The Unbounded-Error Communication Complexity of Symmetric Functions*

ALEXANDER A. SHERSTOV†

Abstract

We prove an essentially tight lower bound on the unbounded-error communication complexity of every symmetric function, i.e., $f(x, y) = D(|x \wedge y|)$, where $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ is a given predicate and x, y range over $\{0, 1\}^n$. Specifically, we show that the communication complexity of f is between $\Theta(k/\log^5 n)$ and $\Theta(k \log n)$, where k is the number of value changes of D in $\{0, 1, \ldots, n\}$. Prior to this work, the problem was solved only for the parity predicate D (Forster 2001).

Our proof is built around two new ideas. First, we show that a predicate D gives rise to a rapidly mixing random walk on $\mathbb{Z}_2^n$, which allows us to reduce the problem to communication lower bounds for "typical" predicates. Second, we use Paturi's approximation lower bounds (1992), suitably generalized here to clusters of real nodes in $[0, n]$ and interpreted in their dual form, to prove that a typical predicate behaves analogous to the parity predicate with respect to a smooth distribution on the inputs.

*An extended abstract of this article appeared in *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science (FOCS),* pages 384–393, 2008.

†The University of Texas at Austin, Department of Computer Sciences, Austin, TX 78712 USA. Email: sherstov@cs.utexas.edu.

1 Introduction

The *unbounded-error* model, due to Paturi and Simon [28], is a rich and elegant model of communication. Fix a function $f\colon X \times Y \to \{0,1\}$, where X and Y are some finite sets. Alice receives an input $x \in X$, Bob receives $y \in Y$, and their objective is to compute $f(x,y)$. To this end, they exchange bits through a shared communication channel according to a strategy, or *protocol,* established in advance. Alice and Bob each have an unlimited private source of random bits which they can use in deciding what messages to send. Eventually, Bob concludes this process by sending Alice a single bit, which is taken to be the output of their joint computation. Let the random variable $P(x,y) \in \{0,1\}$ denote the output bit when the parties receive inputs $x \in X$ and $y \in Y$. Alice and Bob's protocol is said to *compute* f if

$$\mathbf{P}[P(x,y) = f(x,y)] > \frac{1}{2}$$

for all $x \in X$, $y \in Y$. The *cost* of a given protocol is the worst-case number of bits exchanged on any input (x,y). The *unbounded-error communication complexity* of f, denoted $U(f)$, is the least cost of a protocol that computes f.

The unbounded-error model occupies a special place in the study of communication because it is more powerful than any of the usual models (deterministic, nondeterministic, randomized, quantum with or without entanglement). More precisely, the unbounded-error complexity $U(f)$ can be only negligibly greater than the complexity of f in these other models—and often, $U(f)$ is exponentially smaller. For completeness, we provide precise quantitative statements in Section 2.1. The power of the unbounded-error model resides in its very liberal success criterion: it suffices to produce the correct output with probability greater than $1/2$ (say, by an exponentially small amount). This contrasts with the more familiar *bounded-error* models, in which the correct output is expected with probability at least $2/3$.

1.1 Motivation

The additional power of the unbounded-error model has a consequence that proving communication lower bounds in it requires richer mathematical machinery. Furthermore, the resulting lower bounds will have applications that other communication models could not have. Before we state our results, we take a moment to review a few of these applications.

Circuit complexity. Recall that a *threshold gate* g with Boolean inputs $x_1, \ldots, x_n$ is a function of the form $g(x) = \operatorname{sgn}(a_1x_1 + \cdots + a_nx_n - \theta)$, for some

fixed reals $a_1, \ldots, a_n, \theta$. Thus, a threshold gate generalizes the familiar *majority* gate. A major unsolved problem in computational complexity is to exhibit a Boolean function that requires a depth-2 threshold circuit of superpolynomial size.

Communication complexity has been crucial to the progress on this problem. Via randomized communication complexity, many explicit functions have been found [11, 10, 26, 35, 38] that require depth-2 majority circuits of exponential size. By the reductions due to Goldman et al. [10], these lower bounds remain valid for the broader class of *majority-of-threshold* circuits. This solves a special case of the problem. The unbounded-error model solves another special case [8]: it supplies exponential lower bounds against *threshold-of-majority* circuits, i.e., circuits with a threshold gate at the top that receives inputs from majority gates.

Sign-rank and rigidity. Fix a real matrix $M = [M_{ij}]$ without zero entries. The *sign-rank* of M, denoted $\mathrm{rk}_{\pm}(M)$, is the least rank of a matrix $A = [A_{ij}]$ with $M_{ij}A_{ij} > 0$ for all i, j. In other words, sign-rank measures the sensitivity of the rank of M when its entries undergo sign-preserving perturbations. Sensitivity of the rank is a well-studied subject in complexity theory. For example, much work has focused on the closely related concept of *matrix rigidity* [23, 14].

Surprisingly, sign-rank and unbounded-error complexity turn out to be equivalent notions. Specifically, Paturi and Simon [28] showed that every function $f\colon X \times Y \to \{0,1\}$ satisfies $U(f) = \log_2 \mathrm{rk}_{\pm}(M) \pm O(1)$, where $M = [(-1)^{f(x,y)}]_{x \in X,\, y \in Y}$.

PAC learning. In a seminal paper, Valiant [40] formulated the *probably approximately correct* (PAC) model of learning, now a central model in computational learning theory. Research has shown that PAC learning is quite difficult. (By "PAC learning," we mean PAC learning under arbitrary distributions.) Indeed, the learning problem remains unsolved for such natural concept classes as DNF formulas of polynomial size and intersections of two halfspaces, whereas hardness results and lower bounds are abundant [15, 16, 20, 6, 21, 19].

There is, however, an important case when efficient PAC learning is possible. Let $\mathscr{C}$ be a given concept class. For notational convenience, view the functions in $\mathscr{C}$ as mappings $\{0,1\}^n \to \{-1,+1\}$ rather than $\{0,1\}^n \to \{0,1\}$. The *dimension complexity* of $\mathscr{C}$, denoted $\mathrm{dc}(\mathscr{C})$, is the least r for which there exist functions $\phi_1, \ldots, \phi_r\colon \{0,1\}^n \to \mathbb{R}$ such that every $f \in \mathscr{C}$ is expressible as $f(x) \equiv \mathrm{sgn}(a_1\phi_1(x) + \cdots + a_r\phi_r(x))$ for some reals $a_1, \ldots, a_r$. (The functions $\phi_1, \ldots, \phi_r$ themselves need not belong to $\mathscr{C}$, but in practice they often do.) There is a simple and well-known algorithm [18], based on linear programming, that PAC learns $\mathscr{C}$ in time polynomial in $\mathrm{dc}(\mathscr{C})$. To relate this discussion to sign-rank (or

equivalently, to unbounded-error complexity), let $M_{\mathscr{C}} = [f(x)]_{f \in \mathscr{C}, x \in \{0,1\}^n}$ be the characteristic matrix of $\mathscr{C}$. A moment's reflection reveals that $\mathrm{dc}(\mathscr{C}) = \mathrm{rk}_{\pm}(M_{\mathscr{C}})$, i.e., the dimension complexity of a concept class is precisely the sign-rank of its characteristic matrix.

Thus, the study of unbounded-error complexity yields nontrivial PAC learning algorithms. Indeed, the current fastest algorithm for learning polynomial-size DNF formulas in n variables [18] was obtained precisely by placing an upper bound of $2^{\tilde{O}(n^{1/3})}$ on the dimension complexity of that concept class. The dimension-complexity paradigm captures many other efficient PAC learning algorithms designed to date, with the notable exception of learning low-degree polynomials over $\mathrm{GF}(p)$.

1.2 Our Result

To summarize, the unbounded-error model has applications to circuit complexity, matrix analysis, and learning theory, in addition to its intrinsic appeal as a model of communication. Despite this motivation, progress in understanding unbounded-error complexity has been slow and difficult. Indeed, we are aware of only a few nontrivial results on this subject. Alon et al. [1] obtained strong lower bounds for random functions. In a breakthrough result, Forster [7] proved the first strong lower bound for an explicit function. Forster's proof has seen several extensions and refinements [8, 9]. Subsequent to our work, Razborov and Sherstov [33] solved an open problem regarding the comparative power of alternation (the classes Σ_2^{cc} and Π_2^{cc}) and unbounded-error communication, posed by Babai et al. [3].

This paper focuses on *symmetric* functions, i.e., functions $f\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ of the form

$$f(x, y) = D(|x \wedge y|)$$

for a given predicate $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$. Here $|x \wedge y|$ stands for the number of positions where x and y both have a 1. Familiar examples of such functions include DISJOINTNESS (determining if x and y intersect) and INNER PRODUCT MODULO 2 (determining if x and y intersect in an odd number of positions). Symmetric functions have seen much work in communication complexity. An illustrative example is the DISJOINTNESS function, whose study has led to considerable advances [13, 31, 29, 4] in randomized communication complexity. Symmetric functions have also contributed to the progress in quantum communication complexity, starting with the breakthrough result of Razborov [32] and continuing with more recent work, e.g., [17, 37, 39].

Our main result settles the unbounded-error complexity of every symmetric function, to within logarithmic factors. The only symmetric function whose

unbounded-error complexity was known prior to this work was INNER PRODUCT MODULO 2, for which Forster [7] proved a tight lower bound of $\Omega(n)$. The general result that we prove is in terms of the *degree* $\deg(D)$ of a given predicate D, defined as the number of times D changes value in $\{0, 1, \ldots, n\}$. In other words, $\deg(D) = |\{i : D(i) \neq D(i-1)\}|$.

Theorem 1.1 (Main Result). *Let* $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ *be given,* $k = \deg(D)$. *Define* $f(x, y) = D(|x \wedge y|)$. *Then*

$$\Theta(k/\log^5 n) \leqslant U(f) \leqslant \Theta(k \log n).$$

For a somewhat stronger quantitative statement, see Theorem 6.3.

The upper bound in this result has a short and elementary demonstration (see the proof of Theorem 6.3), and this paper is devoted entirely to the proof of the lower bound. The lower bound uses a combination of techniques (random walks on $\mathbb{Z}_2^n$, univariate approximation theory, linear programming duality), with Forster's general method as a starting point.

1.3 Proof Outline

Our proof consists of two independent parts. First, we reduce the original problem to analyzing what we call *dense* predicates. These are predicates $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ that change value frequently and at roughly regular intervals. Dense predicates are highly structured and amenable to direct analysis, unlike general predicates. With this reduction in hand, we complete the proof by solving the problem for every dense predicate. We now describe the two technical components in greater detail.

Reduction to dense predicates. Let $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ be a given predicate that is not dense. Any communication protocol that computes D can clearly compute the restriction of D to a given subinterval $\{i, i+1, \ldots, j\} \subset \{0,1,\ldots,n\}$. Now, let $\mathscr{D}$ denote the set of all restrictions of D to subintervals of a given length. Using a probabilistic argument, we show that a dense predicate arises as the XOR of a small number T of predicates from $\mathscr{D}$ (where T depends on the degree of D). As a result, if the original predicate D has unbounded-error complexity $\ll \deg(D)$, then some dense predicate will have disproportionately small unbounded-error complexity. This is the desired reduction.

The technical challenge here is to show that a dense predicate can be obtained as the XOR of a *small* number of predicates from $\mathscr{D}$. To this end, we model the

probabilistic argument as a random walk on $\mathbb{Z}_2^n$ and bound its mixing time. Our analysis uses a known bound, due to Razborov [30], on the rate of convergence in terms of the probability of a basis for $\mathbb{Z}_2^n$ (see Lemma 3.3).

Solution for dense predicates. Using Chebyshev polynomials and the Markov-Bernstein inequalities, Paturi [27] determined the least degree of a polynomial that approximates any given Boolean predicate on $\{0, 1, \ldots, n\}$ pointwise to within $1/3$. A starting point in our analysis is a related approximation problem, in which the nodes are no longer $\{0, 1, \ldots, n\}$ but are some arbitrary reals $\{a_1, a_2, \ldots, a_n\} \subset [0, n]$. Provided that the nodes are not too clumped together, we are able to prove strong lower bounds on the degree for a relevant class of approximation problems $f : \{a_1, a_2, \ldots, a_n\} \to \{0, 1\}$. Paturi's proof technique does not apply in this more general setting, and we give a direct analysis using fundamentals of approximation theory.

The next step is to show that computation of dense predicates corresponds to the approximation problem just described, where the real nodes $a_1, a_2, \ldots, a_n$ are allowed to form clusters but must still cover much of the interval $[0, n]$. Linear programming duality now tells us that, in a well-defined technical sense, a dense predicate behaves like the PARITY function with respect to a smooth distribution on the inputs. This enables us to bound the spectral norm of relevant matrices using the *pattern matrix method* [37]. In a final step, we invoke Forster's generalized theorem [8] to obtain our main result.

Comparison with related work. Alon et al. [1] introduced ideas from real algebraic geometry to the study of the unbounded-error complexity of random functions. Forster [7] used matrix analysis and a compactness argument to give the first strong lower bound for an explicit function. Follow-up work gave several matrix-analytic improvements [8, 9] on Forster's method. The recent result due to Razborov and Sherstov [33] is built around a new method of analyzing multivariate forms p on $\mathbb{R}^n$, whereby one projects p in several ways to a univariate polynomial, analyzes these simpler objects, and recombines the results using Fourier-theoretic tools.

Our approach is quite different from these works. To our knowledge, we give the first application of random walks to unbounded-error complexity. This technique generalizes beyond symmetric functions and, in fact, applies whenever one seeks a lower bound on the unbounded-error complexity of a set $\mathscr{F}$ of functions. The quality of the resulting communication lower bound will depend on how fast a random XOR-walk on $\mathscr{F}$ mixes to a hard function.

The second part of our proof is also based on a new idea, which effectively

allows us to treat a dense predicate as if it were the PARITY function. The insight here is that Paturi's approximation lower bounds [27], suitably generalized to clusters of real nodes in $[0, n]$ and examined in their dual form, show that a dense predicate behaves analogous to PARITY with respect to a *smooth* distribution on the inputs. We introduce the term *smooth orthogonalizing distribution* to describe this technique. The smoothness property is crucial to applying Forster's method [7].

1.4 Organization

Section 2 provides necessary technical background. Section 3 opens the proof with the reduction to dense predicates. Section 4 solves a certain problem in discrete approximation. Section 5 translates this approximation result, via linear programming duality and the Fourier transform, into an existence proof of *smooth orthogonalizing distributions* for every dense predicate. Section 6 combines the above ingredients to give the final lower bounds on unbounded-error complexity.

2 Preliminaries

A *Boolean function* is a mapping $X \to \{0, 1\}$, where X is a finite set. Typical cases are $X = \{0, 1\}^n$ and $X = \{0, 1\}^n \times \{0, 1\}^n$. The notation $[n]$ stands for the set $\{1, 2, \ldots, n\}$. Throughout this manuscript, "log" refers to the logarithm to base 2. The symbol P_k refers to the set of univariate polynomials of degree up to k.

For $x \in \{0, 1\}^n$, we define $|x| = x_1 + x_2 + \cdots + x_n$. For $x, y \in \{0, 1\}^n$, the notation $x \wedge y$ refers as usual to the component-wise AND of x and y. In particular, $|x \wedge y|$ stands for the number of positions where x and y both have a 1. At several places in this manuscript, it will be important to distinguish between addition over the reals and addition over GF(2). To avoid any confusion, we reserve the operator $+$ for the former and $\oplus$ for the latter.

Random walks on $\mathbb{Z}_2^n$ play an important role in this work. In particular, it will be helpful to recall the following fact.

Proposition 2.1 (Folklore). *For an integer $T \geqslant 1$, let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_T \in \{0, 1\}$ be independent random variables, each taking on 1 with probability p. Then*

$$\mathbf{E}\left[\mathbf{b}_1 \oplus \mathbf{b}_2 \oplus \cdots \oplus \mathbf{b}_T\right] = \frac{1}{2} - \frac{1}{2}(1 - 2p)^T.$$

Proof. Straightforward by induction on T. □

Predicates. A *predicate* is a mapping $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$. We say that a *value change* occurs at index $t \in \{1, 2, \ldots, n\}$ if $D(t) \neq D(t-1)$. The *degree* of D, denoted $\deg(D)$, is the total number of value changes of D. For example, the familiar predicate $\text{PARITY}(t) = t \bmod 2$ has degree n, whereas a constant predicate has degree 0. It is not hard to show [2] that $\deg(D)$ is the least degree of a real univariate polynomial p such that $\operatorname{sgn}(p(t)) = (-1)^{D(t)}$, $t = 0, 1, \ldots, n$, hence the term *degree.* Finally, given two predicates $D_1, D_2\colon \{0, 1, \ldots, n\} \to \{0, 1\}$, recall that their XOR is the predicate $D_1 \oplus D_2\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ defined by $(D_1 \oplus D_2)(t) = D_1(t) \oplus D_2(t)$.

Matrices. The symbol $\mathbb{R}^{m\times n}$ refers to the family of all $m \times n$ matrices with real entries. The (i, j)th entry of a matrix A is denoted by A_{ij}. We frequently use "generic-entry" notation to specify a matrix succinctly: we write $A = [F(i, j)]_{i,j}$ to mean that the (i, j)th entry of A is given by the expression $F(i, j)$. In most matrices that arise in this work, the exact ordering of the columns (and rows) is irrelevant. In such cases we describe a matrix by the notation $[F(i, j)]_{i\in I,\, j\in J}$, where I and J are some index sets. In specifying matrices, we will use the symbol $*$ for entries whose values are irrelevant, as in the proofs of Lemmas 3.2 and 3.5. Recall that the *spectral norm* of a matrix $A \in \mathbb{R}^{m\times n}$ is given by

$$\|A\| = \max_{x\in\mathbb{R}^n,\ \|x\|_2=1} \|Ax\|_2,$$

where $\|\cdot\|_2$ is the Euclidean norm on vectors.

Fourier transform over $\mathbb{Z}_2^n$. Consider the vector space of functions $\{0, 1\}^n \to \mathbb{R}$, equipped with the inner product

$$\langle f, g\rangle = 2^{-n} \sum_{x\in\{0,1\}^n} f(x)g(x).$$

For $S \subseteq [n]$, define $\chi_S\colon \{0, 1\}^n \to \{-1, +1\}$ by $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. Then $\{\chi_S\}_{S\subseteq[n]}$ is an orthonormal basis for the inner product space in question. As a result, every function $f\colon \{0, 1\}^n \to \mathbb{R}$ has a unique representation of the form

$$f(x) = \sum_{S\subseteq[n]} \hat{f}(S)\, \chi_S(x),$$

where $\hat{f}(S) = \langle f, \chi_S\rangle$. The reals $\hat{f}(S)$ are called the *Fourier coefficients of* f. The following fact is immediate from the definition of $\hat{f}(S)$.

Proposition 2.2. *Let $f\colon \{0,1\}^n \to \mathbb{R}$ be given. Then*

$$\max_{S\subseteq[n]} |\hat{f}(S)| \leqslant 2^{-n} \sum_{x\in\{0,1\}^n} |f(x)|.$$

Symmetric functions. Let S_n denote the group of permutations $[n] \to [n]$. A function $\phi\colon \{0,1\}^n \to \mathbb{R}$ is called *symmetric* if $\phi(x)$ is uniquely determined by $x_1 + \cdots + x_n$. Equivalently, ϕ is symmetric if $\phi(x) = \phi(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for every $x \in \{0,1\}^n$ and every $\sigma \in S_n$. Observe that for every $\phi\colon \{0,1\}^n \to \mathbb{R}$ (symmetric or not), the derived function

$$\phi_{\mathrm{sym}}(x) = \frac{1}{n!} \sum_{\sigma\in S_n} \phi(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

is symmetric. Symmetric functions on $\{0,1\}^n$ are intimately related to univariate polynomials, as demonstrated by Minsky and Papert's *symmetrization argument* [24]:

Proposition 2.3 (Minsky and Papert). *Let $\phi\colon \{0,1\}^n \to \mathbb{R}$ be symmetric with $\hat{\phi}(S) = 0$ for $|S| > r$. Then there is a polynomial $p \in P_r$ with $\phi(x) = p(|x|)$ for all $x \in \{0,1\}^n$.*

2.1 The Unbounded-Error Model of Communication

We continue the review started in the introduction. Readers with background in communication complexity will note that the unbounded-error model is exactly the same as the *private-coin randomized model* [22, Chap. 3], with one exception: in the latter case the correct answer is expected with probability at least $2/3$, whereas in the former case the correctness probability need only *exceed* $1/2$ (say, by an exponentially small amount). This difference has far-reaching implications. For example, the fact that the parties in the unbounded-error model do not have a *shared* source of random bits is crucial: allowing shared randomness would make the complexity of every function a constant, as one can easily verify. By contrast, introducing shared randomness into the randomized model has minimal impact on the complexity of any given function [25].

As one might expect, the weaker success criterion in the unbounded-error model has a drastic impact on the complexity of certain functions. For example, the well-known DISJOINTNESS function on n-bit strings has complexity $\Theta(\log n)$ in the unbounded-error model (see Proposition 6.5) and $\Theta(n)$ in the randomized

model [13, 31]. Furthermore, explicit functions are known [5, 36] with unbounded-error complexity $O(\log n)$ that require $\Omega(\sqrt{n})$ communication in the randomized model to even achieve advantage $2^{-\sqrt{n}/5}$ over random guessing.

More generally, the unbounded-error complexity of a function $f\colon X \times Y \to \{0,1\}$ is never much more than its complexity in the other standard models. For example, it is not hard to see that

$$\begin{aligned} U(f) &\leqslant \min\{N^0(f), N^1(f)\} + O(1) \\ &\leqslant D(f) + O(1), \end{aligned}$$

where D, N^0, and N^1 refer to communication complexity in the *deterministic, co-nondeterministic,* and *nondeterministic* models, respectively. Continuing,

$$\begin{aligned} U(f) &\leqslant R_{1/3}(f) + O(1) \\ &\leqslant O\left(R^{\text{pub}}_{1/3}(f) + \log\log[|X| + |Y|]\right), \end{aligned}$$

where $R_{1/3}$ and $R^{\text{pub}}_{1/3}$ refer to the *private-* and *public-coin randomized* models, respectively. As a matter of fact, one can show that

$$U(f) \leqslant O\left(Q^*_{1/3}(f) + \log\log[|X| + |Y|]\right),$$

where $Q^*_{1/3}$ refers to the *quantum model with prior entanglement*. An identical inequality is clearly valid for the quantum model *without* prior entanglement. See [22, 41] for rigorous definitions of these various models; our sole intention was to point out that the unbounded-error model is at least as powerful.

A compelling aspect of the unbounded-error model is that it has an exact interpretation in matrix-analytic terms. Specifically, let $M = [M_{ij}]$ be a real matrix without zero entries. Define the *sign-rank* of M by:

$$\text{rk}_{\pm}(M) = \min_A \{\text{rk}\, A : M_{ij}A_{ij} > 0 \quad \text{for all } i, j\}.$$

In words, $\text{rk}_{\pm}(M)$ is the least rank of a real matrix A whose entries each have the same sign as the corresponding entry of M. Paturi and Simon made the following important observation [28, Thm. 2].

Theorem 2.4 (Paturi and Simon). *Let X, Y be finite sets and $f\colon X \times Y \to \{0,1\}$ a given function. Put $M = [(-1)^{f(x,y)}]_{x\in X, y\in Y}$. Then*

$$U(f) \;=\; \log \text{rk}_{\pm}(M) \;\pm\; O(1).$$

Paturi and Simon's original observation concerned $X = Y = \{0,1\}^n$, but their proof readily extends to arbitrary sets. In words, the unbounded-error complexity of a function essentially equals the logarithm of the sign-rank of its communication matrix. This equivalence is often helpful: sometimes it is more convenient to reason in terms of communication protocols, and sometimes the matrix formulation offers more insight.

The power of the unbounded-error model makes it a challenging model in which to prove communication lower bounds. In a breakthrough result, Forster [7] proved the first strong lower bound in the unbounded-error model for an explicit function. Forster's proof generalizes to yield the following result [8, Thm. 3], which serves as a crucial starting point for our work.

Theorem 2.5 (Forster et al.). *Let X, Y be finite sets and $M = [M_{xy}]_{x \in X, y \in Y}$ a real matrix without zero entries. Then*

$$\mathrm{rk}_{\pm}(M) \geqslant \frac{\sqrt{|X|\,|Y|}}{\|M\|} \min_{x,y} |M_{xy}|.$$

We close this overview by discussing some closure properties of the unbounded-error model. Given functions $f, g\colon X \times Y \to \{0,1\}$, recall that their XOR is the function $f \oplus g\colon X \times Y \to \{0,1\}$ defined by $(f \oplus g)(x,y) = f(x,y) \oplus g(x,y)$. We have:

Proposition 2.6 (Folklore). *Let $f, g\colon X \times Y \to \{0,1\}$ be arbitrary. Then*

$$U(f \oplus g) \leqslant U(f) + U(g).$$

Proof. Alice and Bob can evaluate f and g individually and output the XOR of the two answers. It is straightforward to verify that this strategy is correct with probability greater than $1/2$. □

In what follows, we will be interested primarily in the complexity of predicates $D\colon \{0,1,\ldots,n\} \to \{0,1\}$. Specifically, we define $U(D)$ to be the unbounded-error communication complexity of the function $f\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ given by $f(x,y) = D(|x \wedge y|)$.

2.2 Pattern Matrices

Pattern matrices were introduced in [38, 37] and proved useful in obtaining strong lower bounds on communication. Relevant definitions and results from [37] follow.

Let t and n be positive integers with $t \mid n$. Split $[n]$ into t contiguous blocks, each with n/t elements:

$$[n] = \left\{1, 2, \ldots, \frac{n}{t}\right\} \cup \left\{\frac{n}{t} + 1, \ldots, \frac{2n}{t}\right\} \cup \cdots \cup \left\{\frac{(t-1)n}{t} + 1, \ldots, n\right\}.$$

Let $\mathscr{V}(n,t)$ denote the family of subsets $V \subseteq [n]$ that have exactly one element in each of these blocks (in particular, $|V| = t$). Clearly, $|\mathscr{V}(n,t)| = (n/t)^t$. For a bit string $x \in \{0,1\}^n$ and a set $V \in \mathscr{V}(n,t)$, define the *projection of x onto V* by

$$x|_V = (x_{i_1}, x_{i_2}, \ldots, x_{i_t}) \in \{0,1\}^t,$$

where $i_1 < i_2 < \cdots < i_t$ are the elements of V.

Definition 2.7 (Pattern matrix). For $\phi\colon \{0,1\}^t \to \mathbb{R}$, the (n,t,ϕ)*-pattern matrix* is the real matrix A given by

$$A = \Big[\phi(x|_V \ \oplus \ w)\Big]_{x \in \{0,1\}^n,\ (V,w) \in \mathscr{V}(n,t) \times \{0,1\}^t}.$$

In words, A is the matrix of size 2^n by $2^t(n/t)^t$ whose rows are indexed by strings $x \in \{0,1\}^n$, whose columns are indexed by pairs $(V,w) \in \mathscr{V}(n,t) \times \{0,1\}^t$, and whose entries are given by $A_{x,(V,w)} = \phi(x|_V \ \oplus \ w)$.

The logic behind the term "pattern matrix" is as follows: a mosaic arises from repetitions of a pattern in the same way that A arises from applications of ϕ to various subsets of the variables. We will need the following expression for the spectral norm of a pattern matrix [37, Thm. 4.3].

Theorem 2.8 (Sherstov). *Let $\phi\colon \{0,1\}^t \to \mathbb{R}$ be given. Let A be the (n,t,ϕ)-pattern matrix. Then*

$$\|A\| \ = \ \sqrt{2^{n+t}\left(\frac{n}{t}\right)^t} \ \max_{S \subseteq [t]} \left\{|\hat{\phi}(S)| \left(\frac{t}{n}\right)^{|S|/2}\right\}.$$

3 Reduction to Dense Predicates

For a predicate D, recall that $U(D)$ stands for its unbounded-error communication complexity. Let $U(n,k)$ denote the minimum $U(D)$ over the set of predicates $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ with $\deg(D) = k$. In this notation, our ultimate goal will be to bound $U(n,k)$ from below. This section takes a step in that direction.

First, we reduce the task of analyzing $U(n,k)$ to that of analyzing $U(n,\lceil \alpha n \rceil)$, where $\alpha \geqslant 1/4$. This focuses our efforts on high-degree predicates. We then further reduce the problem to *dense* predicates, i.e., high-degree predicates that change value at more or less even intervals in $\{0,1,\ldots,n\}$. These reductions are essential because dense predicates behave more predictably and are much easier to analyze than arbitrary predicates. Dense predicates will be the focus of all later sections.

We start with some preparatory work. For a predicate $D\colon \{0,1,\ldots,n\} \to \{0,1\}$, we define its *flip vector* $v = (v_0, v_1, \ldots, v_n) \in \{0,1\}^{n+1}$ by

$$v_i = \begin{cases} D(0) & \text{if } i = 0, \\ D(i) \oplus D(i-1) & \text{if } i = 1,2,\ldots,n. \end{cases}$$

Note that $\deg(D) = v_1 + v_2 + \cdots + v_n$. Also, if D_1 and D_2 are predicates with flip vectors $v^{(1)}$ and $v^{(2)}$, then $D_1 \oplus D_2$ has flip vector $v^{(1)} \oplus v^{(2)}$. Finally, given a predicate $D\colon \{0,1,\ldots,n\} \to \{0,1\}$, consider a derived predicate $D'\colon \{0,1,\ldots,m\} \to \{0,1\}$ given by $D'(t) \equiv D(t+\Delta)$, where $m \geqslant 1$ and $\Delta \geqslant 0$ are integers with $m + \Delta \leqslant n$. Then the flip vectors v and v' of D and D', respectively, are related as follows: $v' = (v_0 \oplus \cdots \oplus v_\Delta, v_{\Delta+1}, \ldots, v_{\Delta+m})$. From the standpoint of communication complexity, D' can be computed by hardwiring some inputs to a protocol for D:

$$D'\Big(\Big|x_1x_2\ldots x_m \bigwedge y_1y_2\ldots y_m\Big|\Big) = D\Big(\Big|x_1x_2\ldots x_m 1^\Delta 0^{n-m-\Delta} \bigwedge y_1y_2\ldots y_m 1^\Delta 0^{n-m-\Delta}\Big|\Big).$$

Therefore, $U(D') \leqslant U(D)$.

3.1 Reduction from Arbitrary to High-Degree Predicates

We start with a technical lemma. Consider a Boolean vector $v = (v_1, v_2, \ldots, v_n)$. We show that there is a subvector $(v_i, v_{i+1}, \ldots, v_j)$ that is reasonably far from both endpoints of v and yet contains many of the "1" bits present in v.

Lemma 3.1. *Let $v \in \{0,1\}^n$, $v \neq 0^n$. Put $k = v_1 + \cdots + v_n$. Then there are indices i, j with $i \leqslant j$ such that*

$$v_i + \cdots + v_j \geqslant \frac{1}{14}\,\frac{k}{1+\log(n/k)} \tag{3.1}$$

and

$$\min\{i-1, n-j\} \geqslant j - i. \tag{3.2}$$

Proof. By symmetry, we can assume that $v_1 + v_2 + \cdots + v_m \geqslant \frac{1}{2}k$ for some index $m \leqslant \lceil n/2 \rceil$. Let $\alpha \in (0, \frac{1}{2})$ be a parameter to be fixed later. Let $T \geqslant 0$ be the smallest integer such that

$$v_1 + v_2 + \cdots + v_{\lfloor m/2^T \rfloor} < (1-\alpha)^T (v_1 + v_2 + \cdots + v_m).$$

Clearly, $T \geqslant 1$. Since $v_1 + v_2 + \cdots + v_{\lfloor m/2^T \rfloor} \leqslant m/2^T$, we further obtain

$$1 \leqslant T \leqslant 1 + \frac{1 + \log(n/k)}{\log(2-2\alpha)}.$$

Now,

$$\begin{aligned} v_{\lfloor m/2^T \rfloor + 1} + \cdots + v_{\lfloor m/2^{T-1} \rfloor} &= \underbrace{(v_1 + \cdots + v_{\lfloor m/2^{T-1} \rfloor})}_{\geqslant (1-\alpha)^{T-1}(v_1+v_2+\cdots+v_m)} - \underbrace{(v_1 + \cdots + v_{\lfloor m/2^T \rfloor})}_{< (1-\alpha)^T (v_1+v_2+\cdots+v_m)} \\ &> \frac{1}{2}\alpha(1-\alpha)^{T-1}k \\ &\geqslant \frac{1}{2}\alpha(1-\alpha(T-1))k \\ &\geqslant \frac{1}{2}\alpha\left(1 - \alpha \cdot \frac{1+\log(n/k)}{\log(2-2\alpha)}\right)k. \end{aligned} \tag{3.3}$$

Set $\alpha = 0.23/(1 + \log(n/k))$, $i = \lfloor m/2^T \rfloor + 1$, and $j = \lfloor m/2^{T-1} \rfloor$. Then one easily verifies (3.2), while (3.1) is immediate from (3.3). □

We are now ready to prove the desired reduction to high-degree predicates. Throughout this proof, we will freely use the opening remarks of Section 3, often without mention.

Lemma 3.2 (Reduction from arbitrary to high-degree predicates). *For all integers n, k with $1 \leqslant k \leqslant n$,*

$$U(n,k) \geqslant \frac{5}{6} K \min_{\substack{m=K,\ldots,n, \\ 1/4 \leqslant \alpha \leqslant 1}} \left\{ \frac{1}{m} U(m, \lceil \alpha m \rceil) \right\},$$

where

$$K = \left\lceil \frac{1}{14} \frac{k}{1+\log(n/k)} \right\rceil.$$

Proof. Let $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ be any predicate with $\deg(D) = k$. As outlined in the Introduction, the intuition is to express some complicated (i.e., high-degree) predicate as the XOR of a small number of predicates derived from D. The details follow.

Let $v = (v_0, v_1, \ldots, v_n)$ be the flip vector of D. Apply Lemma 3.1 to $(v_1, \ldots, v_n)$ and let i, j be the resulting indices, $i \leqslant j$. Put $m = j - i + 1$. Since $v_i + \cdots + v_j \geqslant K$, we have

$$K \leqslant m \leqslant n. \tag{3.4}$$

Define predicates $D^{-(m-1)}, \ldots, D^0, \ldots, D^{m-1}$, each a mapping $\{0, 1, \ldots, m\} \to \{0, 1\}$, by $D^r(t) \equiv D(t + i - 1 + r)$. Then (3.2) shows that each of these predicates can be computed by taking a protocol for D and fixing all but the first m variables to appropriate values. Thus,

$$U(D) \geqslant U(D^r), \qquad r = -(m-1), \ldots, (m-1). \tag{3.5}$$

The flip vector of D^0 is $(*, v_i, \ldots, v_j)$ for some $* \in \{0, 1\}$, which means that $\deg(D^0) = v_i + \cdots + v_j$. If $\deg(D^0) > m/2$, then the theorem is true for D in view of (3.4) and (3.5). Thus, we can assume the contrary:

$$K \leqslant v_i + \cdots + v_j \leqslant \frac{1}{2}m. \tag{3.6}$$

If we write the flip vectors of $D^{-(m-1)}, \ldots, D^{m-1}$ one after another as row vectors, we obtain the following matrix A:

$$A = \begin{bmatrix}
* & * & * & * & * & \cdots & * & * & * & v_i \\
* & * & * & * & * & \cdots & * & * & v_i & v_{i+1} \\
* & * & * & * & * & \cdots & * & v_i & v_{i+1} & v_{i+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
* & v_i & v_{i+1} & v_{i+2} & v_{i+3} & \cdots & v_{j-3} & v_{j-2} & v_{j-1} & v_j \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
* & v_{j-2} & v_{j-1} & v_j & * & \cdots & * & * & * & * \\
* & v_{j-1} & v_j & * & * & \cdots & * & * & * & * \\
* & v_j & * & * & * & \cdots & * & * & * & *
\end{bmatrix}.$$

Let T be a suitably large integer to be named later, and let $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(T)}$ be independent random vectors, each selected uniformly from among the rows of A. Put $\mathbf{u} = \mathbf{u}^{(1)} \oplus \mathbf{u}^{(2)} \oplus \cdots \oplus \mathbf{u}^{(T)}$. We will index the columns of A and the components of all these vectors by $0, 1, \ldots, m$ (left to right). Let p_r stand for the fraction of 1s in the rth column of A. Every column of A, except the zeroth, contains $v_i, \ldots, v_j$ and some $m - 1$ additional values. One infers from (3.6) that

$$\frac{K}{2m} \leqslant p_r \leqslant \frac{3}{4}, \qquad r = 1, 2, \ldots, m. \tag{3.7}$$

Therefore,

$$\begin{aligned}\mathbf{E}\Big[(\mathbf{u})_1 + \cdots + (\mathbf{u})_m\Big] &= \sum_{r=1}^{m} \mathbf{E}\Big[(\mathbf{u}^{(1)})_r \ \oplus \ \cdots \ \oplus \ (\mathbf{u}^{(T)})_r\Big] \\ &= \sum_{r=1}^{m} \left(\frac{1}{2} - \frac{1}{2}(1-2p_r)^T\right) && \text{by Proposition 2.1} \\ &\geqslant \frac{1}{2}m\left(1 - \frac{1}{\mathrm{e}^{TK/m}}\right) && \text{by (3.6), (3.7).}\end{aligned}$$

Fix $T = \lceil (\ln 2)m/K \rceil$. Then by the last calculation, there is a vector $u = (u_0, u_1, \ldots, u_m)$ that satisfies $u_1 + \cdots + u_m \geqslant m/4$ and is the XOR of some T rows of A. In other words, there is a predicate $D^{\oplus}\colon \{0, 1, \ldots, m\} \to \{0, 1\}$ that satisfies $\deg(D^{\oplus}) \geqslant m/4$ and is the XOR of some $T \leqslant \frac{6m}{5K}$ predicates from among $D^{-(m-1)}, \ldots, D^{m-1}$. This completes the proof in view of (3.5) and Proposition 2.6. □

3.2 Reduction from High-Degree to Dense Predicates

The proof in this section uses the same setup as Lemma 3.2, except the argument is now more involved. The reason is that the previous averaging argument is not strong enough to yield a dense predicate, which is a highly structured object. To overcome this, we recast the previous argument as a random walk on $\mathbb{Z}_2^n$ and show that it mixes rapidly. In particular, we will need the following lemma that bounds the mixing time of a random walk [30, Lem. 1]; for an English translation, see Jukna [12, Lem. 24.3].

Lemma 3.3 (Razborov). *Fix a probability distribution μ on $\{0,1\}^n$. Let $\{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\}$ be a basis for $\{0,1\}^n$ as a vector space over GF(2). Put*

$$p = \min\left\{\mu(0^n), \mu(v^{(1)}), \mu(v^{(2)}), \ldots, \mu(v^{(n)})\right\}.$$

Let $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(T)}$ be independent random vectors, each distributed according to μ. Then for every $v \in \{0,1\}^n$,

$$\left|\mathbf{P}\left[\mathbf{u}^{(1)} \oplus \cdots \oplus \mathbf{u}^{(T)} = v\right] - 2^{-n}\right| \leqslant \mathrm{e}^{-2Tp}.$$

We are ready to formally define dense predicates and give the promised reduction.

Definition 3.4 (Dense predicate). Let n, b be positive integers and $d \geqslant 0$ a real number. A predicate D is called (n, b, d)-*dense* if D is a predicate $\{0, 1, \ldots, n\} \to \{0, 1\}$ with flip vector $(v_0, v_1, \ldots, v_n)$ satisfying

$$v_{rb+1} + v_{rb+2} + \cdots + v_{(r+1)b} \geqslant d, \qquad r = 0, 1, 2, \ldots, \left\lfloor \frac{n}{b} \right\rfloor - 1.$$

Lemma 3.5 (Reduction from high-degree to dense predicates). *Let $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ be a predicate with $\deg(D) \geqslant \frac{1}{4}n$. Let b be any integer with $1 \leqslant b \leqslant \frac{1}{350}n$. Then*

$$U(D) \geqslant \frac{b}{n \log n}\, U(D'),$$

where D' is some $(m, \lceil \log n \rceil b, \frac{1}{700}b)$-dense predicate and $\frac{1}{350}n \leqslant m \leqslant n$.

Proof. Let $(v_0, v_1, \ldots, v_n)$ be the flip vector of D. Apply Lemma 3.1 to $(v_1, \ldots, v_n)$ and let i, ℓ be the resulting indices ($i \leqslant \ell$). It will be convenient to work with a somewhat smaller subvector $v = (v_i, \ldots, v_j)$, where we define $j \in \{i, \ldots, \ell\}$ to be the largest integer so that $b \mid (j - i + 1)$. Since $b \leqslant \frac{1}{350}n$ and $v_i + \cdots + v_\ell \geqslant \frac{1}{168}n$, this gives:

$$v_i + \cdots + v_j \geqslant \frac{1}{350}n. \tag{3.8}$$

Defining $m = j - i + 1$, we infer that $\frac{1}{350}n \leqslant m \leqslant n$, as desired. We view $v = (v_i, \ldots, v_j)$ as composed of consecutive blocks, each b bits long:

$$v = \Bigg(\underbrace{\boxed{v_i, \ldots, v_{i+b-1}}}_{\text{block 1}}, \underbrace{\boxed{v_{i+b}, \ldots, v_{i+2b-1}}}_{\text{block 2}}, \cdots\cdots, \underbrace{\boxed{v_{j-b+1}, \ldots, v_j}}_{\text{block } m/b} \Bigg). \tag{3.9}$$

For $r = 1, 2, \ldots, b$, define the *rth layer of v*, denoted $z^{(r)}$, to be the vector obtained by taking the rth component from each of the above blocks:

$$z^{(r)} = (v_{i-1+r}, v_{i-1+b+r}, \ldots, v_{j-b+r}) \in \{0, 1\}^{m/b}.$$

We say of a layer z that it is *perfect* if it does not have $\lceil \log n \rceil$ consecutive components equal to 0. If more than $\frac{1}{700}b$ of the layers are perfect, take D' to be the predicate with flip vector $(v_0 \oplus \cdots \oplus v_{i-1}, v_i, \ldots, v_j)$. Clearly, D' is $(m, \lceil \log n \rceil b, \frac{1}{700}b)$-dense. Furthermore, $U(D') \leqslant U(D)$, by the same argument as in Lemma 3.2. As a result, the theorem holds in this case.

Thus, we may assume that at least $(1 - \frac{1}{700})b$ of the layers are not perfect. In view of (3.8), at most $(1 - \frac{1}{350})b$ layers can be zero vectors. Therefore, $\frac{1}{700}b$ or more

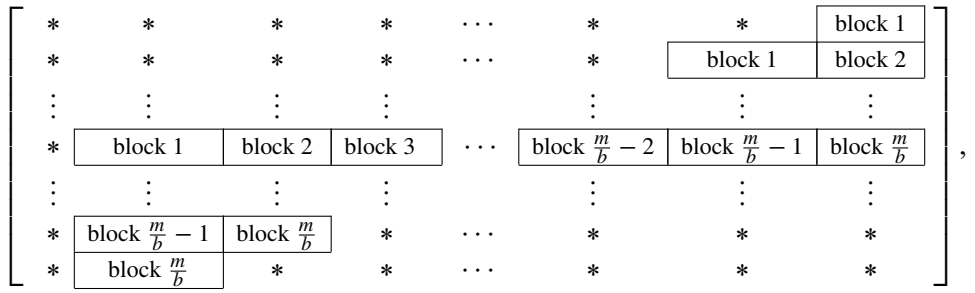
layers are nonzero *and* not perfect. These are the only layers we will consider in the remainder of the proof.

Define predicates $D^{-(m-b)}, D^{-(m-2b)}, \ldots, D^{-b}, D^0, D^b, \ldots, D^{m-2b}, D^{m-b}$, each a mapping $\{0, 1, \ldots, m\} \to \{0, 1\}$, by $D^r(t) \equiv D(t + i - 1 + r)$. These are a subset of the predicates from the proof of Lemma 3.2, and again

$$U(D) \geqslant U(D^r) \qquad \text{for each } r. \tag{3.10}$$

Writing the flip vectors of these predicates one after another as row vectors yields the following matrix B:

$B =$

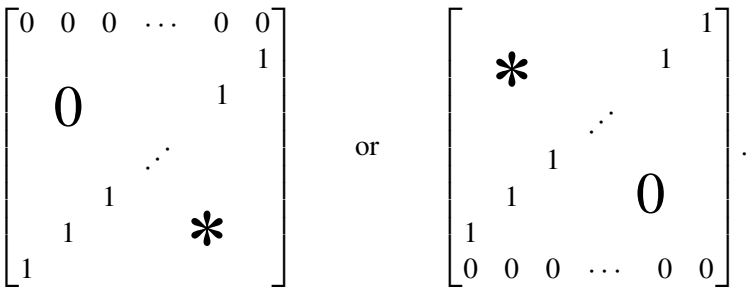


where the blocks refer to the partition in (3.9). Let T be a suitably large integer to be named later, and let $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(T)}$ be independent random vectors, each selected uniformly from among the rows of B. Put $\mathbf{u} = \mathbf{u}^{(1)} \oplus \mathbf{u}^{(2)} \oplus \cdots \oplus \mathbf{u}^{(T)}$. We will index the columns of B and the components of $\mathbf{u}$ by $0, 1, \ldots, m$ (left to right). Key to analyzing the distribution of $\mathbf{u}$ is the following claim.

Claim 3.5.1. *Let $T \geqslant (m/b) \ln n$. Let $\Delta \in \{1, 2, \ldots, b\}$ be such that the layer $z^{(\Delta)}$ is nonzero and not perfect. Let $s \in \{0, b, 2b, 3b, \ldots\}$ be such that $s + \lceil \log n \rceil b \leqslant m$. Then*

$$\mathbf{P}\Big[(\mathbf{u})_{s+\Delta} = (\mathbf{u})_{s+b+\Delta} = \cdots = (\mathbf{u})_{s+(\lceil \log n \rceil - 1)b+\Delta} = 0\Big] \leqslant \frac{2}{n}.$$

Proof. Let B' be the matrix whose columns are the following columns of B: $s + \Delta, s + b + \Delta, \ldots, s + (\lceil \log n \rceil - 1)b + \Delta$, in that order. Since $z^{(\Delta)}$ is nonzero and not perfect, $z^{(\Delta)}$ has $\lceil \log n \rceil + 1$ consecutive components with values either $0, 0, \ldots, 0, 1$ or $1, 0, 0, \ldots, 0$. Consequently, B' must contain one of the following submatrices, each of size $(\lceil \log n \rceil + 1) \times \lceil \log n \rceil$:



The claim now follows from Lemma 3.3, since $2^{-\lceil \log n \rceil} + \mathrm{e}^{-2T \cdot \frac{b}{2m}} \leqslant 2/n$. □

We return to the proof of the lemma. Fix $T = \lceil (m/b) \ln n \rceil$. Let $s = 0$ and apply Claim 3.5.1 with every $\Delta \in \{1, 2, \ldots, b\}$ for which the layer $z^{(\Delta)}$ is nonzero and not perfect. Since there are at least $\frac{1}{700}b$ such choices for Δ, we conclude by the union bound that

$$\mathbf{P}\left[(\mathbf{u})_1 + (\mathbf{u})_2 + \cdots + (\mathbf{u})_{\lceil \log n \rceil b} < \frac{1}{700} b \right] \leqslant b \cdot \frac{2}{n}.$$

The same calculation applies to the next set of $\lceil \log n \rceil b$ components of $\mathbf{u}$ (i.e., $s = \lceil \log n \rceil b$), and so on. Applying a union bound across all these $m/(\lceil \log n \rceil b)$ calculations, we find that with probability

$$1 - \frac{m}{\lceil \log n \rceil b} \left(b \cdot \frac{2}{n} \right) > 0,$$

the predicate whose flip vector is $\mathbf{u}$ is $(m, \lceil \log n \rceil b, \frac{1}{700}b)$-dense. Fix any such predicate D'. Since D' is the XOR of $T \leqslant (n \log n)/b$ predicates from among $D^{-(m-b)}, \ldots, D^{m-b}$, the lemma follows by (3.10) and Proposition 2.6. □

4 Univariate Approximation with Clusters of Nodes

Crucial to our study of dense predicates are certain approximation problems to which they give rise. Roughly speaking, the hardness of such an approximation problem for low-degree polynomials translates into the communication hardness of the associated predicate. This section carries out the first part of the program, namely, showing that the approximation task at hand is hard for low-degree polynomials. We examine this question in its basic mathematical form, with no extraneous considerations to obscure our view. How communication fits in this picture will become clear in the next two sections.

For a finite set $X \subset \mathbb{R}$, a function $f : X \to \mathbb{R}$, and an integer $r \geqslant 0$, define

$$\epsilon^*(f, X, r) = \min_{p \in P_r} \max_{x \in X} |p(x) - f(x)|.$$

In words, $\epsilon^*(f, X, r)$ is the least error (in the uniform sense) to which a degree-r polynomial can approximate f on X. The following well-known fact from approximation theory is useful in estimating this error.

Fact 4.1 (see, e.g., [34, Thm. 1.15]). *Let $X = \{x_1, x_2, \ldots, x_{r+2}\}$ be given reals, where $x_1 < x_2 < \cdots < x_{r+2}$. Let $f : X \to \mathbb{R}$ be given. Put*

$$\omega(x) = (x - x_1)(x - x_2) \cdots (x - x_{r+2}).$$

Then

$$\epsilon^*(f, X, r) = \frac{\left| \sum_{i=1}^{r+2} [f(x_i)/\omega'(x_i)] \right|}{\sum_{i=1}^{r+2} [1/|\omega'(x_i)|]}.$$

To develop some intuition for the work in this section, consider the following approximation problem. Let $f : \{0, 1, \ldots, n\} \to \{0, 1\}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x = \lfloor n/2 \rfloor, \\ 0 & \text{otherwise.} \end{cases}$$

It is well-known that any polynomial that approximates f within $1/3$ has degree $\Omega(n)$. For example, this follows from work by Paturi [27]. The approximation problem of interest to us is similar, except that our points need not be as evenly spaced as $0, 1, \ldots, n$ but rather may form clusters. As a result, Paturi's results and methods do not apply, and we approach this question differently, using the first-principles formula of Fact 4.1. Specifically, our main result in this section is as follows.

Lemma 4.2 (Inapproximability by low-degree polynomials). *Let positive integers L, d and a real number $B \geqslant d$ be given. Let $\{x_{ij} : i = 1, \ldots, L; j = 1, \ldots, d\}$ be a set of Ld distinct reals, where $x_{ij} \in [(i-1)B, iB]$ and*

$$|x_{ij} - x_{i'j'}| \geqslant 1 \qquad \text{for } (i, j) \neq (i', j'). \tag{4.1}$$

Let $x_0 \in [\frac{1}{4}LB, \frac{3}{4}LB]$. Then any polynomial p with

$$p(x_0) = 1, \qquad |p(x_{ij})| < \frac{1}{2}\left(\frac{1}{LB}\right)^{4d+1} \qquad \text{for all } i, j$$

has degree at least $\left(\frac{1}{2}L - 1\right)d$.

Proof. Define $f(x)$ by

$$f(x) = \begin{cases} 1 & \text{if } x = x_0, \\ 0 & \text{if } x = x_{ij} \text{ for some } i,\, j. \end{cases}$$

By symmetry, we can assume that $x_0 \in [\frac{1}{4}LB, \frac{1}{2}LB]$. Fix an integer $\ell \leqslant \lceil \frac{1}{2}L \rceil$ so that $x_0 \in [(\ell-1)B, \ell B]$. Put

$$X = \{x_0\} \cup \{x_{ij} : i = 1, \ldots, 2\ell - 1; \quad j = 1, \ldots, d\}.$$

With $\omega(x) = \prod_{y \in X}(x - y)$, Fact 4.1 implies that

$$\epsilon^*(f, X, |X| - 2) \geqslant \frac{1}{|X|} \, \frac{\min_{x \in X} |\omega'(x)|}{|\omega'(x_0)|}. \tag{4.2}$$

We proceed to estimate the denominator and numerator of (4.2). Since x_0 is distinct from each x_{ij}, the quantity

$$\delta = \min_{\substack{i=1,\ldots,2\ell-1, \\ j=1,\ldots,d}} |x_0 - x_{ij}|$$

satisfies $\delta > 0$. We have:

$$\begin{aligned} |\omega'(x_0)| \;&=\; \prod_{j=1}^{d} \prod_{i=1}^{2\ell-1} |x_0 - x_{ij}| \;\leqslant\; \delta \prod_{j=1}^{d} \prod_{i=1}^{2\ell-1} B \underbrace{\left\lceil \frac{|x_0 - x_{ij}|}{B} \right\rceil}_{\leqslant |i-\ell|+1} \\ &\leqslant\; \delta \cdot \left(\ell!\, \ell!\, B^{2\ell-1}\right)^d. \end{aligned} \tag{4.3}$$

On the other hand, every $x_{i'j'} \in X$ satisfies:

$$\begin{aligned} |\omega'(x_{i'j'})| \;&=\; \prod_{x \in X \setminus \{x_{i'j'}\}} |x - x_{i'j'}| \\ &\geqslant\; \delta \prod_{j=1}^{d} \prod_{\substack{i=1,\ldots,2\ell-1 \\ i \notin \{i'-1, i', i'+1\}}} |x_{ij} - x_{i'j'}| \qquad \text{by (4.1)} \\ &\geqslant\; \delta \prod_{j=1}^{d} \prod_{\substack{i=1,\ldots,2\ell-1 \\ i \notin \{i'-1, i', i'+1\}}} B \underbrace{\left\lfloor \frac{|x_{ij} - x_{i'j'}|}{B} \right\rfloor}_{\geqslant |i-i'|-1} \\ &\geqslant\; \delta \cdot \left(\frac{\ell!\, \ell!\, B^{2\ell-4}}{\ell^4} \right)^d. \end{aligned} \tag{4.4}$$

Now (4.2) yields, in view of (4.3) and (4.4):

$$\epsilon^*(f, X, |X| - 2) \geqslant \frac{1}{2}\left(\frac{1}{LB}\right)^{4d+1},$$

which concludes the proof since $|X| \geqslant \left(\frac{1}{2}L - 1\right)d + 1$. □

5 Key Analytic Property of Dense Predicates

We now transition to the final ingredient of our proof, *smooth orthogonalizing distributions* for a given predicate D. This informal term refers to a distribution on $\{0,1\}^n$ that does not put too little weight on any point (the *smooth* part) and under which $(-1)^{D(x_1+\cdots+x_n)}$ is approximately orthogonal to all low-degree characters χ_S (the *orthogonalizing* part). Our task is to establish the existence of such distributions for every dense predicate. Once this is accomplished, we will be able to treat a dense predicate as if it were the familiar PARITY function (whose defining analytic property is precisely its orthogonality to the lower-order characters under the uniform distribution). Crucial to the development below will be the inapproximability result proved in Section 4.

For a polynomial p, a predicate $D\colon \{0,1,\ldots,n\} \to \{0,1\}$, and a number $N > 0$, define the *advantage* of p in computing D by

$$\mathrm{adv}(p, N, D) = N \min_{t=0,\ldots,n}\left\{(-1)^{D(t)}p(t)\right\} + \sum_{t=0}^{n}\frac{\binom{n}{t}}{2^n}(-1)^{D(t)}p(t).$$

This quantity is conceptually close to the correlation of p and D with respect to the binomial distribution. There is a substantial difference, however: if p and D differ in sign at some point, this causes a penalty term to be subtracted. We will be interested in values $N \gg 1$, when even a single error of p results in a large penalty. Define

$$\mathrm{adv}_r(N, D) = \max_p \mathrm{adv}(p, N, D),$$

where the maximization is over $p \in P_r$ with $|p(t)| \leqslant 1$ for $t = 0, 1, \ldots, n$. As we now show, this quantity is closely related to smooth orthogonalizing distributions for D.

Theorem 5.1 (Smooth distributions vs. approximation by polynomials). *Fix a predicate $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ and an integer $r \geqslant 0$. Then for every $N > 1$, there is a distribution μ on $\{0,1\}^n$ such that $\mu(x) \geqslant \frac{1}{2^n N}$ for each x and*

$$\left|\mathop{\mathbf{E}}_{x}\left[(-1)^{D(|x|)}\mu(x)\,\chi_S(x)\right]\right| \leqslant \frac{1}{2^n N}\mathrm{adv}_r(N-1, D) \qquad \text{for } |S| \leqslant r.$$

Proof. Put $f(x) = (-1)^{D(|x|)}$ and consider the following linear program:

$$\begin{array}{lll}
\text{variables:} & \mu(x) \text{ for all } x; \quad \epsilon & \\
\text{minimize:} & \epsilon & \\
\text{subject to:} & \left| \sum_{x \in \{0,1\}^n} \mu(x) f(x) \chi_S(x) \right| \leqslant \epsilon & \text{for } |S| \leqslant r, \\
& \sum_{x \in \{0,1\}^n} \mu(x) = 1, & \\
& \mu(x) \geqslant \dfrac{1}{2^n N} & \text{for each } x.
\end{array} \tag{LP1}$$

It suffices to show that the optimum of this program is at most $\frac{1}{N} \operatorname{adv}_r(N-1, D)$. For this, we pass to the dual:

$$\begin{array}{lll}
\text{variables:} & \alpha_S \text{ (for } |S| \leqslant r); \quad \xi_x \text{ (for all } x); \quad \Delta & \\
\text{maximize:} & \dfrac{1}{N} \left((N-1)\Delta + \dfrac{1}{2^n} \sum_{x \in \{0,1\}^n} (\Delta + \xi_x) \right) & \\
\text{subject to:} & f(x) \sum_{|S| \leqslant r} \alpha_S \chi_S(x) \geqslant \Delta + \xi_x & \text{for all } x, \\
& \sum_{|S| \leqslant r} |\alpha_S| \leqslant 1, & \\
& \alpha_S \in \mathbb{R} & \text{for } |S| \leqslant r, \\
& \xi_x \geqslant 0 & \text{for all } x, \\
& \Delta \in \mathbb{R}. &
\end{array} \tag{LP2}$$

The dual programs (LP1) and (LP2) are both feasible and thus have the same finite optimum. Therefore, our task reduces to proving that the optimum of (LP2) is at most $\frac{1}{N} \operatorname{adv}_r(N-1, D)$. Fix an optimal solution to (LP2). Then

$$f(x) \sum_{|S| \leqslant r} \alpha_S \chi_S(x) = \Delta + \xi_x \qquad \text{for all } x, \tag{5.1}$$

since in case of a strict inequality ($>$) we could increase the corresponding variable ξ_x by a small amount to obtain a feasible solution with greater value. Furthermore, we claim that

$$\Delta = \min_{x \in \{0,1\}^n} \left\{ f(x) \sum_{|S| \leqslant r} \alpha_S \chi_S(x) \right\}. \tag{5.2}$$

Indeed, let m stand for the right-hand side of (5.2). Then $\Delta \leqslant m$ because each ξ_x is nonnegative. It remains to show that $\Delta \geqslant m$. If we had $\Delta < m$, then (5.1) would imply that $\xi_x \geqslant m - \Delta$ for all x. As a result, we could obtain a new feasible solution $\xi'_x = \xi_x + (\Delta - m)$ and $\Delta' = m$. This new solution satisfies $\Delta' + \xi'_x = \Delta + \xi_x$ for all x. Moreover, $\Delta' > \Delta$, which results in a greater objective value and yields the desired contradiction. In summary, $\Delta = m$.

In view of (5.1) and (5.2), the optimum of (LP2) is

$$\frac{1}{N} \max_{\phi} \left\{ (N-1) \min_x \{ f(x)\phi(x) \} \;\; + \;\; \frac{1}{2^n} \sum_x f(x)\phi(x) \right\}, \tag{5.3}$$

where the maximization is over functions ϕ of the form

$$\phi(x) = \sum_{|S| \leqslant r} \alpha_S \, \chi_S(x), \qquad \text{where } \sum_{|S| \leqslant r} |\alpha_S| \leqslant 1. \tag{5.4}$$

Fix ϕ that optimizes (5.3). By (5.4),

$$\max_{x \in \{0,1\}^n} |\phi(x)| \leqslant 1.$$

Put

$$\phi_{\mathrm{sym}}(x) = \frac{1}{n!} \sum_{\sigma \in S_n} \phi(x_{\sigma(1)}, \ldots, x_{\sigma(n)}).$$

Since f is symmetric, ϕ and ϕ_{sym} have the same objective value in (5.3). By the symmetrization argument (Proposition 2.3), there is a univariate polynomial $p \in P_r$ with

$$\phi_{\mathrm{sym}}(x) = p(x_1 + \cdots + x_n) \qquad \text{for all } x \in \{0,1\}^n.$$

For $t = 0, 1, \ldots, n$,

$$|p(t)| = |p(\underbrace{1 + \cdots + 1}_{t \text{ times}} + 0 + \cdots + 0)| \leqslant \max_{x \in \{0,1\}^n} |\phi_{\mathrm{sym}}(x)| \leqslant \max_{x \in \{0,1\}^n} |\phi(x)| \leqslant 1.$$

Replacing $\phi(x)$ by $p(x_1 + \cdots + x_n)$ in (5.3), we see that the optimum of (LP2) is at most

$$\frac{1}{N} \max_{p} \left\{ (N-1) \min_{t=0,\ldots,n} \left\{ (-1)^{D(t)} p(t) \right\} \;\; + \;\; \frac{1}{2^n} \sum_{t=0}^{n} \binom{n}{t} (-1)^{D(t)} p(t) \right\},$$

where the maximization is over $p \in P_r$ with $|p(t)| \leqslant 1$ for $t = 0, 1, \ldots, n$. This latter quantity is $\frac{1}{N} \mathrm{adv}_r(N-1, D)$, by definition. □

Theorem 5.1 states that a smooth orthogonalizing distribution for D exists whenever low-degree polynomials have negligible advantage in computing D. Accordingly, we proceed to examine the advantage achievable by low-degree polynomials.

Lemma 5.2 (Each dense predicate induces a hard approximation problem). *Let D be an $(n, B, 2d+1)$-dense predicate, where n, B, d are positive integers. Assume that $\mathrm{adv}_r(N, D) \geqslant n2^{-n/6}$, where $r < \deg(D)$ and $N > 0$ are given. Then there are $\lfloor \frac{n}{B} \rfloor d$ distinct reals $\{x_{ij} : i = 1, \ldots, \lfloor \frac{n}{B} \rfloor;\ j = 1, \ldots, d\}$ and a polynomial $p \in P_r$ such that:*

$$
\begin{aligned}
&x_{ij} \in [(i-1)B, iB] && \text{for all } i, j,\\
&|x_{ij} - x_{i'j'}| \geqslant 1 && \text{for all } (i, j) \neq (i', j'),\\
&|p(x_{ij})| \leqslant \sqrt{n}/N && \text{for all } i, j,\\
&p(x_0) = 1 && \text{for some } x_0 \in [\tfrac{1}{4}n, \tfrac{3}{4}n].
\end{aligned}
$$

Proof. Fix $q \in P_r$ with $|q(t)| \leqslant 1$ for $t = 0, 1, \ldots, n$ and $\mathrm{adv}(q, N, D) = \mathrm{adv}_r(N, D)$. Fix $k \in \{0, 1, \ldots, n\}$ with

$$\binom{n}{k}(-1)^{D(k)}q(k) = \max_{t=0,\ldots,n}\left\{\binom{n}{t}(-1)^{D(t)}q(t)\right\}.$$

Since $\deg(q) < \deg(D)$, the quantity $\binom{n}{t}(-1)^{D(t)}q(t)$ is positive for at most n values of $t = 0, 1, \ldots, n$. Therefore,

$$\mathrm{adv}(q, N, D) \leqslant n \cdot \frac{\binom{n}{k}}{2^n}(-1)^{D(k)}q(k) \leqslant n \cdot \frac{\binom{n}{k}}{2^n}.$$

Recalling that $\mathrm{adv}(q, N, D) \geqslant n2^{-n/6}$, we infer that $\frac{1}{4}n \leqslant k \leqslant \frac{3}{4}n$. Put

$$p(t) = \frac{1}{q(k)}\,q(t).$$

Taking $x_0 = k$, we have $\frac{1}{4}n \leqslant x_0 \leqslant \frac{3}{4}n$ and $p(x_0) = 1$, as desired. It remains to find the points x_{ij}. For this, we need the following claim.

Claim 5.2.1. *Let a, b be integers with $a < b$ and $D(a) \neq D(b)$. Then $|p(\xi)| \leqslant \sqrt{n}/N$ for some $\xi \in [a, b]$.*

Proof. If q vanishes at some point in $[a, b]$, we are done. In the contrary case, q is nonzero and has the same sign at every point of $[a, b]$, which means that either $q(a)(-1)^{D(a)} < 0$ or $q(b)(-1)^{D(b)} < 0$. Since $\mathrm{adv}(q, N, D) \geqslant 0$, we have:

$$\begin{aligned}\min\{|q(a)|, |q(b)|\} &\leqslant \frac{n}{N} \max_{t=0,\ldots,n} \left\{ \frac{\binom{n}{t}}{2^n} (-1)^{D(t)} q(t) \right\} = \frac{n}{N} \cdot \frac{\binom{n}{k}}{2^n} \cdot |q(k)| \\ &\leqslant \frac{\sqrt{n}}{N} |q(k)|,\end{aligned}$$

and hence $\min\{|p(a)|, |p(b)|\} \leqslant \sqrt{n}/N$. □

Fix an integer $i = 1, 2, \ldots, \lfloor \frac{n}{B} \rfloor$. Since D is $(n, B, 2d + 1)$-dense, D changes value at least $2d$ times in $[(i - 1)B + 1, iB]$. As a result, there are at least d pairs of integers $(a_1, b_1), \ldots, (a_d, b_d)$ with

$$D(a_1) \neq D(b_1), \quad D(a_2) \neq D(b_2), \quad \ldots, \quad D(a_d) \neq D(b_d)$$

and

$$(i - 1)B + 1 \leqslant a_1 < b_1 < a_2 < b_2 < \cdots < a_d < b_d \leqslant iB.$$

In view of Claim 5.2.1, this provides the desired d points in $[(i-1)B+1, iB]$. □

We have reached the main result of this section.

Theorem 5.3 (Smooth orthogonalizing distributions for dense predicates). *Let D be an $(n, B, 2d + 1)$-dense predicate, where n, B, d are positive integers with $B \mid n$ and $n \geqslant 3B$. Then there is a distribution μ on $\{0, 1\}^n$ such that:*

$$\begin{aligned}&\mu(x) \geqslant \frac{1}{2^n} \frac{1}{3n^{4d+1.5}} && \text{for each } x, \\ &\left| \mathop{\mathbf{E}}_{x} \left[(-1)^{D(|x|)} \mu(x) \chi_S(x) \right] \right| \leqslant 2^{-7n/6} && \text{for } |S| < \frac{nd}{6B}.\end{aligned}$$

Proof. Put $N = 3n^{4d+1.5}$. In view of Theorem 5.1, it is sufficient to show that $\mathrm{adv}_r(N - 1, D) < n2^{-n/6}$ for all $r < \frac{nd}{6B}$. So assume, for the sake of contradiction, that $\mathrm{adv}_r(N - 1, D) \geqslant n2^{-n/6}$ for some $r < \frac{nd}{6B}$. Since $\deg(D) \geqslant \frac{n}{B}(2d + 1)$, we have $r < \deg(D)$. Thus, Lemma 5.2 is applicable and yields $\frac{nd}{B}$ distinct reals $\{x_{ij} : i = 1, \ldots, \frac{n}{B};\ j = 1, \ldots, d\}$ and a polynomial $p \in P_r$ such that:

$$\begin{aligned}&x_{ij} \in [(i - 1)B, iB] && \text{for all } i, j, \\ &|x_{ij} - x_{i'j'}| \geqslant 1 && \text{for all } (i, j) \neq (i', j'), \\ &|p(x_{ij})| < \tfrac{1}{2} \left(\tfrac{1}{n}\right)^{4d+1} && \text{for all } i, j, \\ &p(x_0) = 1 && \text{for some } x_0 \in [\tfrac{1}{4}n, \tfrac{3}{4}n].\end{aligned}$$

Applying Lemma 4.2 with $L = \frac{n}{B}$, we infer that $r \geqslant \left(\frac{1}{2}\frac{n}{B} - 1\right) d$, which yields $r \geqslant \frac{nd}{6B}$ since $\frac{n}{B} \geqslant 3$. We have reached the desired contradiction to $r < \frac{nd}{6B}$. □

6 Proof of the Main Result

This section consolidates the preceding developments into our main result, a near-optimal lower bound on the unbounded-error communication complexity of every symmetric function. As outlined earlier, we will first solve this problem for dense predicates and then extend our work to the general case via the reductions of Section 3.

Theorem 6.1 (Communication complexity of dense predicates). *Let $\alpha > 0$ be a sufficiently small absolute constant. Let D be an $(m, b\lceil \log n \rceil, \frac{1}{700}b)$-dense predicate, where $\frac{1}{350}n \leqslant m \leqslant n$ and $b = \lfloor \alpha n / \log^2 n \rfloor$. Then*

$$U(D) \geqslant \Omega\left(\frac{n}{\log n}\right).$$

Proof. Throughout the proof we will, without mention, use the assumption that n is large enough. This will simplify the setting of parameters, the manipulation of floors and ceilings, and generally make the proof easier to follow.

Fix an integer $v \in [\frac{1}{8}m, \frac{1}{4}m]$ with $b\lceil \log n \rceil \mid v$. Clearly, $v \gg 3b\lceil \log n \rceil$. Define $D' \colon \{0, 1, \ldots, v\} \to \{0, 1\}$ by $D'(t) \equiv D(t)$. Since D' is $(v, b\lceil \log n \rceil, \frac{1}{700}b)$-dense, Theorem 5.3 provides a distribution μ on $\{0, 1\}^v$ with

$$\mu(z) \geqslant 2^{-v} \cdot 2^{-\alpha n / 350 \log n} \qquad \text{for each } z \in \{0, 1\}^v, \tag{6.1}$$

$$\left| \mathop{\mathbf{E}}_{z} \left[(-1)^{D(|z|)} \mu(z) \chi_S(z) \right] \right| \leqslant 2^{-7v/6} \qquad \text{for } |S| < \frac{v}{6 \cdot 1401 \lceil \log n \rceil}. \tag{6.2}$$

Define $\phi \colon \{0, 1\}^v \to \mathbb{R}$ by $\phi(z) = (-1)^{D(|z|)} \mu(z)$. Restating (6.2),

$$|\hat{\phi}(S)| \leqslant 2^{-7v/6} \qquad \text{for } |S| < \frac{v}{6 \cdot 1401 \lceil \log n \rceil}. \tag{6.3}$$

Furthermore, Proposition 2.2 reveals that

$$\max_{S \subseteq [v]} |\hat{\phi}(S)| \leqslant 2^{-v}. \tag{6.4}$$

Let A be the $(2v, v, 8^{-v}\phi)$-pattern matrix. By (6.3), (6.4), and Theorem 2.8,

$$\|A\| \leqslant 4^{-v} \cdot 2^{-v / 12 \cdot 1401 \lceil \log n \rceil}. \tag{6.5}$$

By (6.1), every entry of A has absolute value at least $16^{-v}\, 2^{-\alpha n/350 \log n}$. Combining this observation with (6.5) and Theorem 2.5,

$$\mathrm{rk}_{\pm}(A) \;\geqslant\; 2^{v/12 \cdot 1401 \lceil \log n \rceil}\;\; 2^{-\alpha n/350 \log n}.$$

Recall that $v \geqslant \frac{1}{8} m \geqslant \frac{1}{8 \cdot 350} n$. Hence, for a suitably small constant $\alpha > 0$,

$$\mathrm{rk}_{\pm}(A) \geqslant 2^{\Omega(n/\log n)}.$$

It remains to relate the sign-rank of A to the communication complexity of D. Let F be the $(2v, v, f)$-pattern matrix, where $f(z) = (-1)^{D(|z|)}$. Then $\mathrm{rk}_{\pm}(A) = \mathrm{rk}_{\pm}(F)$ because A and F have the same sign pattern. But F is a submatrix of the communication matrix of D, namely,

$$M = \Big[(-1)^{D(|x \wedge y|)}\Big]_{x \in \{0,1\}^m, y \in \{0,1\}^m}.$$

Thus,

$$\mathrm{rk}_{\pm}(M) \geqslant \mathrm{rk}_{\pm}(F) = \mathrm{rk}_{\pm}(A) \geqslant 2^{\Omega(n/\log n)}.$$

In view of Theorem 2.4, the proof is complete. □

Corollary 6.2 (Communication complexity of high-degree predicates). *Let* $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ *be a predicate with* $\deg(D) \geqslant \frac{1}{4} n$. *Then*

$$U(D) \geqslant \Omega\left(\frac{n}{\log^4 n}\right).$$

Proof. Immediate from Lemma 3.5 and Theorem 6.1. □

At last, we arrive at the main result of this paper, cf. Theorem 1.1 in the Introduction.

Theorem 6.3 (Main Result). *Let* $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ *be a nonconstant predicate,* $k = \deg(D)$. *Then*

$$\Omega\left(\frac{k}{\{1 + \log(n/k)\} \log^4 n}\right) \leqslant U(D) \leqslant O\left(k \log \frac{2n}{k}\right).$$

Proof. The lower bound is immediate from Lemma 3.2 and Corollary 6.2. To prove the upper bound, fix $p \in P_k$ with $\operatorname{sgn}(p(t)) = (-1)^{D(t)}$ for $t = 0, 1, \ldots, n$. Put

$$M = \Big[(-1)^{D(|x \wedge y|)}\Big]_{x,y}, \qquad R = \Big[p(x_1 y_1 + \cdots + x_n y_n)\Big]_{x,y},$$

where the indices run as usual: $x, y \in \{0,1\}^n$. Then $M_{xy} R_{xy} > 0$ for all x and y. Thus, the sign-rank of M does not exceed $\sum_{i=0}^{k} \binom{n}{i}$. In view of Theorem 2.4, this completes the proof. □

Remark 6.4. Immediate consequences of Theorem 6.3 are near-tight lower bounds on the size of threshold-of-majority and majority-of-threshold circuits for every function $f(x, y) = D(|x \wedge y|)$, where $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ is a given predicate. Similarly, Theorem 6.3 yields near-tight lower bounds on the dimension complexity of every concept class $\mathscr{C}_D = \{f_{D,y} : y \in \{0,1\}^n\}$, where $f_{D,y}(x) = D(|x \wedge y|)$ for a fixed predicate $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$. These applications follow from well-known black-box arguments involving sign-rank [8, Lem. 5], and we do not spell them out here.

On Logarithmic Factors in Theorem 6.3

It is natural to wonder whether the logarithmic factors in Theorem 6.3 can be eliminated. The answer varies from one predicate to another. There are indeed predicates $D\colon \{0, 1, \ldots, n\} \to \{0, 1\}$ for which $U(D) = \Theta(\deg(D))$. For example, the conjunction predicate, given by $\mathrm{AND}_n(t) = 1 \Leftrightarrow t = n$, has degree 1 and unbounded-error complexity $\Theta(1)$, as one can verify from the representation $\mathrm{AND}_n(|x \wedge y|) = \prod x_i \cdot \prod y_i$. Similarly, the familiar predicate $\mathrm{PARITY}_n(t) = t \bmod 2$ has degree n and unbounded-error complexity $\Theta(n)$ by Forster's result [7]. At the same time, there are predicates D for which a logarithmic gap exists between $\deg(D)$ and $U(D)$. One such predicate is disjointness, given by $\mathrm{DISJ}_n(t) = 1 \Leftrightarrow t = 0$, which has degree 1 and unbounded-error complexity $\Theta(\log n)$:

Proposition 6.5. $U(\mathrm{DISJ}_n) = \Theta(\log n)$.

Proof. The upper bound is immediate from Theorem 6.3. For the lower bound, note that

$$\bigoplus_{i=1}^{n} x_i \wedge y_i = \bigvee_{i=1}^{4^n} f_i(x_1, \ldots, x_n) \wedge g_i(y_1, \ldots, y_n),$$

where f_i, g_i are suitable Boolean functions (in fact, conjunctions of literals). This yields the inequality $U(\mathrm{PARITY}_n) \leqslant U(\mathrm{DISJ}_{4^n})$, which completes the proof since $U(\mathrm{PARITY}_n) = \Theta(n)$ by Forster's result [7]. □

The lower bound of Proposition 6.5 is of course valid for any predicate D that contains disjointness or its negation as a subfunction. More precisely:

Proposition 6.6. *Let* $D\colon \{0,1,\ldots,n\} \to \{0,1\}$ *be a predicate with flip vector* v. *If* v *contains the subvector* $(1, \underbrace{0, 0, \ldots, 0}_{m})$, *then* $U(D) \geqslant \Omega(\log m)$.

To illustrate, Proposition 6.6 shows that the majority predicate $\textsc{maj}_n(t) = 1 \Leftrightarrow t > n/2$ has degree 1 and unbounded-error complexity $\Theta(\log n)$. Other threshold predicates can be handled analogously.

Acknowledgments

I would like to thank Anna Gál, Adam Klivans, and the anonymous reviewers for their useful comments on an earlier version of this manuscript. This research was supported by Adam Klivans' NSF CAREER Award and NSF Grant CCF-0728536.

References

[1] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *Proc. of the 26th Symposium on Foundations of Computer Science (FOCS)*, pages 277–280, 1985.

[2] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

[3] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proc. of the 27th Symposium on Foundations of Computer Science (FOCS)*, pages 337–347, 1986.

[4] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[5] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proc. of the 22nd Conf. on Computational Complexity (CCC)*, pages 24–32, 2007.

[6] V. Feldman, P. Gopalan, S. Khot, and A. K. Ponnuswami. New results for learning noisy parities and halfspaces. In *Proceedings of the 47th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 563–574, 2006.

[7] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[8] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H.-U. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Proc. of the 21st Conf. on Foundations of Software Technology and Theoretical Computer Science (FST TCS)*, pages 171–182, 2001.

[9] J. Forster and H. U. Simon. On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes. *Theor. Comput. Sci.*, 350(1):40–48, 2006.

[10] M. Goldmann, J. Håstad, and A. A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[11] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[12] S. Jukna. *Extremal Combinatorics with Applications in Computer Science*. Springer-Verlag, Berlin, 2001.

[13] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5(4):545–557, 1992.

[14] B. S. Kashin and A. A. Razborov. Improved lower bounds on the rigidity of Hadamard matrices. *Matematicheskie zametki*, 63(4):535–540, 1998. In Russian.

[15] M. Kearns and L. Valiant. Cryptographic limitations on learning Boolean formulae and finite automata. *J. ACM*, 41(1):67–95, 1994.

[16] M. Kharitonov. Cryptographic hardness of distribution-specific learning. In *Proc. of the 25th Symposium on Theory of Computing*, pages 372–381, 1993.

[17] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007.

[18] A. R. Klivans and R. A. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. *J. Comput. Syst. Sci.*, 68(2):303–318, 2004.

[19] A. R. Klivans and A. A. Sherstov. A lower bound for agnostically learning disjunctions. In *Proc. of the 20th Conf. on Learning Theory (COLT)*, pages 409–423, 2007.

[20] A. R. Klivans and A. A. Sherstov. Unconditional lower bounds for learning intersections of halfspaces. *Machine Learning*, 69(2–3):97–114, 2007.

[21] A. R. Klivans and A. A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. *J. Comput. Syst. Sci.*, 75(1):2–12, 2009.

[22] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, New York, 1997.

[23] S. V. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. *J. Comput. Syst. Sci.*, 63(3):449–473, 2001.

[24] M. L. Minsky and S. A. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, Mass., 1969.

[25] I. Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, 1991.

[26] N. Nisan. The communication complexity of threshold gates. In *Combinatorics, Paul Erdős is Eighty*, pages 301–315, 1993.

[27] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions. In *Proc. of the 24th Symposium on Theory of Computing (STOC)*, pages 468–474, 1992.

[28] R. Paturi and J. Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[29] R. Raz. Fourier analysis for probabilistic communication complexity. *Comput. Complex.*, 5(3/4):205–221, 1995.

[30] A. A. Razborov. Bounded-depth formulae over the basis $\{\&, \oplus\}$ and some combinatorial problems. *Complexity Theory and Applied Mathematical Logic*, vol. "Problems of Cybernetics":146–166, 1988. In Russian.

[31] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[32] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[33] A. A. Razborov and A. A. Sherstov. The sign-rank of AC^0. In *Proc. of the 49th Symposium on Foundations of Computer Science (FOCS)*, pages 57–66, 2008.

[34] T. J. Rivlin. *An Introduction to the Approximation of Functions*. Dover Publications, New York, 1981.

[35] A. A. Sherstov. Powering requires threshold depth 3. *Inf. Process. Lett.*, 102(2–3):104–107, 2007.

[36] A. A. Sherstov. Halfspace matrices. *Comput. Complex.*, 17(2):149–178, 2008. Preliminary version in 22nd CCC, 2007.

[37] A. A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 2010. To appear. Preliminary version in 40th STOC, 2008.

[38] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009. Preliminary version in 39th STOC, 2007.

[39] Y. Shi and Y. Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5–6):444–460, 2009.

[40] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

[41] R. de Wolf. *Quantum Computing and Communication Complexity*. PhD thesis, University of Amsterdam, 2001.