only $\text{polylog}(n, s)$ random bits. Our construction is essentially identical to that of [AW15], except that we use far fewer random bits to sample entries from the input vector in the recursive step of the construction.

For the analysis, we need a Chernoff bound for bits with limited independence:

Lemma 3.1 ([SSS95] Theorem 5 (I)(b)). *If X is the sum of k -wise independent random variables, each of which is confined to the interval $[0, 1]$, with $\mu = \mathbb{E}[X]$, $\delta \leq 1$, and $k = \lfloor \delta^2 \mu e^{-1/3} \rfloor$, then*

$$\Pr[|X - \mu| \geq \delta\mu] \leq e^{-\delta^2\mu/3}.$$

In particular, the following inequality appears in the analysis of [AW15]:

Corollary 3.1. *If $x \in \{0, 1\}^n$ with $|x|/n = w$, and $\tilde{x} \in \{0, 1\}^{n/10}$ is a vector each of whose entries is k -wise independently chosen entry of x , where $k = \lfloor 20e^{-1/3}\log(1/\varepsilon) \rfloor$, with $|\tilde{x}|/(n/10) = v$, then for every $\varepsilon < 1/4$,*

$$\Pr\left[v \leq w - \frac{a}{\sqrt{n}}\right] \leq \frac{\varepsilon}{4},$$

where $a = \sqrt{10} \cdot \sqrt{\ln(1/\varepsilon)}$.

Proof. Apply Lemma 3.1 with $X = |\tilde{x}|$, $\mu = \mathbb{E}[|\tilde{x}|] = wn$, and $\delta = \sqrt{40\log(1/\varepsilon)/n}$. □

Reminder of Theorem 1.1. For any $0 \leq \theta \leq 1$, there is a probabilistic polynomial for the threshold function TH_θ of degree $O(\sqrt{n \log s})$ on n bits with error $1/s$ that can be randomly sampled using $O(\log(n) \log(ns))$ random bits.

Proof. Our polynomial is defined recursively, just as in [AW15]. Set $\varepsilon = 1/s$. Using their notation, the polynomial $M_{n,\theta,\varepsilon}$ for computing TH_θ on n bits with error ε is defined by:

$$M_{n,\theta,\varepsilon}(x) := A_{n,\theta,2a}(x) \cdot S_{n/10,\theta,a/\sqrt{n},\varepsilon/4}(\tilde{x}) + M_{n/10,\theta,\varepsilon/4}(\tilde{x}) \cdot (1 - S_{n/10,\theta,a/\sqrt{n},\varepsilon/4}(\tilde{x})).$$

In [AW15], \tilde{x} was a sample of $n/10$ bits of x , chosen independently at random. Here, we pick \tilde{x} to be a sample of $n/10$ bits chosen k -wise independently, for $k = \lfloor 20e^{-1/3} \log(1/\varepsilon) \rfloor$. The other polynomials in this recursive definition are as in [AW15]:

- $M_{m,\theta,\varepsilon}$ for $m < n$ is the (recursively defined) probabilistic polynomial for TH_θ on m bits and ε error
- $S_{m,\theta,\delta,\varepsilon}(x) := (1 - M_{m,\theta+\delta,\varepsilon}(x)) \cdot M_{m,\theta-\delta,\varepsilon}(x)$ for $m < n$
- $A_{n,\theta,g} : \{0, 1\}^n \rightarrow \mathbb{Z}$ is an exact polynomial of degree at most $2g\sqrt{n} + 1$ which gives the correct answer to TH_θ for any vector x with $|x| \in [\theta n - g\sqrt{n}, \theta n + g\sqrt{n}]$, and may give arbitrary answers on other vectors.

Examining the proof of correctness in Alman and Williams [AW15], we see that the only requirement of the randomness is that it satisfies their Lemma 3.4, a concentration inequality for sampling \tilde{x} from x . Our Corollary 3.1 is identical to their Lemma 3.4, except that it replaces their method of sampling \tilde{x} with k -wise sampling; the remainder of the proof of correctness is exactly as before.

Our polynomial construction is recursive: we divide n by 10 and divide ε by 4, each time we move from one recursive layer to the next. At the j th recursive level of our construction, for $1 \leq j < \log_{10}(n)$, we need to $O(\log(4^j/\varepsilon))$ -wise independently sample $n/10^j$ entries from a vector of length $n/10^{j-1}$. Summing across all of the layers, we need a total of $O(n)$ samples from a k -wise independent space, where k is never more than $O(n/\varepsilon)$. This can be done all together using $O(n)$ samples from $\{1, 2, \dots, n\}$ which are $O(n/\varepsilon)$ -wise independent. Using standard constructions, this requires $O(\log(n) \log(n/\varepsilon))$ random bits. \square

4 PTFs for ORs of Threshold Functions

In this section, we show how to construct low-degree PTFs representing threshold functions that have good threshold behavior, and consequently obtain low-degree PTFs for an OR of many threshold functions.

4.1 Deterministic Construction

We begin by reviewing some basic facts about Chebyshev polynomials. The *degree- q Chebyshev polynomial of the first kind* is

$$T_q(x) := \sum_{i=0}^{\lfloor q/2 \rfloor} \binom{q}{2i} (x^2 - 1)^i x^{q-2i}.$$

Fact 4.1. For any $\varepsilon \in (0, 1)$,

- if $x \in [-1, 1]$, then $|T_q(x)| \leq 1$;
- if $x \in (1, 1 + \varepsilon)$, then $T_q(x) > 1$;
- if $x \geq 1 + \varepsilon$, then $T_q(x) \geq \frac{1}{2}e^{q\sqrt{\varepsilon}}$.

Proof. The first property easily follows from the known formula $T_q(x) = \cos(q \arccos(x))$ for $x \in [-1, 1]$. The second and third properties follow from another known formula $T_q(x) = \cosh(q \operatorname{arcosh}(x))$ for $x > 1$, which for $x \geq 1 + \varepsilon$ implies $T_q(x) \geq \cosh(q\sqrt{\varepsilon}) = \frac{1}{2}(e^{q\sqrt{\varepsilon}} + e^{-q\sqrt{\varepsilon}})$. \square

In certain scenarios, we obtain slightly better results using a (lesser known) family of *discrete Chebyshev polynomials* defined as follows [Hir03, page 59]:

$$D_{q,t}(x) := \sum_{i=0}^q (-1)^i \binom{q}{i} \binom{t-x}{q-i} \binom{x}{i}.$$

(See also [Sze75, pages 33–34] or Chebyshev’s original paper [Che99] with an essentially equivalent definition up to rescaling.)

Fact 4.2. Let $c_{q,t} = (t+1)^{q+1}/q!$. For all $t > q \geq \sqrt{8(t+1)\ln(t+1)}$,

- if $x \in \{0, 1, \dots, t\}$, then $|D_{q,t}(x)| \leq c_{q,t}$;
- if $x \leq -1$, then $D_{q,t}(x) \geq e^{q^2/(8(t+1))} c_{q,t}$.

Proof. From [Hir03, page 61],

$$\begin{aligned} \sum_{k=0}^t D_{q,t}(k)^2 &= \binom{2q}{q} \binom{t+1+q}{2q+1} \\ &= \frac{2q(2q-1)\cdots q}{q(q-1)\cdots 1} \cdot \frac{(t+1+q)(t+q)\cdots(t+1-q)}{(2q+1)(2q)\cdots 1} \\ &= \frac{(t+1)((t+1)^2-1^2)((t+1)^2-2^2)\cdots((t+1)^2-q^2)}{(2q+1)(q!)^2} \leq \frac{(t+1)^{2q+2}}{(q!)^2}. \end{aligned}$$

Thus, for every integer $x \in [0, t]$, we have $|D_{q,t}(x)| \leq (t+1)^{q+1}/q! = c_{q,t}$.

For $x \leq -1$, we have $(-1)^i \binom{x}{i} = \frac{(-x)(-x+1)\cdots(-x+i-1)}{1\cdot 2\cdots i} \geq 1$, and by the Chu–Vandermonde identity,

$$\begin{aligned} D_{q,t}(x) &\geq \sum_{i=0}^q \binom{q}{i} \binom{t+1}{q-i} = \binom{t+1+q}{q} \\ &= \frac{(t+1)^q (1 + \frac{1}{t+1})(1 + \frac{2}{t+1})\cdots(1 + \frac{q}{t+1})}{q!} \\ &\geq \frac{c_{q,t}}{t+1} e^{\frac{1+2+\cdots+q}{2(t+1)}} = e^{q(q+1)/(4(t+1)) - \ln(t+1)} c_{q,t} \geq e^{q^2/(8(t+1))} c_{q,t}. \end{aligned} \quad \square$$

Reminder of Theorem 1.2. We can construct a polynomial $P_{s,t,\varepsilon} : \mathbb{R} \rightarrow \mathbb{R}$ of degree $O(\sqrt{1/\varepsilon} \log s)$, such that

- if $x \in \{0, 1, \dots, t\}$, then $|P_{s,t,\varepsilon}(x)| \leq 1$;
- if $x \in (t, (1+\varepsilon)t)$, then $P_{s,t,\varepsilon}(x) > 1$;
- if $x \geq (1+\varepsilon)t$, then $P_{s,t,\varepsilon}(x) \geq s$.

For the “exact” setting with $\varepsilon = 1/t$, we can alternatively bound the degree by $O(\sqrt{t \log(st)})$.

Proof. Set $P_{s,t,\varepsilon}(x) := T_q(x/t)$ for a parameter q to be determined. The first two properties are obvious from Fact 4.1. On the other hand, if $x \geq (1 + \varepsilon)t$, then Fact 4.1 shows that $P_{s,t,\varepsilon}(x) \geq \frac{1}{2}e^{q\sqrt{\varepsilon}} \geq s$, provided we set $q = \left\lceil \sqrt{1/\varepsilon} \ln(2s) \right\rceil$. This achieves $O(\sqrt{1/\varepsilon} \log s)$ degree.

When $\varepsilon = 1/t$ the above yields $O(\sqrt{t} \log s)$ degree; we can reduce the $\log s$ factor by instead defining $P_{s,t,\varepsilon}(x) := D_{q,t}(t - x)/c_{q,t}$. Now, if $x \geq t + 1$, then $P_{s,t,\varepsilon}(x) \geq e^{q^2/(8(t+1))} \geq s$ by setting $q = \left\lceil \sqrt{8(t+1) \ln(\max\{s, t+1\})} \right\rceil$. \square

Using Theorem 1.2, we can construct a low-degree PTF for computing an OR of s thresholds of n bits:

Corollary 4.1. *Given n, s, t, ε , we can construct a polynomial $P : \{0, 1\}^{ns} \rightarrow \mathbb{R}$ of degree at most $\Delta := O(\sqrt{1/\varepsilon} \log s)$ and at most $s \cdot \binom{n}{\Delta}$ monomials, such that*

- if the formula $\bigvee_{i=1}^s [\sum_{j=1}^n x_{ij} > t]$ is false, then $|P(x_{11}, \dots, x_{1n}, \dots, x_{s1}, \dots, x_{sn})| \leq s$;
- if the formula $\bigvee_{i=1}^s [\sum_{j=1}^n x_{ij} \geq t + \varepsilon n]$ is true, then $P(x_{11}, \dots, x_{1n}, \dots, x_{s1}, \dots, x_{sn}) > 2s$.

For the exact setting with $\varepsilon = 1/n$, we can alternatively bound Δ by $O(\sqrt{n \log(ns)})$.

Proof. Define $P(x_{11}, \dots, x_{1n}, \dots, x_{s1}, \dots, x_{sn}) := \sum_{i=1}^s P_{n,3s,t,\varepsilon}(\sum_{j=1}^n x_{ij})$, where $P_{n,3s,t,\varepsilon}$ is from Theorem 1.2. The stated properties clearly hold. (In the second case, the output is at least $3s - (s - 1) > 2s$.) \square

4.2 Probabilistic Construction

Allowing ourselves a distribution of PTFs to randomly draw from, we can achieve noticeably lower degree than the previous section. We start with a fact which follows easily from the (tight) probabilistic polynomial for MAJORITY:

Fact 4.3. (Alman–Williams [AW15], or Theorem 1.1) *We can construct a probabilistic polynomial $Q_{n,s,t} : \{0, 1\}^n \rightarrow \mathbb{R}$ of degree $O(\sqrt{n \log s})$, such that*

- if $\sum_{i=1}^n x_i \leq t$, then $Q_{n,s,t}(x_1, \dots, x_n) = 0$ with probability at least $1 - 1/s$;
- if $\sum_{i=1}^n x_i > t$, then $Q_{n,s,t}(x_1, \dots, x_n) = 1$ with probability at least $1 - 1/s$.

Reminder of Theorem 1.3. *We can construct a probabilistic polynomial $\tilde{P}_{n,s,t,\varepsilon} : \{0, 1\}^n \rightarrow \mathbb{R}$ of degree $O((1/\varepsilon)^{1/3} \log s)$, such that*

- if $\sum_{i=1}^n x_i \leq t$, then $|\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n)| \leq 1$ with probability at least $1 - 1/s$;
- if $\sum_{i=1}^n x_i \in (t, t + \varepsilon n)$, then $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) > 1$ with probability at least $1 - 1/s$;
- if $\sum_{i=1}^n x_i \geq t + \varepsilon n$, then $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) \geq s$ with probability at least $1 - 1/s$.

For the “exact” setting with $\varepsilon = 1/n$, we can alternatively bound the degree by $O(n^{1/3} \log^{2/3}(ns))$.

Proof. Let r and q be parameters to be set later. Draw a random sample $R \subseteq \{1, \dots, n\}$ of size r . Let

$$t_R := \frac{tr}{n} - c_0 \sqrt{r \log s} \quad \text{and} \quad t^- := t - 2c_0 \left(\frac{n}{\sqrt{r}} \right) \sqrt{\log s}$$

for a sufficiently large constant c_0 . Define

$$\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_d) := Q_{r,2s,t_R}(\{x_i\}_{i \in R}) \cdot P_{s,t',\varepsilon'}\left(\sum_{i=1}^n x_i - t^-\right),$$

where $P_{s,t',\varepsilon'}$ is the polynomial from Theorem 1.2, with $t' := t - t^- = \Theta((n/\sqrt{r})\sqrt{\log s})$ and $\varepsilon' := \varepsilon n/t' = \Theta(\varepsilon\sqrt{r}/\sqrt{\log s})$.

To verify the stated properties, consider three cases:

- CASE 1: $\sum_{i=1}^n x_i < t^-$. By a standard Chernoff bound, with probability at least $1 - 1/(2s)$, we have $\sum_{i \in R} x_i < t^- r/n + c_0 \sqrt{r \log s} \leq t_R$ (assuming that $r \geq \log s$). Thus, with probability at least $1 - 1/s$, we have $Q_{r,2s,t_R}(\{x_i\}_{i \in R}) = 0$ and so $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) = 0$.
- CASE 2: $\sum_{i=1}^n x_i \in [t^-, t]$. With probability at least $1 - 1/s$, we have $Q_{r,2s,t_R}(\{x_i\}_{i \in R}) \in \{0, 1\}$ and so $|\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n)| \leq 1$.
- CASE 3: $\sum_{i=1}^n x_i > t$. By a standard Chernoff bound, with probability at least $1 - 1/(2s)$, we have $\sum_{i \in R} x_i \geq tr/n + c_0 \sqrt{r \log s} = t_R$. Thus, with probability at least $1 - 1/s$, we have $Q_{r,2s,t_R}(\{x_i\}_{i \in R}) = 1$ and so $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) > 1$ for $\sum_{i=1}^n x_i \in (t, t + \varepsilon n)$, or $\tilde{P}_{n,s,t,\varepsilon}(x_1, \dots, x_n) \geq s$ for $\sum_{i=1}^n x_i \geq t + \varepsilon n$.

The degree of $\tilde{P}_{n,s,t,\varepsilon}$ is

$$O\left(\sqrt{r \log s} + \sqrt{(1/(\varepsilon\sqrt{r}))\sqrt{\log s} \log s}\right)$$

and we can set $r = \lceil (1/\varepsilon)^{2/3} \log s \rceil$. For the exact setting, the degree is

$$O\left(\sqrt{r \log s} + \sqrt{(n/\sqrt{r})\sqrt{\log s} \cdot \log(ns)}\right)$$

and we can set $r = \lceil n^{2/3} \log^{1/3}(ns) \rceil$. □

Remark 1. Using the same techniques as in Theorem 1.1, we can sample a probabilistic polynomial from Theorem 1.3 with only $O(\log(n) \log(ns))$ random bits.

Corollary 4.2. Given d, s, t, ε , we can construct a probabilistic polynomial $\tilde{P} : \{0, 1\}^{ns} \rightarrow \mathbb{R}$ of degree at most $\Delta := O((1/\varepsilon)^{1/3} \log s)$ with at most $s \cdot \binom{n}{D}$ monomials, such that

- if $\bigvee_{i=1}^s [\sum_{j=1}^n x_{ij} \geq t]$ is false, then $|\tilde{P}(x_{11}, \dots, x_{1n}, \dots, x_{s1}, \dots, x_{sn})| \leq s$ with probability at least $2/3$;
- if $\bigvee_{i=1}^s [\sum_{j=1}^d x_{ij} \geq t + \varepsilon n]$ is true, then $\tilde{P}(x_{11}, \dots, x_{1n}, \dots, x_{s1}, \dots, x_{sn}) > 2s$ with probability at least $2/3$.

For the exact setting with $\varepsilon = 1/n$, we can alternatively bound Δ by $O(n^{1/3} \log^{2/3}(ns))$.

Proof. Define $\tilde{P}(x_{11}, \dots, x_{1n}, \dots, x_{s1}, \dots, x_{sn}) := \sum_{i=1}^s \tilde{P}_{n,3s,t_i,\varepsilon}(x_{i1}, \dots, x_{in})$. □

Remark 2. The coefficients of the polynomials from Fact 4.3 are $\text{poly}(n)$ -bit integers, and it can be checked that the coefficients of all our deterministic and probabilistic polynomials are rational numbers with $\text{poly}(n)$ -bit numerators and a common $\text{poly}(n)$ -bit denominator, and that the same bound for the number of monomials holds for the construction time, up to $\text{poly}(n)$ factors. That is, computations with these polynomials have low computational overhead relative to n .

5 Exact and Approximate Offline Nearest Neighbor Search

We now apply our new probabilistic PTF construction to obtain a faster algorithm for offline exact nearest/farthest neighbor search in Hamming space:

Reminder of Theorem 1.4. *Given n red and n blue points in $\{0,1\}^d$ for $d = c \log n \ll \log^3 n / \log^5 \log n$, we can find an (exact) Hamming nearest/farthest blue neighbor for every red point in randomized time $n^{2-1/O(\sqrt{c} \log^{3/2} c)}$.*

Proof. We proceed as in Abboud, Williams, and Yu’s algorithm for Boolean orthogonal vectors [AWY15] or Alman and Williams’ algorithm for Hamming closest pair [AW15]. For a fixed t , we first solve the decision problem of testing whether the nearest neighbor distance is less than t for each red point. (Farthest neighbors are similar.) Let $s = n^\alpha$ for some parameter α to be set later. Arbitrarily divide the blue point set into n/s groups of s points. For every group G of blue points and every red point q , we want to test whether

$$F(G, q) := \left[\min_{p \in G} \|p - q\|_1 < t \right] = \bigvee_{p \in G} \left[\sum_{i=1}^d (p_i q_i + (1 - p_i)(1 - q_i)) > d - t \right]$$

(where p_i denotes the i -th coordinate of a point p). By Corollary 4.2, we can express $F(G, q)$ as a probabilistic polynomial that has the following number of monomials:

$$\begin{aligned} s \cdot \binom{O(d)}{O(d^{1/3} \log^{2/3}(ds))} &\leq n^\alpha \cdot O\left(\frac{c \log n}{c^{1/3} \alpha^{2/3} \log n}\right)^{O(c^{1/3} \alpha^{2/3} \log n)} \\ &\leq n^\alpha \cdot n^{O(c^{1/3} \alpha^{2/3} \log \frac{c}{\alpha})} \ll (n/s)^{0.1} \end{aligned}$$

for large enough n , by setting α to be a sufficiently small constant times $1/(c^{1/3} \log^{3/2} c)$. The same bound holds for the construction time of the polynomial.

We can rewrite the polynomial for $F(G, q)$ as the dot product of two vectors $\phi(G)$ and $\psi(q)$ in $(n/s)^{0.1}$ dimensions over \mathbb{R} . The problem of evaluating $F(G, q)$ over all n/s groups G of blue points and all red points q then reduces to multiplying an $n/s \times (n/s)^{0.1}$ with an $(n/s)^{0.1} \times n$ matrix over \mathbb{R} . This in turn reduces to s instances of multiplication of $n/s \times (n/s)^{0.1}$ with $(n/s)^{0.1} \times n/s$ matrices, each of which can be done in $\tilde{O}(n/s)^2$ arithmetic operations on poly(d)-bit numbers over an appropriately large field (Lemma 2.1). The total time is $\tilde{O}(\text{poly}(d)n^2/s) = O(n^{2-1/O(c^{1/3} \log^{3/2} c)})$.

The error probability for each pair (G, q) is at most $1/3$, which can be lowered to $O(1/n^3)$, for example, by repeating $O(\log n)$ times (and taking the majority of the answers). The overall error probability is then $O(1/n)$. This solves the decision problem for a fixed t , but we can compute all nearest neighbor distances by calling the decision algorithm d times for all values of t . For each red point, we can find an actual nearest neighbor in additional $O(s)$ time, since we know which group achieves the nearest neighbor distance. \square

The same approach can be applied to solve *approximate* nearest neighbor search in Hamming space:

Theorem 5.1. *Given n red and n blue points in $\{0,1\}^d$ and $\varepsilon \gg \log^6(d \log n) / \log^3 n$, we can find an approximate Hamming nearest/farthest blue neighbor with additive error at most εd for each red point in randomized time $n^{2-\Omega(\varepsilon^{1/3} / \log(\frac{d}{\varepsilon \log n}))}$.*

Proof. We mimic the proof of Theorem 1.4 up to the definition of the polynomial $F(G, q)$. However, instead of applying the exact polynomial of Corollary 4.2, we insert the *approximate* polynomial construction from

the same corollary. While the exact polynomial had degree $O(d^{1/3} \log^{2/3}(ds))$, the approximate one has degree $O((1/\varepsilon)^{1/3} \log s)$. Setting

$$s := n^\alpha := n^{\Omega(\varepsilon^{1/3}/\log(\frac{d}{\varepsilon \log n}))},$$

the number of monomials in the new polynomial is now

$$\begin{aligned} s \cdot \left(\begin{array}{c} O(d) \\ O((1/\varepsilon)^{1/3} \log s) \end{array} \right) &\leq n^\alpha \cdot O\left(\frac{d}{(\alpha/\varepsilon^{1/3}) \log n} \right)^{O((\alpha/\varepsilon^{1/3}) \log n)} \\ &\leq n^\alpha \cdot n^{O((\alpha/\varepsilon^{1/3}) \log \frac{d}{\alpha \log n})} \ll (n/s)^{0.1}, \end{aligned}$$

for large enough n . The remainder of the algorithm is the same as the proof of Theorem 1.4, and the running time is $\tilde{O}(n^2/s^2) \leq n^{2-\Omega(\varepsilon^{1/3}/\log(\frac{d}{\varepsilon \log n}))}$. \square

Remark 3. For deterministic algorithms, using Corollary 4.1 instead, the time bounds for Theorems 1.4 and 1.5 become $n^{2-1/O(c \log^2 c)}$ and $n^{2-\Omega(\sqrt{\varepsilon}/\log(\frac{d}{\varepsilon \log n}))}$ respectively.

The algorithm of Theorem 5.1 still has three drawbacks: (i) the exponent in the time bound depends on the dimension d , (ii) the result requires additive instead of multiplicative error, and (iii) the result is for Hamming space instead of more generally ℓ_1 or ℓ_2 . We can resolve all three issues at once, by using known dimension reduction techniques:

Reminder of Theorem 1.5. *Given n red and n blue points in $[U]^d$ and $\varepsilon \gg \frac{\log^6 \log n}{\log^3 n}$, we can find a $(1 + \varepsilon)$ -approximate ℓ_1 or ℓ_2 nearest/farthest blue neighbor for each red point in $(dn + n^{2-\Omega(\varepsilon^{1/3}/\log(1/\varepsilon))}) \cdot \text{poly}(\log(nU))$ randomized time.*

Proof. (The ℓ_1 case.) We first solve the decision problem for a fixed threshold value t . We use a variant of ℓ_1 locality-sensitive hashing (see [And05]) to map points from ℓ_1 into low-dimensional Hamming space (providing an alternative to Kushilevitz, Ostrovsky, and Rabani's dimension reduction technique for Hamming space [KOR00]). For each red/blue point p and each $i \in \{1, \dots, k\}$, define $h_i(p) = (h_{i1}(p), \dots, h_{id}(p))$ with $h_{ij}(p) = \lfloor (p_{a_{ij}} + b_{ij}) / (2t) \rfloor$ where $a_{ij} \in \{1, \dots, d\}$ and $b_{ij} \in [0, 2t)$ are independent uniformly distributed random variables. For each of the $O(n)$ hashed values of h_i , pick a random bit; let $f_i(p)$ be the random bit associated with $h_i(p)$. Finally, define $f(p) = (f_1(p), \dots, f_k(p)) \in \{0, 1\}^k$. For any fixed p, q ,

$$\begin{aligned} \Pr[h_{ij}(p) \neq h_{ij}(q)] &= \frac{1}{d} \sum_{a=1}^d \min \left\{ \frac{|p_a - q_a|}{2t}, 1 \right\} \\ \Pr[f_i(p) \neq f_i(q)] &= \frac{1}{2} \Pr[h_i(p) \neq h_i(q)] = \frac{1}{2} \Pr \left[\bigvee_{j=1}^k [h_{ij}(p) \neq h_{ij}(q)] \right]. \end{aligned}$$

- If $\|p - q\|_1 \leq t$, then $\Pr[h_{ij}(p) \neq h_{ij}(q)] \leq \frac{\|p - q\|_1}{2dt} \leq \frac{1}{2d}$ and $\Pr[f_i(p) \neq f_i(q)] \leq \alpha_0 := \frac{1}{2}(1 - (1 - \frac{1}{2d})^d)$;
- if $\|p - q\|_1 \geq (1 + \varepsilon)t$, then $\Pr[h_{ij}(p) \neq h_{ij}(q)] \geq \min\{\frac{\|p - q\|_1}{2dt}, \frac{1}{d}\} \geq \frac{1 + \varepsilon}{2d}$ and $\Pr[f_i(p) \neq f_i(q)] \geq \alpha_1 := \frac{1}{2}(1 - (1 - \frac{1 + \varepsilon}{2d})^d)$.

Note that $\alpha_1 - \alpha_0 = \Omega(\varepsilon)$. By a Chernoff bound, it follows (assuming $k \geq \log n$) that

- if $\|p - q\|_1 \leq t$, then $\|f(p) - f(q)\|_1 \leq A_0 := \alpha_0 k + O(\sqrt{k \log n})$ with probability $1 - O(1/n^3)$;
- if $\|p - q\|_1 \geq (1 + \varepsilon)t$, then $\|f(p) - f(q)\|_1 \geq A_1 := \alpha_1 k - O(\sqrt{k \log n})$ with probability $1 - O(1/n^3)$.

Note that $A_1 - A_0 = \Omega(\varepsilon k)$ by setting k to be a sufficiently large constant times $(1/\varepsilon)^2 \log n$. We have thus reduced the problem to an approximate problem with additive error $O(\varepsilon k)$ for Hamming space in $k = O((1/\varepsilon^2) \log n)$ dimensions, which by Theorem 5.1 requires $n^{2-\Omega(\varepsilon^{1/3}/\log(1/\varepsilon))}$ time. The initial cost of applying the mapping f is $O(dkn)$.

This solves the decision problem; we can solve the original problem by calling the decision algorithm $O(\log_{1+\varepsilon} U)$ times for all t 's that are powers of $1 + \varepsilon$. \square

Proof. (The ℓ_2 case.) We use a version of the Johnson–Lindenstrauss lemma to map from ℓ_2 to ℓ_1 (see for example [Mat08]). For each red/blue point p , define $f(p) = (f_1(p), \dots, f_k(p)) \in \mathbb{R}^k$ with $f_i(p) = \sum_{j=1}^k a_{ij} p_j$, where the a_{ij} 's are independent normally distributed random variables with mean 0 and variance 1. For each fixed $p, q \in \mathbb{R}^d$, it is known that after rescaling by a constant, $\|f(p) - f(q)\|_1$ approximates $\|p - q\|_2$ to within $1 \pm O(\varepsilon)$ factor with probability $1 - O(1/n^3)$, by setting $k = O((1/\varepsilon)^2 \log n)$. It suffices to keep $O(\log U)$ -bit precision of the mapped points. The initial cost of applying the mapping f is $O(dkn)$ (which can be slightly improved by utilizing a sparse Johnson–Lindenstrauss transform [AC09]). \square

Numerous applications to high-dimensional computational geometry now follow. We briefly mention just one such application, building on the work of [IM98, HIM12]:

Corollary 5.1. *Given n points in $[U]^d$ and $\varepsilon \gg \log^6 \log n / \log^3 n$, we can find a $(1 + \varepsilon)$ -approximate ℓ_1 or ℓ_2 minimum spanning tree in $(dn + n^{2-\Omega(\varepsilon^{1/3}/\log(1/\varepsilon))}) \cdot \text{poly}(\log(nU))$ randomized time.*

Proof. Let G_r denote the graph where the vertex set is the given point set P and an edge pq is present whenever p and q have distance at most r . Har-Peled, Indyk, and Motwani [HIM12] gave a reduction of the approximate minimum spanning tree problem to the following *approximate connected components* problem:

Given a value r , compute a partition of P into subsets with the properties that (i) two points in the same subset must be in the same component in $G_{(1+\varepsilon)r}$, and (ii) two points in different subsets must be in different components in G_r .

The reduction is based on Kruskal's algorithm and increases the running time by a logarithmic factor.

To solve the approximate connected components problem, Har-Peled, Indyk, and Motwani gave a further reduction to online dynamic approximate nearest neighbor search. Since we want a reduction to offline static approximate nearest neighbor search, we proceed differently.

We first reduce the approximate connected components problem to the *offline approximate nearest foreign neighbors* problem:

Given a set P of n colored points with colors from $[n]$, for each point $q \in P$, find a $(1 + \varepsilon)$ -approximate nearest neighbor NFN_q among all points in P with color different from q 's color.

The reduction can be viewed as a variant of Boruvka's algorithm and is as follows: Initially assign each point a unique color and mark all colors as active. At each iteration, solve the offline approximate nearest foreign neighbors problem for points with active colors. For each q , if NFN_q and q have distance at most $(1 + \varepsilon)r$ and have different colors, merge the color class of NFN_q and q . If a color class has not been merged to other color classes during the iteration, mark its color as inactive. When all colors are inactive, output the color classes. Otherwise, proceed to the next iteration. The correctness of the algorithm is obvious. Since each iteration decreases the number of active colors by at least a half, the number of iterations is bounded by $O(\log n)$. Thus, the reduction increases the running time by a logarithmic factor.

To finish, we reduce the offline approximate nearest foreign neighbors problem to the standard (red/blue) offline approximate nearest neighbors problem by a standard trick: For each $j = 1, \dots, \lceil \log n \rceil$, for each point $q \in P$ where the j -th bit of q 's color is 0 (resp. 1), compute an approximate nearest neighbor of q among all points $p \in P$ where the j -th bit of p 's color is 1 (resp. 0). Record the nearest among all approximate nearest neighbors found for each point q . The final reduction increases the running time by another logarithmic factor. \square

6 Faster Algorithms For MAX-SAT

Next, we apply our improved probabilistic PTFs to obtain faster algorithms for MAX-SAT for sparse instances with cn clauses. We first consider MAX- k -SAT for small k before solving the general problem:

Theorem 6.1. *Given a k -CNF formula F (or k -CSP instance) with n variables and $cn \ll n^4 / (k^4 \log^6 n)$ clauses, we can find an assignment that satisfies the maximum number of clauses (constraints) of F in randomized $2^{n-n/O(k^{4/3}c^{1/3}\log(kc))}$ time.*

Proof. We proceed as in the $\#k$ -SAT algorithm of Chan and Williams [CW16]. We first solve the decision problem of testing whether there is a variable assignment satisfying more than t clauses for a fixed $t \in [cn]$. Let $s = \alpha n$ for some parameter $\alpha < 1/2$ to be set later.

For $j \in [cn]$, define the function $C_j(x_1, \dots, x_n) = 1$ if the j -th clause of the given formula is satisfied, and 0 otherwise. Note that each C_j can be expressed as a polynomial of degree at most k .

Say that a variable is *good* if it occurs in at most $2kc$ clauses. By the pigeonhole principle, at least half of the variables are good, so we can find s good variables x_1, \dots, x_s . Let x_{s+1}, \dots, x_n be the remaining variables, and let $J \subset [cn]$ be the set of indices of all clauses C_j that contain some occurrence of a good variable; note that $|J| = O(kcs)$. Now for every variable assignment $(x_{s+1}, \dots, x_n) \in \{0, 1\}^{n-s}$, we want to compute

$$F(x_{s+1}, \dots, x_n) := \sum_{(a_1, \dots, a_s) \in \{0, 1\}^s} \left[\sum_{j=1}^{cn} C_j(a_1, \dots, a_s, x_{s+1}, \dots, x_n) > t \right].$$

We will achieve this by computing for every $t' \in [cn]$:

$$G_{t'}(x_{s+1}, \dots, x_n) := \sum_{(a_1, \dots, a_s) \in \{0, 1\}^s} \left[\sum_{j \in J} C_j(a_1, \dots, a_s, x_{s+1}, \dots, x_n) > t' \right].$$

Let us define $T[x_{s+1}, \dots, x_n] := t - \sum_{j \notin J} C_j(0, \dots, 0, x_{s+1}, \dots, x_n)$. (Observe that it is OK to zero out the good variables x_1, \dots, x_s here, because we are only summing over clauses that *do not* contain them.) Note that T can be viewed as a polynomial in $n - s$ variables with only $\text{poly}(n)$ monomials. Therefore for all $(x_{s+1}, \dots, x_n) \in \{0, 1\}^{n-s}$, these T -values can be precomputed in $\text{poly}(n)2^{n-s}$ time. As these T -values are measuring the contribution from the variables x_{s+1}, \dots, x_n to the number of satisfied clauses, we have

$$F(x_{s+1}, \dots, x_n) = G_{T[x_{s+1}, \dots, x_n]}(x_{s+1}, \dots, x_n).$$

Applying Corollary 4.2 (in the exact setting), we can express any $G_{t'}$ as a sum of 2^s probabilistic polynomials of degree $k \cdot O((kcs)^{1/3}(s + \log(kcs))^{2/3})$, where each probabilistic polynomial computes an expression of the form $[\sum_{j \in J} p_j(x_{s+1}, \dots, x_n)]$ with error probability at most $1/(10 \cdot 2^s)$, and for all $j \in J$ we have $\deg(p_j(x_{s+1}, \dots, x_n)) \leq k$. The number of monomials in our probabilistic polynomial for $G_{t'}$ is at most

$$\begin{aligned} 2^s \cdot \binom{n-s}{k \cdot O((kcs)^{1/3}(s + \log(kcs))^{2/3})} &\leq 2^{\alpha n} \cdot O\left(\frac{n}{k^{4/3}c^{1/3}\alpha n}\right)^{O(k^{4/3}c^{1/3}\alpha n)} \\ &\leq 2^{\alpha n} \cdot 2^{O(k^{4/3}c^{1/3}\alpha \log \frac{1}{\alpha})n} \ll 2^{0.1n} \end{aligned}$$

by setting α to be a sufficiently small constant times $1/(k^{4/3}c^{1/3}\log(kc))$. The same bound holds for the construction time of the polynomial.

For each t' , we can evaluate the polynomial for $G_{t'}$ at all 2^{n-s} input values by divide-and-conquer or dynamic programming using $\text{poly}(n)2^{n-s}$ arithmetic operations [Yat37, Wil14c] on $\text{poly}(n)$ -bit numbers. The total time is $2^{n-n/O(k^{4/3}c^{1/3}\log(kc))}$. As before, the error probability can be lowered by taking the majority values over $O(n)$ repetitions, and the original problem can be solved by calling the decision algorithm for at most cn times. \square

Reminder of Theorem 1.6. *Given a CNF formula with n variables and $cn \ll n^4/\log^{10} n$ clauses, we can find an assignment that satisfies the maximum number of clauses in randomized $2^{n-n/O(c^{1/3}\log^{7/3}c)}$ time.*

Proof. We use a standard width reduction technique [SST15] originally observed by Schuler [Sch05] and studied closely by Calabro, Impagliazzo, and Paturi [CIP06]. Consider the following recursive algorithm:

- If all clauses have length at most k , then call the algorithm from Theorem 6.1 and return its output.
- Otherwise, pick a clause $(\alpha_1 \vee \dots \vee \alpha_\ell)$ with $\ell > k$. Return “SAT” if at least one of the two following calls return “SAT”:
 - Recursively solve the instance in which $(\alpha_1 \vee \dots \vee \alpha_\ell)$ is replaced by $(\alpha_1 \vee \dots \vee \alpha_k)$, and
 - recursively solve the instance in which $\alpha_1, \dots, \alpha_k$ are all assigned *false*.

Sakai, Seto, and Tamaki’s analysis for MAX-SAT [SST15] can be directly modified to show that the total time of this algorithm remains $2^{n-n/O(k^{4/3}c^{1/3}\log(kc))}$, when the parameter k is set to be a sufficiently large constant times $\log c$. \square

For MAX- k -SAT with $k \leq 4$, we can obtain a much better dependency on the sparsity parameter c ; in fact, we obtain significant speedup even for general dense instances. The approach this time requires only the previous probabilistic polynomials by Alman and Williams [AW15]. Naively, the dense case seems to require threshold functions with superlinearly many arguments, but by incorporating a few new ideas, we manage to solve MAX-4-SAT using only $O(n)$ -variate threshold functions.

Reminder of Theorem 1.7. *Given a weighted 4-CNF formula F with n variables with positive integer weights bounded by $\text{poly}(n)$, we can find an assignment that maximizes the total weight of clauses satisfied in F , in randomized $2^{n-n/O(\log^2 n \log^2 \log n)}$ time. In the sparse case when the clauses have total weight cn , the time bound improves to $2^{n-n/O(\log^2 c \log^2 \log c)}$.*

Proof. (Dense case.) Let $s = \alpha n$ for some parameter α to be set later. Arbitrarily divide the n variables of F into three groups: $x = \{x_1, \dots, x_{(n-s)/2}\}$, $y = \{y_1, \dots, y_{(n-s)/2}\}$, and $z = \{z_1, \dots, z_s\}$. As in Theorem 6.1, it suffices to solve the decision problem of whether there exist $x, y \in \{0, 1\}^{(n-s)/2}$ and $z \in \{0, 1\}^s$ such that $f(x, y, z) > t$, for a given degree-4 polynomial f and a fixed $t \in [n^{c_0}]$ (for an appropriately large constant c_0). Since f has degree 4, observe that each term has either (a) at most one y variable, (b) at most one x variable, or (c) no z variable. We can thus write

$$f(x, y, z) = \sum_{i=1}^{(n-s)/2} f_i(x, z)y_i + \sum_{i=1}^{(n-s)/2} g_i(y, z)x_i + h(x, y)$$

where the f_i ’s and g_i ’s are degree-3 polynomials, and h is a degree-4 polynomial.

For every $x, y \in \{0, 1\}^{(n-s)/2}$, it suffices to compute

$$F(x, y) := \sum_{z \in \{0, 1\}^s} [f(x, y, z) > t].$$

More generally, we compute for every $t' \in [n^{c_0}]$:

$$G_{t'}(x, y) := \sum_{z \in \{0,1\}^s} H_{z,t'}(x, y), \quad \text{with } H_{z,t'}(x, y) := \left[\sum_{i=1}^{(n-s)/2} f_i(x, z) y_i + \sum_{i=1}^{(n-s)/2} g_i(y, z) x_i > t' \right].$$

Then $F(x, y) = G_{t-h(x,y)}(x, y)$; we can precompute all $h(x, y)$ values in $\text{poly}(n)2^{n-s}$ time.

The $H_{z,t'}(x, y)$ predicate can be viewed as a *weighted* threshold function with $O(n)$ arguments. To further complicate matters, these weights are not fixed: they depend on x and y . We resolve the issue by extending the vectors x and y and using a binary representation trick.

For each vector $x \in \{0, 1\}^{(n-s)/2}$, define an *extended vector* x^* where $x_i^* = x_i$ for each $i = 1, \dots, (n-s)/2$ and $x_{i,j,z}^*$ is the j -th least significant bit in the binary representation of $f_i(x, z)$ for each $i = 1, \dots, (n-s)/2$, $j = 0, \dots, \ell$ and $z \in \{0, 1\}^s$, with $\ell = O(\log n)$. Note that x^* is a vector in $O(n \cdot \log n \cdot 2^s)$ dimensions. Similarly, for each vector $y \in \{0, 1\}^{(n-s)/2}$, define an extended vector y^* where $y_i^* = y_i$ for each $i = 1, \dots, (n-s)/2$ and $y_{i,j,z}^*$ is the j -th least significant bit in the binary representation of $g_i(y, z)$ for each $i = 1, \dots, (n-s)/2$, $j = 0, \dots, \ell$ and $z \in \{0, 1\}^s$. We can precompute all extended vectors in $2^{(n-s)/2} \cdot \text{poly}(n)2^s$ time.

Then

$$H_{z,t'}(x, y) := \sum_{(t_0, \dots, t_\ell)} \prod_{j=0}^{\ell} \left[\sum_{i=1}^{(n-s)/2} x_{i,j,z}^* y_i + \sum_{i=1}^{(n-s)/2} y_{i,j,z}^* x_i = t_j \right],$$

where the outer sum is over all tuples $(t_0, \dots, t_\ell) \in [n^{c_0}]^\ell$ with $\sum_{j=0}^{\ell} 2^j \cdot t_j > t'$.

By Fact 4.3, for each $z \in \{0, 1\}^s$, $j = 0, \dots, \ell$, and $t_j \in [n^{c_0}]$, we can construct a probabilistic polynomial (over \mathbb{R} or \mathbb{F}_2) for the predicate $\left[\sum_i x_{i,j,z}^* y_i + \sum_i y_{i,j,z}^* x_i = t_j \right]$ with degree $O(\sqrt{n \log S})$ with error probability at most $1/S$. By the union bound, the probability that there is an error for some z, j, t_j is at most $O((1/S) \cdot 2^s \cdot \log n \cdot n^{O(1)})$, which can be made at most $1/4^s$, for example, by setting $S = n^{c_0} 2^s$ for a sufficiently large constant c_0 . Thus, the degree for each predicate is $O(\sqrt{ns})$ (assuming $s \geq \log n$).

For each $z \in \{0, 1\}^s$ and $t' \in [n^{c_0}]$, by distributing over the product $\prod_{j=0}^{\ell}$ we can then construct a probabilistic polynomial for $H_{z,t'}(x, y)$ with degree $O(\sqrt{ns\ell}) \leq O(\sqrt{ns} \log n)$. For a fixed z and t' , such a polynomial is a function of $O(n \log n)$ free variables in x^* and y^* , and therefore has at most $\binom{O(n \log n)}{O(\sqrt{ns} \log n)}$ monomials. The same bound holds for the time needed to construct the probabilistic polynomial (note the number of tuples (t_0, \dots, t_ℓ) is $n^{O(\log n)}$, which is a negligible factor).

For each $t' \in [n^{c_0}]$, we can thus construct a probabilistic polynomial for $G_{t'}(x, y)$ with degree $O(\sqrt{ns} \log n)$ over x^* and y^* , with the following number of monomials:

$$\begin{aligned} 2^s \cdot \binom{O(n \log n)}{O(\sqrt{ns} \log n)} &\leq 2^{\alpha n} \cdot O\left(\frac{n \log n}{\sqrt{\alpha n \log n}}\right)^{O(\sqrt{\alpha n \log n})} \\ &\leq 2^{\alpha n} \cdot 2^{\sqrt{\alpha n}(\log(n)) \log(1/\alpha)} \lll 2^{0.1(n-s)/2} \end{aligned}$$

by setting α to be a sufficiently small constant times $1/(\log n \cdot \log \log n)^2$. The same bound holds for the construction time.

We can rewrite the polynomial for $G_{t'}(x, y)$ as the dot product of two vectors $\phi(x^*)$ and $\psi(y^*)$ of $2^{0.1(n-s)/2}$ dimensions. The problem of evaluating $G_{t'}(x, y)$ over all $x, y \in \{0, 1\}^{(n-s)/2}$ then reduces to multiplying a $2^{(n-s)/2} \times 2^{0.1(n-s)/2}$ with a $2^{0.1(n-s)/2} \times 2^{(n-s)/2}$ matrix (over \mathbb{R} or \mathbb{F}_2), which can be done in $\text{poly}(n)2^{n-s}$ time (Lemma 2.1). The total time is $2^{n-n/O(\log^2 n \log^2 \log n)}$. \square

Proof. (Sparse case.) If the clauses have total weight cn , we can refine the analysis above, in the following way. Let μ_i and ν_i be the maximum value of $f_i(x, z)$ and $g_i(y, z)$ respectively. We know that $\sum_i (\mu_i + \nu_i) \leq cn$.

The variable $x_{i,j,z}^*$ is needed only when $j \leq \log(\mu_i)$, and the variable $y_{i,j,z}^*$ is needed only when $j \leq \log(v_i)$. For each z, j, t_j , the probabilistic polynomial for the predicate

$$\left[\sum_i x_{i,j,z}^* y_i + \sum_i y_{i,j,z}^* x_i = t_j \right]$$

has degree $O(\sqrt{n_j s})$, where n_j is the number of i 's with $\mu_i \geq 2^j$ or $v_i \geq 2^j$.

Observe that $n_j = O(cn/2^j)$. It follows that the degree for the $H_{z,t'}(x, y)$ polynomial is $O(\sum_{j=0}^{\ell} \sqrt{n_j s}) = O(\sqrt{ns} \log c + \sum_{j > \log c} \sqrt{(cn/2^j)s}) = O(\sqrt{ns} \log c)$. The number of variables in $H_{z,t'}(x, y)$ is at most $O(\sum_{j=0}^{\ell} n_j) = O(n \log c + \sum_{j > \log c} (cn/2^j)) = O(n \log c)$.

Thus, the bound on the total number of monomials becomes

$$\begin{aligned} 2^s \cdot \left(\frac{O(n \log c)}{O(\sqrt{ns} \log c)} \right) &\leq 2^{\alpha n} \cdot O\left(\frac{n \log c}{\sqrt{\alpha n \log c}} \right)^{O(\sqrt{\alpha n \log c})} \\ &\leq 2^{\alpha n} \cdot 2^{\sqrt{\alpha n \log c} \log(1/\alpha)} \ll 2^{0.1(n-s)/2} \end{aligned}$$

by setting α to be a sufficiently small constant times $1/(\log c \log \log c)^2$. □

7 Circuit Satisfiability Algorithms

In this section, we give new algorithms for solving the SAT problem on some rather expressive circuit classes. First, we outline some notions used in both algorithms.

7.1 Satisfiability on a Cartesian Product

In intermediate stages of our SAT algorithms, we will study the following generalization of SAT, where the task is to find a SAT assignment in a ‘‘Cartesian product’’ of possible assignments.

Definition 7.1. *Let n be even, and let $A, B \subseteq \{0, 1\}^{n/2}$ be arbitrary. The SAT problem on the set $A \times B$ is to determine if a given n -input circuit has a satisfying assignment contained in the set $A \times B$.*

Recall that a Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ is a linear threshold function (LTF) if there are $a_1, \dots, a_n, t \in \mathbb{R}$ such that for all $x \in \{0, 1\}^n$, $f(x) = 1 \iff \sum_i a_i x_i \geq t$.

Let $\text{Circuit} \circ \text{LTF}[Z, S]$ be the class of circuits with a layer of S LTFs at the bottom (nearest the inputs), with Z additional arbitrary gates above that layer. Let $\text{Circuit} \circ \text{SUM} \circ \text{AND}[Z, S]$ be the analogous circuit class, but with S DNFs at the bottom layer with property that each DNF always has at most *one* conjunct true for every variable assignment. (Thus we may think of the DNF as simply an *integer sum*.) We first prove that the SAT problem for $\text{Circuit} \circ \text{LTF}$ can be reduced to the SAT problem for $\text{Circuit} \circ \text{SUM} \circ \text{AND}$, utilizing a weight reduction trick that can be traced back to Matoušek’s algorithm for computing dominances in high dimensions [Mat91, Wil14b]:

Lemma 7.1. *Let $A, B \subseteq \{0, 1\}^{n/2}$, with $|A| = |B| = N \leq 2^n$. Let $K \in [1, N]$ be an integer parameter. The SAT problem for $\text{Circuit} \circ \text{LTF}[Z, S]$ circuits on the set $A \times B$ can be reduced to the SAT problem for $\text{Circuit} \circ \text{SUM} \circ \text{AND}[Z, S]$ where each DNF has at most $O(\log K)$ terms and each AND has fan-in at most $2 \log K$, on a prescribed set $A' \times B'$ with $|A'| = |B'| = N$ and $A', B' \subseteq \{0, 1\}^{2S \log K}$. The reduction has the property that if the latter SAT problem can be solved in time T , then the former SAT problem can be solved in time $(T + N^2 \cdot Z^2 / K + N \cdot S) \cdot \text{poly}(n)$.*

Proof. For a given circuit C of type $\text{Circuit} \circ \text{LTF}[Z, S]$, let the j th LTF in the bottom layer have weights $\alpha_{j,1}, \dots, \alpha_{j,n}, t_j$. Let the assignments in A be a_1, \dots, a_N , and let the assignments in B be b_1, \dots, b_N . Denote the k th bit of a_i and b_i as $a_i[k]$ and $b_i[k]$, respectively.

Make $N \times S$ matrices M_A and M_B , where

$$M_A[i, j] = \sum_{k=1}^{n/2} \alpha_{j,k} \cdot a_i[k]$$

and

$$M_B[i, j] = t_j - \sum_{k=1}^{n/2} \alpha_{j,n/2+k} \cdot b_i[k].$$

The key property of these matrices is that $M_A[i, j] \geq M_B[i', j]$ if and only if the n -variable assignment $(a_i, b_{i'})$ makes the j th LTF output 1.

For each $j = 1, \dots, S$, let L_j be the list of all $2 \cdot N$ entries in the j th column of M_A and the j th column of M_B , sorted in increasing order. Partition L_j into K contiguous parts of $O(N/K)$ entries each, and think of each part of L_j as containing a set of $O(N/K)$ assignments from $A \cup B$. (So, the partition of L_j is construed as a partition of the assignments in $A \cup B$.) There are two possible cases for a satisfying assignment to the circuit C :

1. *There is a satisfying assignment $(a_i, b_{i'}) \in A \times B$ such that for some $j = 1, \dots, S$, a_i and $b_{i'}$ are in the same part of L_j .* By enumerating every $a_i \in A$, every $j = 1, \dots, S$, and all $O(N/K)$ assignments $b_{i'}$ of B which are in the same part of L_j as a_i , then evaluating the circuit C on the assignment $(a_i, b_{i'})$ in $Z^2 \cdot \text{poly}(n)$ time, we can determine satisfiability for this case in $O(N \cdot N/K \cdot Z^2) \cdot \text{poly}(n)$ time. If this does not uncover a SAT assignment, we move to the second case.
2. *There is a satisfying assignment $(a_i, b_{i'}) \in A \times B$ such that for every $j = 1, \dots, S$, a_i and $b_{i'}$ are different parts of L_j .* Then for every LTF gate $j = 1, \dots, S$ on the bottom layer of the circuit, we claim that the j -th LTF can be replaced by a sum of $O(\log K)$ ANDs on $2 \log K$ new variables. In particular, for the j -th LTF we define one new set of $\log K$ variables which encodes the index $k = 1, \dots, K$ such that a_i is in part k of L_j , and another set of $\log K$ variables which encodes the index k' such that $b_{i'}$ is in part k' of L_j . Then, determining $[k \geq k']$ is equivalent to determining whether $(a_i, b_{i'})$ satisfies the j -th LTF gate. Finally, note that the predicate $[k \geq k']$ can be computed by a DNF of $O(\log K)$ conjuncts. (Take an OR over all $\ell = 0, \dots, \log K$, guessing that the ℓ -th bit is the most significant bit in which k and k' differ; we can verify that guess with a conjunction on $2 \log K$ variables.) On every possible input $(k, k') \in \{0, 1\}^{2 \log K}$, the DNF has at most *one* true conjunction. Thus we can construe the OR as simply an *integer sum* of ANDs, as desired. Preparing these new assignments for this new SAT problem takes time $O(N \cdot S) \cdot \text{poly}(n)$. □

7.2 Simulating LTFs with AC0 of MAJORITY

In our SAT algorithms, we will need a way to simulate LTFs with bounded-depth circuits with MAJORITY gates. This was also used in Williams' work on solving ACC-LTF SAT [Wil14b], as a black box. However, here we must pay careful attention to the details of the construction. In fact, we will actually have to modify the construction slightly in order for our circuit conversion to work out. Let us review the construction here, and emphasize the parts that need modification for this paper. Recall that MAJ denotes the majority function.

Theorem 7.1 (Follows from [MT98], Theorem 3.3). *Every LTF can be computed by polynomial-size $AC^0 \circ MAJ$ circuits. Furthermore, the circuits can be constructed in polynomial time given the weights of the LTF, and the fan-in of each MAJ gate can be made $n^{1+\varepsilon}$, for every desired $\varepsilon > 0$, and the circuit has depth $O(\log(1/\varepsilon))$.*

It will be crucial for our final results that the fan-in of the MAJ gates can be made arbitrarily close to linear.

Proof. We begin by revisiting the circuit construction of Maciel and Thérien [MT98], which shows that the addition of n distinct n -bit numbers can be performed with polynomial-size $AC^0 \circ MAJ$ circuits. The original construction of Maciel and Thérien yields MAJ gates of fan-in $\tilde{O}(n^2)$, which is too large for our purposes. We can reduce the fan-in of MAJ gates to $O(n^{1+\varepsilon})$ by setting the parameters differently in their construction. Let us sketch their construction in its entirety, then describe how to modify it.

Recall that SYM denotes the class of symmetric functions. First, we show that addition of n n -bit numbers can be done in $AC^0 \circ SYM$. Suppose the n -bit numbers to be added are A_1, \dots, A_n , where $A_i = A_{i,n} \cdots A_{i,1}$ for $A_{j,i} \in \{0, 1\}$. Maciel and Thérien partition each A_i into m blocks of ℓ bits, where $m \cdot \ell = n$. They compute the sum S_k of the n ℓ -bit numbers in each block $k = 1, \dots, m$, i.e.

$$S_k = \sum_{i=1}^n \sum_{j=1}^{\ell} A_{i,(k-1)\ell+j} \cdot 2^{j-1},$$

and note that the desired sum is

$$z = \sum_{k=1}^m S_k \cdot 2^{(k-1)\ell}.$$

Each S_k can be represented in $\ell + \log n$ bits. Maciel and Thérien set $\ell = \log n$, so that each S_k is represented by 2ℓ bits. They then split each S_k into ℓ -bit numbers H_k and L_k such that

$$S_k = H_k \cdot 2^{\ell} + L_k.$$

Note that the ‘‘high’’ part H_k corresponds to the ‘‘carry bits’’ of S_k . They then note that if

$$y_1 := \sum_{k=1}^m H_k \cdot 2^{k\ell}, \quad y_2 := \sum_{k=1}^m L_k \cdot 2^{(k-1)\ell},$$

we have

- (a) $z = y_1 + y_2$, and
- (b) each bit of y_i is a function of exactly one H_k or L_k for some k . In turn, each L_k, H_k is a sum of $n \cdot \ell$ $A_{i,j}$'s where each $A_{i,j}$ is multiplied by a power of two in $[0, 2^{\ell}]$. Therefore, each bit of y_i can be computed by a SYM gate of fan-in at most $n \cdot \ell \cdot 2^{\ell} \leq n^2$.

We have therefore reduced the addition of n n -bit numbers to adding the two $O(n)$ -bit numbers y_1 and y_2 , with a layer of SYM gates. Adding two numbers can be easily computed in AC^0 (see for example [CFL85]), so the whole circuit is of the form $AC^0 \circ SYM$.

We wish to reduce the fan-in of the SYM gates to $O(n^{1+\varepsilon})$ for arbitrary $\varepsilon > 0$. To reduce the fan-in further, it suffices to find a construction that lets us reduce ℓ . Naturally, we can try to set $\ell = \varepsilon \log n$ for arbitrarily small $\varepsilon \in (0, 1)$. Without loss of generality, let us assume $1/\varepsilon$ is an integer. Then, each S_k is represented in $\ell + \log n \leq (1 + 1/\varepsilon)\ell$ bits. Let $t = 1 + 1/\varepsilon$. If we then split each S_k into t ℓ -bit numbers T_k^{t-1}, \dots, T_k^0 , ranging from high-order to low-order bits, we then have

$$S_k = T_k^{t-1} \cdot 2^{(t-1)\ell} + \dots + T_k^1 \cdot 2^{\ell} + T_k^0.$$

Defining the t numbers

$$y_i := \sum_{k=1}^m T_k^i \cdot 2^{(k+i-1)\ell},$$

the desired sum is $z = \sum_{i=0}^{t-1} y_i$. Just as before, each bit of y_i is a function of exactly one T_k^i for some k , which is a sum of $n \cdot \ell$ $A_{i,j}$'s where each $A_{i,j}$ is multiplied by an integer in $[0, 2^\ell]$. Hence each bit of y_i can be computed by a SYM gate of fan-in at most $n \cdot \ell \cdot 2^\ell \leq \tilde{O}(n^{1+\varepsilon})$. So with one layer of SYM gates, we have reduced the n number n -bit addition problem to the addition of t $O(n)$ -bit numbers y_0, \dots, y_{t-1} . But for $t \leq \log n$, addition of t n -bit numbers can be computed by AC^0 circuits of $\text{poly}(n)$ -size and *fixed* depth independent of t (see e.g. [Vol99], p.14-15). This completes the description of our $AC^0 \circ \text{SYM}$ circuit.

Observe that each SYM gate can be easily represented by an $OR \circ AND \circ MAJ$ circuit. In particular, the OR is over all $j \in \{0, 1, \dots, n\}$ such that the SYM gate outputs 1 when given j inputs are equal to 1, and the $AND \circ MAJ$ part computes $\sum_j x_j = j$. Again, the fan-in of each MAJ here is $\tilde{O}(n^{1+\varepsilon})$.

We now apply the addition circuits to show how every LTF on n variables can be represented by a polynomial-size $AC^0 \circ MAJ$ circuit. Suppose our LTF has weights w_1, \dots, w_{n+1} , computing $\sum_{j=1}^n w_j x_j \geq w_{n+1}$. By standard facts about LTFs, we may assume for all j that $|w_j| \leq 2^{bn \log_2 n}$ for some constant $b > 0$. Set $W = bn \log_2 n$.

Let D be a $AC^0 \circ MAJ$ circuit for adding n W -bit numbers as described above, where each MAJ gate has fan-in $\tilde{O}(n^{1+\varepsilon})$. For all $j = 1, \dots, n$, connect to the j th W -bit input of D a circuit which, given x_j , feeds w_j to D if the input bit $x_{i_j} = 1$, and the all-zero W -bit string if $x_j = 0$. Observe this extra circuitry is only wires, no gates: we simply place a wire from x_j to all bits of the j th W -bit input where the corresponding bit of w_j equals 1.

This new circuit D' clearly computes the linear form $\sum_{j=1}^n w_j x_j$. The linear form can then be compared to w_{n+1} with an AC^0 circuit, since the ‘‘less-than-or-equal-to’’ comparison of two integers can be performed in AC^0 . Indeed, this function can be represented as a quadratic-size DNF ($\text{SUM} \circ \text{AND}$), as was noticed in Lemma 7.1. We now have an $AC^0 \circ MAJ$ circuit D'' of size $\text{poly}(W, t) \leq n^b$ computing the LTF, where the MAJ gates have fan-in $\tilde{O}(n^{1+\varepsilon})$. \square

7.3 Satisfiability Algorithm for ACC of LTF of LTF

Let $AC^0[d, m] \circ \text{LTF} \circ \text{LTF}[S_1, S_2, S_3]$ be the class of circuits with a layer of S_3 LTFs at the bottom layer (nearest the inputs), a layer of S_2 LTFs above the bottom layer, and a size S_1 $AC^0[m]$ circuit of depth d above the two LTF layers.

Reminder of Theorem 1.8. *For every integer $d > 0$, $m > 1$, and $\delta > 0$, there is an $\varepsilon > 0$ and an algorithm for satisfiability of $AC^0[d, m] \circ \text{LTF} \circ \text{LTF}[2^{n^\varepsilon}, 2^{n^\varepsilon}, n^{2-\delta}]$ circuits that runs in deterministic 2^{n-n^ε} time.*

We use the following depth-reduction theorem of Beigel and Tarui (with important constructibility issues clarified by Allender and Gore [AG94], and recent size improvements by Chen and Papakonstantinou [CP16]):

Theorem 7.2 ([BT94, AG94]). *Every $\text{SYM} \circ \text{ACC}$ circuit of size s can be simulated by a $\text{SYM} \circ \text{AND}$ circuit of $2^{(\log s)^{c'}}$ size for some constant c' depending only on the depth d and $\text{MOD}m$ gates of the ACC part. Moreover, the AND gates of the final circuit have only $(\log s)^{c'}$ fan-in, the final circuit can be constructed from the original in $2^{O((\log s)^{c'})}$ time, and the final symmetric function at the output can be computed in $2^{O((\log s)^{c'})}$ time.*

Proof of Theorem 1.8. Let $\varepsilon > 0$ be a parameter to be set later. The plan is to start with a circuit as specified in the theorem statement, and slowly convert into a nice form that can be evaluated efficiently on many inputs.

1. Trade Variables for Circuit Size. Our first step is standard for ACC-SAT algorithms [Wil14b, Wil14c]: given an $\text{AC}^0[d, m] \circ \text{LTF} \circ \text{LTF}[2^{n^\varepsilon}, 2^{n^\varepsilon}, n^{2-\delta}]$ circuit C with n variables, create a copy of the circuit $C_v := C(v, \cdot)$ for all possible assignments $v \in \{0, 1\}^{n^\varepsilon}$ to the first n^ε variables of C , and define

$$C'(x_{n^\varepsilon+1}, \dots, x_n) := \bigvee_v C_v(x_{n^\varepsilon+1}, \dots, x_n).$$

Observe that C' is satisfiable if and only if C is satisfiable, C' has size at most $2^{O(n^\varepsilon)}$, C' is also an $\text{AC}^0 \circ \text{LTF} \circ \text{LTF}$ circuit, and C' has only $n - n^\varepsilon$ variables.

2. Replace the middle LTFs with MAJORITYs (Theorem 7.1). Note that each LTF on the second layer of C' has fan-in at most $n^{2-\delta} + n$, since the number of LTFs on the first layer is $n^{2-\delta}$. Applying the low fan-in transformation of Theorem 7.1, we can replace each of the LTFs on the second layer of C' with $\text{poly}(n)$ -size $\text{AC}^0 \circ \text{MAJ}$ circuits where each MAJ has fan-in at most $n^{2-\delta/2}$. This generates at most 2^{dn^ε} new MAJ gates in the circuit C' , for some constant $d > 0$, and produces a circuit of type

$$\text{ACC}^0 \circ \text{MAJ} \circ \text{LTF}.$$

3. Replace those MAJORITYs with (derandomized) probabilistic polynomials over \mathbb{F}_2 (Theorem 1.1). We replace each of these new MAJ gates with our low-randomness probabilistic polynomials for the MAJORITY function, as follows. Recall from Theorem 1.1 that we can construct a probabilistic polynomial over \mathbb{F}_2 for k -bit MAJORITY with degree $O(\sqrt{k \log(1/\varepsilon')})$ and error at most ε' , using a distribution of $k^{O(\log(k/\varepsilon'))}$ uniformly chosen \mathbb{F}_2 -polynomials. Setting $k := n^{2-\delta/2}$ for the fan-in of the MAJ gates, and the error to be $\varepsilon' := 1/2^{2dn^\varepsilon}$, the degree becomes

$$D := O\left(\sqrt{n^{2-\delta/2} \cdot 2dn^\varepsilon}\right) \leq O(n^{1-\delta/4+\varepsilon/2})$$

and the sample space has size $S = n^{O(n^\varepsilon)}$. For $\varepsilon \ll \delta/4$, we have $D := O(n^{1-\delta/8})$, and each polynomial in our sample space has at most $\binom{n^{2-\delta}}{n^{1-\delta/8}} \leq 2^{O(n^{1-\delta/8} \log n)}$ monomials. For every choice of the random seed r to the probabilistic polynomial, let C'_r be the circuit C' with the corresponding \mathbb{F}_2 polynomial P_r substituted in place of each MAJ gate. That is, each MAJ gate is substituted by an XOR of $2^{O(n^{1-\delta/8} \log n)}$ ANDs of fan-in at most $O(n^{1-\delta/8})$.

We now form a circuit C'' which takes a majority vote over all $2^{O(n^\varepsilon \log n)}$ circuits C'_r . The new circuit C'' therefore has the form

$$\text{MAJ} \circ \text{ACC}^0 \circ \text{XOR} \circ \text{AND} \circ \text{LTF},$$

where the $\text{MAJ} \circ \text{ACC}^0$ part has size $2^{O(n^\varepsilon \log n)}$, and each $\text{XOR} \circ \text{AND} \circ \text{LTF}$ subcircuit has size $2^{O(n^{1-\delta/8} \log n)}$. Since our probabilistic polynomial computes MAJORITY with $1/2^{2dn^\varepsilon}$ error and there are at most 2^{dn^ε} MAJ gates in C' , the new circuit C'' is equivalent to the original circuit C' .

4. Apply Beigel–Tarui to the top of the circuit, and distribute. It is very important to observe that we *cannot* apply Beigel–Tarui (Theorem 7.2) to the *entire* circuit C'' , as its total size is $2^{\Omega(n^{1-\delta/8} \log n)}$, and the quasi-polynomial blowup of Beigel–Tarui would generate a huge circuit of size $\Omega(2^n)$, rendering our conversion intractable.

However, the top $\text{MAJ} \circ \text{ACC}^0$ part is still small. Invoking the depth reduction lemma of Beigel and Tarui (Theorem 7.2 above), we can replace the $\text{MAJ} \circ \text{ACC}^0$ part in C'' of size $2^{O(n^\varepsilon \log n)}$ (even though it has $2^{O(n^\varepsilon \log n)}$ inputs from the XOR layer!) with a $\text{SYM} \circ \text{AND}$ circuit of size $2^{n^{a\varepsilon}}$ for a constant $a \geq 1$, where

each AND has fan-in at most $n^{a\varepsilon}$, and a depends only on the (constant) depth d and (constant) modulus m of the ACC^0 subcircuit.

The resulting circuit C_3 now has the form

$$\text{SYM} \circ \text{AND} \circ \text{XOR} \circ \text{AND} \circ \text{LTF}.$$

Applying the distributive law to the $\text{AND} \circ \text{XOR}$ parts, where the ANDs have fan-in at most $n^{a\varepsilon}$ and the XORs have fan-in $2^{O(n^{1-\delta/8} \log n)}$, each $\text{AND} \circ \text{XOR}$ parts can be converted into an $\text{XOR} \circ \text{AND}$ circuit of size $2^{O(n^{1-\delta/8+a\varepsilon} \log n)}$, where the fan-in of ANDs is at most $n^{a\varepsilon}$. Letting $\varepsilon \ll \delta/(ca)$ for sufficiently large $c \geq 1$, the fan-in of the new XORs is at most $2^{O(n^{1-\varepsilon})}$. We now have a circuit C_4 of the form

$$\text{SYM} \circ \text{XOR} \circ \text{AND} \circ \text{LTF}.$$

Note that the fan-in of the SYM gate is at most $2^{n^{a\varepsilon}}$, and the fan-in of the (merged) ANDs is $O(n^{1-\delta/8+a\varepsilon})$.

5. Apply modulus-amplifying polynomials to eliminate the XOR layer. We'd like to remove the XOR layer, to further reduce the depth of the circuit. But as the gates of this layer have very high fan-in, we must be careful not to blow the circuit size up to $\Omega(2^n)$. The following construction will take advantage of the fact that we have only $\text{poly}(n)$ total gates in the bottom LTF layer.

We apply one step of Beigel-Tarui's transformation [BT94] (from ACC^0 to $\text{SYM} \circ \text{AND}$) to the $\text{SYM} \circ \text{XOR} \circ \text{AND}$ part of our circuit. In particular, we apply a modulus-amplifying polynomial P (over the integers) of degree $2D' = 2n^{a\varepsilon}$ to each of the $\text{XOR} \circ \text{AND}$ parts. Construing the $\text{XOR} \circ \text{AND}$ as a sum of products $\sum \Pi$, the polynomial P has the property:

- If the $\sum \Pi = 1 \pmod 2$, then $P(\sum \Pi) = 1 \pmod{2^{D'}}$.
- If the $\sum \Pi = 0 \pmod 2$, then $P(\sum \Pi) = 0 \pmod{2^{D'}}$.

So, composing P with each $\text{XOR} \circ \text{AND}$ part, each P outputs either 0 or 1 modulo $2^{n^{a\varepsilon}}$. The key property here is that the modulus exceeds the fan-in of the SYM gate, so the sum of all $P(\sum \Pi)$ simply counts the number of $\text{XOR} \circ \text{AND}$ s which are true; this is enough to determine the output of the SYM gate. Construing the output of each bottom LTF gate as a variable, there are at most $n^{2-\varepsilon}$ variables. Expressing each $P(\sum \Pi)$ (expanded as a sum of products) as a multilinear polynomial in these LTF variables, the total number of terms is at most

$$\binom{n^{2-\varepsilon}}{D' \cdot n^{1-\delta/8+a\varepsilon}} \leq 2^{O(D' \cdot n^{1-\delta/8+a\varepsilon} \log n)} \leq 2^{O(n^{2a\varepsilon+1-\delta/8} \log n)}.$$

Let $\varepsilon := \delta/(ca)$ for a sufficiently large constant $c > 1$ so that $2a\varepsilon + 1 - \delta/8 < 1 - \varepsilon$. We can then merge the sum of all $P(\sum \Pi)$'s into the SYM gate, and obtain a $\text{SYM} \circ \text{AND}$ circuit where the SYM has fan-in

$$2^{O(n^{2a\varepsilon+(1-\delta/8)})} \leq 2^{O(n^{1-\varepsilon})},$$

and the AND gates have fan-in $O(n^{2a\varepsilon+(1-\delta/8)}) \leq O(n^{1-\varepsilon})$. The result is a circuit C_4 of the form

$$\text{SYM} \circ \text{AND} \circ \text{LTF}.$$

6. Replace the bottom threshold gates with DNFs (Theorem 7.1), and distribute. Note that the circuit C_4 has $n - n^\varepsilon$ variables, so our SAT algorithm would follow if we could evaluate C_4 on all of its variable assignments in $2^{n-n^\varepsilon} \cdot \text{poly}(n)$ time. We are now in a position to apply Lemma 7.1, which lets us reduce the evaluation problem for $\text{SYM} \circ \text{AND} \circ \text{LTF}$ circuits to the evaluation problem for $\text{SYM} \circ \text{AND} \circ \text{SUM} \circ \text{AND}$ circuits, with a parameter K that needs setting. Recall the middle AND gates have fan-in $O(n^{1-\varepsilon})$, and the fan-in of the SUM is $O(\log K)$. Therefore by the distributive law, we can rewrite the circuit as a $\text{SYM} \circ \text{SUM} \circ \text{AND}$ circuit, where each SUM gate has $(\log K)^{O(n^{1-\varepsilon})}$ ANDs below it, and at most *one*

AND below each SUM is true. Thus we can wire these AND gates directly into the top SYM gate without changing the output.

In more detail, let $A, B = \{0, 1\}^{(n-n^\varepsilon)/2}$, and set $N = 2^{(n-n^\varepsilon)/2}$ and the integer parameter $K := 2^{b \cdot n^{1-\varepsilon}}$ for a sufficiently large constant $b > 1$. By Lemma 7.1, we can reduce the SAT problem for $\text{SYM} \circ \text{AND} \circ \text{LTF}$ circuits of size $2^{O(n^{1-\varepsilon})}$ on the set $A \times B = \{0, 1\}^{n-n^\varepsilon}$ to the SAT problem for $\text{SYM} \circ \text{SUM} \circ \text{AND}$ circuits of size

$$2^{O(n^{1-\varepsilon})} \cdot 2^{2bn^{1-\varepsilon}} \cdot n^{2-\delta} \leq 2^{O(n^{1-\varepsilon})}$$

on a prescribed set $A' \times B'$ with $|A'| = |B'| = N$ and $A', B' \subseteq \{0, 1\}^{2bn^{2-\delta} \cdot n^{1-\varepsilon}}$. By the distributive argument from the previous paragraph, we can convert the $\text{SYM} \circ \text{SUM} \circ \text{AND}$ circuit into a $\text{SYM} \circ \text{AND}$ circuit of size at most

$$2^{O(n^{1-\varepsilon})} \cdot 2^{O(n^{1-\varepsilon} \log \log K)} \leq 2^{O(n^{1-\varepsilon} \log(n))}.$$

By Lemma 7.1, we know that if the $\text{SYM} \circ \text{AND}$ SAT problem is solvable in time T on the set $A' \times B'$, then the SAT problem for C_4 on the set $A \times B$ can be solved in time $O(T + N^2 \cdot Z/K + N \cdot S) \cdot \text{poly}(n)$.

7. Evaluate the depth-two circuit on many pairs of points. By applying fast rectangular matrix multiplication in a now-standard way [Wil14c, Wil14b], the resulting $\text{SYM} \circ \text{AND}$ circuit of $2^{\tilde{O}(n^{1-\varepsilon})}$ size can be evaluated on all points in $A' \times B'$, in time $\text{poly}(n) \cdot 2^{n-n^\varepsilon}$, thus solving its SAT problem. Therefore, the SAT problem for C_4 can be solved in time

$$\text{poly}(n) \cdot 2^{n-n^\varepsilon} + \frac{2^{n-n^\varepsilon} \cdot 2^{O(n^{1-\varepsilon})}}{2^{b \cdot n^{1-\varepsilon}}} + 2^{\frac{n-n^\varepsilon}{2}} \cdot 2^{O(n^{1-\varepsilon} \log(n))}.$$

Setting $b > 1$ to be sufficiently large, we obtain a SAT algorithm for C_4 (and hence the original circuit C) running in $\text{poly}(n) \cdot 2^{n-n^\varepsilon}$ time. \square

7.4 Satisfiability for Three Layers of Majority + AC0

In this section, we give our SAT algorithm for $\text{MAJ} \circ \text{AC}^0 \circ \text{LTF} \circ \text{AC}^0 \circ \text{LTF}$ circuits with low-polynomial fan-in at the output gate and the middle LTF layer:

Reminder of Theorem 1.9. *For all $\varepsilon > 0$ and integers $d \geq 1$, there is a $\delta > 0$ and a randomized satisfiability algorithm for $\text{MAJ} \circ \text{AC}^0 \circ \text{LTF} \circ \text{AC}^0 \circ \text{LTF}$ circuits of depth d running in $2^{n-\Omega(n^\delta)}$ time, on circuits with the following properties:*

- the top MAJ gate, along with every LTF on the middle layer, has $O(n^{6/5-\varepsilon})$ fan-in, and
- there are $O(2^{n^\delta})$ many AND/OR gates (anywhere) and LTF gates at the bottom layer.

We need one more result concerning probabilistic polynomials over the integers:

Theorem 7.3 ([BRS91, Tar93]). *For every AC^0 circuit C with n inputs and size s , there is a distribution of n -variate polynomials \mathcal{D} over \mathbb{Z} such that every p has degree $\text{poly}(\log s)$ (depending on the depth of C) and for all $x \in \{0, 1\}^n$, $\Pr_{p \sim \mathcal{D}}[C(x) = p(x)] \geq 1 - 1/2^{\text{poly}(\log s)}$.*

Proof of Theorem 1.9. The SAT algorithm is somewhat similar in structure to Theorem 1.8, but with a few important changes. Most notably, we work with probabilistic polynomials over \mathbb{Z} instead of \mathbb{F}_2 .

Start with a circuit C of the required form. Let s be the number of AND/OR gates in C plus the number of LTF gates on the bottom layer. Let $f \leq n^{6/5-\varepsilon}$ be the maximum fan-in of the top MAJ gate and the LTFs on the middle layer, and recall that we're planning to consider C with size at most 2^{n^δ} where $\delta > 0$ is a sufficiently small constant (depending on $\varepsilon > 0$ and the circuit depth) in the following. Our SAT algorithm runs as follows:

1. By Theorem 7.1, every LTF of fan-in f can be replaced by an $AC^0 \circ MAJ$ of fan-in $f^{1+o(1)}$ and $\text{poly}(f)$ size. Hence we can reduce C to a circuit of similar size, but of the form

$$MAJ \circ AC^0 \circ MAJ \circ AC^0 \circ MAJ.$$

The fan-ins of the majority gates in the middle and bottom layer can be made at most $n^{6/5-\varepsilon'}$, for any $\varepsilon' > 0$ which is smaller than ε . To be concrete, let us set $\varepsilon' := \varepsilon/2$.

2. Replace the “middle” majority gates of fan-in $n^{6/5-\varepsilon/2}$ with probabilistic polynomials (over \mathbb{Z}) of degree $n^{3/5-\varepsilon/4} \text{poly}(\log s)$ and error $1/2^{\text{poly}(\log s)}$ [AW15] (Theorem 1.1 in this paper). Replace all the AC^0 subcircuits of size s by probabilistic polynomials (over \mathbb{Z}) of degree $\text{poly}(\log s)$ and error $1/2^{\text{poly}(\log s)}$, via Lemma 7.3. Note that the latter $\text{poly}(\log s)$ factor depends on the depth of the circuit.
3. Replace the majority gate at the output (of fan-in $f \leq n^{6/5-\varepsilon}$) with the *probabilistic* PTF of Corollary 4.2, setting the threshold parameter s' (which is called s in the statement of the corollary) to be 2^{2n^δ} and setting the error (called ε in the statement of the corollary) to be $1/f$. The resulting polynomial has degree $n^{2/5-\varepsilon/3} \cdot \text{poly}(n^\delta)$.

Applying the distributive law to all the polynomials from steps 2 and 3, the new circuit C' can be viewed as an *integer sum* of at most T $AND \circ LTF$ circuits of at most T size, where

$$T = 2^{n^{3/5-\varepsilon/4} \cdot n^{2/5-\varepsilon/3} \cdot \text{poly}(\log s, n^\delta)} = 2^{n^{1-7\varepsilon/12} \cdot \text{poly}(\log s, n^\delta)}$$

and all AND gates have fan-in at most $n^{1-7\varepsilon/12} \cdot \text{poly}(\log s, n^\delta)$ (because the resulting polynomial has at most this degree).

Now is a good time to mention our choice of δ , as it will considerably clean up the exponents in what follows. We will choose $\delta > 0$ to be sufficiently small so that the $\text{poly}(\log s, n^\delta)$ factor in the exponent of T is less than $n^{\varepsilon/12}$. That is, we take $\delta := \varepsilon/c$ and the size parameter $s < 2^{n^\delta} = 2^{n^{\varepsilon/c}}$, for a sufficiently large constant $c \geq 12$. (Note that c depends on the depth of the circuit, since the degree of the poly log factor depends on the depth.) Thus we have the size bound

$$T = 2^{n^{1-7\varepsilon/12} \cdot \text{poly}(\log s, n^\delta)} \leq O(2^{n^{1-7\varepsilon/12} \cdot n^{\varepsilon/12}}) \leq O(2^{n^{1-\varepsilon/2}}),$$

and all AND gates have fan-in at most $n^{1-\varepsilon/2}$.

4. For all assignments a to the first n^δ variables of C' , plug a into C' , creating a copy C'_a . Let C'' be the integer sum of all 2^{n^δ} circuits C'_a . By the properties of the polynomial constructed in Theorem 1.3 and the chosen parameter $s' = 2^{2n^\delta}$, with probability at least $2/3$ there is a (computable) threshold value $v = 3s/2$ such that

- $C''(x) > v$ when at least one $C'_a(x)$ outputs 1, and
- $C''(x) < v$ when all $C'_a(x)$ output 0.

The circuit C'' is a $\text{Sum-of-AND} \circ \text{LTF}$ circuit; note that C'' has $n - n^\delta$ variables.

5. We now want to evaluate C'' on all of its 2^{n-n^δ} possible variable assignments. Applying Lemma 7.1 for an integer parameter $K \in [2^n]$ (to be determined), $N = 2^{(n-n^\delta)/2}$, and $Z, S = 2^{n^{1-\varepsilon/2}}$, we can convert this evaluation problem for C'' into a corresponding evaluation problem for a $\text{Sum-of-AND} \circ \text{SUM} \circ \text{AND}$ circuit C''' , on an appropriate combinatorial rectangle $A' \times B'$ of 2^{n-n^δ} variable assignments in total.

The relative size of the circuit is unchanged, as each $\text{SUM} \circ \text{AND}$ has size $O(\log^2 K) \leq O(n^2)$. The time for conversion of C'' into C''' is

$$\left(\frac{N^2 Z^2}{K} + N \cdot S \right) \cdot \text{poly}(n) \leq \frac{2^{n-n^\delta} \cdot 2^{2n^{1-\varepsilon/2}} \cdot \text{poly}(n)}{K}.$$

Setting $K := 2^{2n^{1-\varepsilon/2}}$ makes this time bound $2^{n-\Omega(n^\delta)}$.

Recall that in the $\text{Sum-of-AND} \circ \text{SUM} \circ \text{AND}$ circuit C''' , the fan-in of the middle ANDs is at most $n^{1-\varepsilon/2}$, and each SUM has $O(n)$ fan-in. We can therefore apply the distributive law to each $\text{AND} \circ \text{SUM}$ part, and obtain a $\text{SUM} \circ \text{AND}$ of size at most $n^{O(n^{1-\varepsilon/2})}$. Merging the SUMs into the SYM gate, we obtain a $\text{SYM} \circ \text{AND}$ circuit of size at most $n^{O(n^{1-\varepsilon/2})}$.

6. Finally, applying rectangular matrix multiplication (Lemma 2.1) we can evaluate the Sum-of-AND C''' of $n^{O(n^{1-\varepsilon/2})}$ size on the combinatorial rectangle $A' \times B'$ in $2^{n-\Omega(n^\delta)}$ time, by preparing matrices of dimensions $2^{n/2-\Omega(n^\delta)} \times n^{O(n^{1-\varepsilon/2})}$ (for A') and $n^{O(n^{1-\varepsilon/2})} \times 2^{n/2-\Omega(n^\delta)}$ (for B'), then multiplying them. Note that preparing these matrices takes time no more than $2^{n/2+O(n^{1-\varepsilon/2} \log n)}$, which is negligible for us.

After multiplying the matrices, we obtain a value for $C''(x)$ for each assignment x , which is correct with probability at least $2/3$. By repeating steps 2-5 for $100n$ times, we obtain correct values on all 2^{n-n^δ} points with high probability.

This completes the proof. □

8 Conclusion

Our work has led to interesting algorithmic improvements for several core problems. Here are two open problems that we wish to highlight.

First, it would be interesting to understand what are the power and limits of probabilistic polynomial threshold functions representing Boolean functions. How easy/difficult is it to prove degree lower bounds for such representations? In this paper, we have demonstrated how probabilistic PTFs can be significantly better than probabilistic polynomials or deterministic PTFs alone, by combining the strengths of the two representation methods. Informally, a probabilistic polynomial threshold function can be seen as an $\text{Approximate-MAJ} \circ \text{LTF} \circ \text{AND}$ circuit or as an $\text{Approximate-MAJ} \circ \text{LTF} \circ \text{XOR}$ circuit, so we are effectively asking about lower bounds regarding such circuit classes.

Second, can our SAT algorithm for $\text{MAJ} \circ \text{AC}^0 \circ \text{LTF} \circ \text{AC}^0 \circ \text{LTF}$ be derandomized? If so, the derandomization should lead to new circuit lower bounds. Perhaps the ideas in Tamaki's recent work [Tam16] will be helpful here.

Acknowledgments

The authors thank the FOCS referees for their helpful comments.

References

- [ABFR94] James Aspnes, Richard Beigel, Merrick Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

- [AC09] Nir Ailon and Bernard Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.
- [ACR⁺10] Andris Ambainis, Andrew M Childs, Ben W Reichardt, Robert Špalek, and Shengyu Zhang. Any and-or formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. *SIAM J. Computing*, 39(6):2513–2530, 2010.
- [AG94] Eric Allender and Vivek Gore. A uniform circuit lower bound for the permanent. *SIAM J. Computing*, 23(5):1026–1049, 1994.
- [AI06] Alexandr Andoni and Piotr Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In *FOCS*, pages 459–468, 2006.
- [AINR14] Alexandr Andoni, Piotr Indyk, Huy L Nguyen, and Ilya Razenshteyn. Beyond locality-sensitive hashing. In *SODA*, pages 1018–1028, 2014.
- [And05] Alexandr Andoni. Approximate nearest neighbor problem in high dimensions. Master’s thesis, MIT, 2005.
- [AR15] Alexandr Andoni and Ilya Razenshteyn. Optimal data-dependent hashing for approximate near neighbors. In *STOC*, pages 793–801, 2015.
- [AW15] Josh Alman and Ryan Williams. Probabilistic polynomials and Hamming nearest neighbors. In *FOCS*, pages 136–150, 2015.
- [AWY15] Amir Abboud, Ryan Williams, and Huacheng Yu. More applications of the polynomial method to algorithm design. In *SODA*, pages 218–230, 2015.
- [Bei95] Richard Beigel. The polynomial method in circuit complexity. In *IEEE Structure in Complexity Theory Conference*, pages 82–95, 1995.
- [BR02] Omer Barkol and Yuval Rabani. Tighter lower bounds for nearest neighbor search and related problems in the cell probe model. *JCSS*, 64(4):873–896, 2002.
- [BRS91] Richard Beigel, Nick Reingold, and Daniel A. Spielman. The perceptron strikes back. In *IEEE Structure in Complexity Theory Conference*, pages 286–291, 1991.
- [BS92] Jehoshua Bruck and Roman Smolensky. Polynomial threshold functions, AC^0 functions, and spectral norms. *SIAM J. Comput.*, 21(1):33–42, 1992.
- [BT94] Richard Beigel and Jun Tarui. On ACC. *Computational Complexity*, pages 350–366, 1994.
- [CFL85] Ashok K. Chandra, Steven Fortune, and Richard J. Lipton. Unbounded fan-in circuits and associative functions. *JCSS*, 30(2):222–234, 1985.
- [Che99] Pafnuty L. Chebyshev. Sur l’interpolation. In A. Markoff and N. Sonin, editors, *Oeuvres de P. L. Tchebychef*, volume 1, pages 539–560. Commissionnaires de L’Académie Impériale des Sciences, 1899.
- [CIP06] Chris Calabro, Russell Impagliazzo, and Ramamohan Paturi. A duality between clause width and clause density for SAT. In *IEEE Conf. Computational Complexity*, pages 252–260, 2006.
- [Cop82] Don Coppersmith. Rapid multiplication of rectangular matrices. *SIAM J. Comput.*, 11(3):467–471, 1982.

- [CP16] Shiteng Chen and Periklis A. Papakonstantinou. Depth-reduction for composites. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:85, 2016.
- [CS15] Ruiwen Chen and Rahul Santhanam. Improved algorithms for sparse MAX-SAT and MAX- k -CSP. In *SAT*, pages 33–45, 2015.
- [CSS16] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. Average-case lower bounds and satisfiability algorithms for small threshold circuits. In *CCC*, pages 1:1–1:35, 2016.
- [CSV84] Ashok K. Chandra, Larry Stockmeyer, and Uzi Vishkin. Constant depth reducibility. *SIAM J. Computing*, 13(2):423–439, 1984.
- [CW16] Timothy M. Chan and Ryan Williams. Deterministic APSP, orthogonal vectors, and more: Quickly derandomizing Razborov–Smolensky. In *SODA*, pages 1246–1255, 2016.
- [DW06] Evgeny Dantsin and Alexander Wolpert. MAX-SAT for formulas with constant clause density can be solved faster than in $O(2^n)$ time. In *SAT*, pages 266–276, 2006.
- [HIM12] Sariel Har-Peled, Piotr Indyk, and Rajeev Motwani. Approximate nearest neighbor: Towards removing the curse of dimensionality. *Theory of Computing*, 8(1):321–350, 2012.
- [Hir03] Mika Hirvensalo. *Studies on Boolean Functions Related to Quantum Computing*. PhD thesis, University of Turku, 2003.
- [HNO08] Nicholas J. A. Harvey, Jelani Nelson, and Krzysztof Onak. Sketching and streaming entropy via approximation theory. In *FOCS*, pages 489–498, 2008.
- [IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *STOC*, pages 604–613, 1998.
- [Ind04] Piotr Indyk. Nearest neighbors in high-dimensional spaces. In *Handbook of Discrete and Computational Geometry*, pages 877–892. Chapman and Hall, 2nd edition, 2004.
- [IPS13] Russell Impagliazzo, Ramamohan Paturi, and Stefan Schneider. A satisfiability algorithm for sparse depth two threshold circuits. In *FOCS*, pages 479–488, 2013.
- [JMV15] Hamid Jahanjou, Eric Miles, and Emanuele Viola. Local reductions. In *ICALP Part I*, pages 749–760, 2015.
- [KOR00] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Computing*, 30(2):457–474, 2000.
- [KOS04] Adam R. Klivans, Ryan O’Donnell, and Rocco A. Servedio. Learning intersections and thresholds of halfspaces. *J. Comput. Syst. Sci.*, 68(4):808–840, 2004.
- [KS01] Adam R. Klivans and Rocco Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. In *STOC*, pages 258–265, 2001.
- [KW16] Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:188, 2015. To appear in STOC’16.
- [LN90] Nathan Linial and Noam Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10(4):349–365, 1990.

- [Mat91] Jirí Matoušek. Computing dominances in E^n . *Inf. Process. Lett.*, 38(5):277–278, 1991.
- [Mat08] Jirí Matoušek. On variants of the Johnson–Lindenstrauss lemma. *Random Struct. Algorithms*, 33(2):142–156, 2008.
- [MPS16] Daniel Moeller, Ramamohan Paturi, and Stefan Schneider. Subquadratic algorithms for succinct stable matching. In *Computer Science Symposium in Russia*, pages 294–308, 2016.
- [MT98] Alexis Maciel and Denis Thérien. Threshold circuits of small majority-depth. *Information and Computation*, 146(1):55–83, 1998.
- [MT99] Alexis Maciel and Denis Thérien. Efficient threshold circuits for power series. *Information and Computation*, 152(1):62–73, 1999.
- [NS94] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994.
- [OS10] Ryan O’Donnell and Rocco A. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010.
- [Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In *STOC*, pages 468–474, 1992.
- [Raz87] A. A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.
- [Sch05] Rainer Schuler. An algorithm for the satisfiability problem of formulas in conjunctive normal form. *J. Algorithms*, 54(1):40–44, 2005.
- [She08] Alexander A Sherstov. Approximate inclusion-exclusion for arbitrary symmetric functions. In *CCC*, pages 112–123, 2008.
- [She13] Alexander A. Sherstov. Making polynomials robust to noise. *Theory of Computing*, 9:593–615, 2013.
- [She14] Alexander A. Sherstov. Breaking the Minsky–Papert barrier for constant-depth circuits. In *STOC*, pages 223–232, 2014.
- [Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *STOC*, pages 77–82, 1987.
- [SSS95] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff-Hoeffding bounds for applications with limited independence. *SIAM J. Discrete Mathematics*, 8(2):223–250, 1995.
- [SST15] Takayuki Sakai, Kazuhisa Seto, and Suguru Tamaki. Solving sparse instances of Max SAT via width reduction and greedy restriction. *Theory Comput. Syst.*, 57(2):426–443, 2015.
- [SSTT15a] Takayuki Sakai, Kazuhisa Seto, Suguru Tamaki, and Junichi Teruyama. Improved exact algorithms for mildly sparse instances of Max SAT. In *IPEC*, pages 90–101, 2015.
- [SSTT15b] Takayuki Sakai, Kazuhisa Seto, Suguru Tamaki, and Junichi Teruyama. A satisfiability algorithm for depth-2 circuits with a symmetric gate at the top and AND gates at the bottom. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:136, 2015.

- [SV13] Sushant Sachdeva and Nisheeth K Vishnoi. Faster algorithms via approximation theory. *Theoretical Computer Science*, 9(2):125–210, 2013.
- [Sze75] Gabor Szegő. *Orthogonal Polynomials*. American Mathematical Society, 1975.
- [Tam16] Suguru Tamaki. A satisfiability algorithm for depth two circuits with a sub-quadratic number of symmetric and threshold gates. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:100, 2016.
- [Tar93] Jun Tarui. Probabilistic polynomials, AC⁰ functions and the polynomial-time hierarchy. *Theor. Comput. Sci.*, 113(1):167–183, 1993.
- [Val12] Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and the closest pair problem. *J. ACM*, 62(2):13, 2015. Preliminary version in FOCS’12.
- [Vol99] Heribert Vollmer. *Introduction to Circuit Complexity: A Uniform Approach*. Springer, 1999.
- [Wil14a] Ryan Williams. Faster all-pairs shortest paths via circuit complexity. In *STOC*, pages 664–673, 2014.
- [Wil14b] Ryan Williams. New algorithms and lower bounds for circuits with linear threshold gates. In *STOC*, pages 194–202, 2014.
- [Wil14c] Ryan Williams. Nonuniform ACC circuit lower bounds. *J. ACM*, 61(1):2, 2014.
- [Wil14d] Ryan Williams. The polynomial method in circuit complexity applied to algorithm design (invited talk). In *34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS*, pages 47–60, 2014.
- [Wil04] Ryan Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theor. Comput. Sci.*, 348(2-3):357–365, 2005. See also ICALP’04.
- [Wil10] Ryan Williams. Improving exhaustive search implies superpolynomial lower bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013. See also STOC’10.
- [WY14] Ryan Williams and Huacheng Yu. Finding orthogonal vectors in discrete structures. In *SODA*, pages 1867–1877, 2014.
- [Yat37] F. Yates. The design and analysis of factorial experiments. *Technical Communication No. 35, Commonwealth Bureau of Soil Science, Harpenden, UK*, 1937.