On Computation and Communication with Small Bias

Harry Buhrman*
CWI Amsterdam, and
University of Amsterdam
buhrman@cwi.nl

Nikolay Vereshchagin†
Moscow State University
ver@mccme.ru

Ronald de Wolf‡
CWI Amsterdam
rdewolf@cwi.nl

Abstract

We present two results for computational models that allow error probabilities close to 1/2.

First, most computational complexity classes have an analogous class in communication complexity. The class PP in fact has two*, a version with weakly restricted bias called PP^{cc}, and a version with unrestricted bias called UPP^{cc}. Ever since their introduction by Babai, Frankl, and Simon in 1986, it has been open whether these classes are the same. We show that PP$^{cc} \subsetneq$ UPP^{cc}. Our proof combines a query complexity separation due to Beigel with a technique of Razborov that translates the acceptance probability of* quantum *protocols to polynomials.*

Second, we study how small the bias of minimal-degree polynomials that sign-represent Boolean functions needs to be. We show that the worst-case bias is at worst double-exponentially small in the sign-degree (which was very recently shown to be optimal by Podolski), while the average-case bias can be made single-exponentially small in the sign-degree (which we show to be close to optimal).

1 Introduction

Many models in theoretical computer science allow for computations or representations where the answer is only slightly biased in the right direction. The best-known of these is the complexity class PP, for "probabilistic polynomial time". A language is in PP if there is a randomized polynomial-time Turing machine whose acceptance probability is greater than 1/2 if, and only if, its input is in the language. The *bias* of such a computation is how far from the crossover value of 1/2 the actual probability is. This class is quite powerful. For instance, it can compute NP-complete problems, albeit with exponentially small bias. Many analogues of this class exist, for instance for decision trees, communication protocols, polynomial representations, etc. Though not corresponding to "effective" computation (for that we need small error probability), this is still a fundamental mode of computation, giving rise to many interesting questions. Clearly the larger the bias the better, for instance because it is much cheaper to amplify the success probability of an algorithm with large bias than one with small bias. Hence it makes sense to ask how large we can make this bias. In this paper we study this issue in two contexts: communication protocols and sign-representing polynomials over the reals.

1.1 Communication complexity

Communication complexity has been one of the most fruitful areas of theoretical computer science since its introduction by Yao [33]. The model has appeal in its own right as a simple model of distributed computing, and also has found numerous applications, in particular for proving lower bounds on circuits, data structures, etc. [19]. Already 20 years ago, Babai, Frankl, and Simon [6] defined the communication complexity analogues of standard computational complexity classes such as P, BPP, NP, PH, PSPACE, etc. Here "polylog communication" replaces "polynomial-time" as the formalization of "efficient" computation of some function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$.[1] The communication complexity classes are distinguished from their computational cousins by a superscript 'cc'. This framework enables a notion of efficiency-preserving "rectangular

*Partially supported by a Vici grant from the Netherlands Organization for Scientific Research (NWO), by BRICKS Project AFM1, and by the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract Number 015848.

†Work partly done while visiting CWI in Summer 2006. Partially supported by the grant 05-01-02803 from the Russian Federation Basic Research Fund.

‡Partially supported by a Veni grant from the Netherlands Organization for Scientific Research (NWO), and by the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract Number 015848.

[1] For upper and lower bounds depending on the input length n to make sense, we should really be talking about *families* of functions $\{f_n\}$, one for each n, instead of functions f for specific n. We will ignore this technicality here.

reduction" between communication problems, analogous to efficient many-one reductions in computational complexity.

Some relations between complexity classes that are notoriously hard to settle in the computational setting, can be solved in the communication case. For instance, $\mathrm{P}^{cc} \neq \mathrm{NP}^{cc}$, $\mathrm{NP}^{cc} \neq \mathrm{coNP}^{cc}$, $\mathrm{NP}^{cc} \not\subseteq \mathrm{BPP}^{cc}$ (example for these three cases: set intersection [6]), and $\mathrm{P}^{cc} \neq \mathrm{BPP}^{cc}$ and $\mathrm{BPP}^{cc} \not\subseteq \mathrm{NP}^{cc}$ (example: equality [33]). On the other hand, there are also some collapses that we do not expect to hold true in the computational setting, in particular $\mathrm{P}^{cc} = \mathrm{NP}^{cc} \cap \mathrm{coNP}^{cc}$ [2]. Other properties of communication complexity classes may be found in [6, 13, 14, 16, 23, 9, 21, 31].

In some cases the communication framework is richer than the computational framework. For example, Babai et al. introduced two different communication complexity versions of the complexity class PP. The first communication version, called UPP^{cc} for "unrestricted-error probabilistic protocols", just considers all functions computable by protocols with polylogarithmic communication and acceptance probabilities that are above 1/2 if $f(x, y) = 1$, and below 1/2 if $f(x, y) = 0$. Such protocols were first studied in [26]. The second version realizes that efficiency should also involve the number of random bits used. Here we mean *private* coins, not *public* coins. Note that if the number of coin flips is upper bounded by c, then any bias will be lower bounded by 2^{-c}, just because the probability of any event will be a multiple of 2^{-c}. Accordingly, the second kind of communication complexity is defined as the sum of the communication and the log of the reciprocal of the worst-case bias. PP^{cc} is the class of communication problems for which this PP-complexity is polylogarithmic. Note that we allow bias as small as $2^{-polylog(n)}$ here.

Obviously $\mathrm{PP}^{cc} \subseteq \mathrm{UPP}^{cc}$. Ever since the introduction of these two classes by Babai et al., it has been an open question whether this inclusion is strict. In this paper we answer this question in the affirmative. We exhibit a total Boolean function, inspired by a function used earlier by Beigel [7] in the setting of oracle-computations, which can be solved by UPP-protocols with $O(\log n)$ communication, but whose PP-communication complexity is $n^{\Omega(1)}$. In other words, this function can be efficiently computed with some small positive bias, but not with relatively large bias.[2]

Interestingly, our lower bound relies on a result of Razborov [28] which roughly says that the acceptance probability of *quantum* communication protocols can be well-approximated by a polynomial of degree roughly equal to the communication complexity. It should be noted that this connection with quantum is not essential: the special case of Razborov's result that applies to classical protocols would already suffice for our purposes. However, the classical version of Razborov's lemma was not known prior to [28], and arguably would not have been discovered if it weren't for the more general quantum version.

Our separation between UPP^{cc} and PP^{cc} also separates two well-known lower bound techniques in randomized communication complexity. As mentioned in the next section, the UPP-communication complexity of a function f is determined by the minimal rank among all matrices that *sign-represent* f, while the PP-complexity is determined by the *discrepancy* of f under the hardest input distribution. It follows that the second technique can be exponentially stronger than the first. By the recent work of Linial and Shraibman [21, 22] (following up on [20]), discrepancy is equivalent to *margin complexity*, which is an important notion from learning theory (we will not spell out the consequences of our bounds for learning theory here). Hence our result also exponentially separates sign-rank from margin complexity.

Sherstov's results. As we learned recently, an exponential separation between sign-rank and margin complexity has also been obtained independently by Sherstov [31] (in these proceedings), for a different function and with quite different techniques.

In another development, Sherstov [32] recently exhibited a function with exponentially small discrepancy that has depth-3 circuits of polynomially many AND, OR, and NOT-gates. He shows that exponentially small discrepancy implies that depth-2 circuits with majority-gates for the function need exponential size. In other words, he separates AC^0 from $MAJ \circ MAJ$ circuits. This contrasts with a classic result by Allender [3], who showed that all languages in AC^0 have quasipolynomial-sized majority-circuits of *depth 3*. As Sherstov noticed, the function we analyze in Section 3 has the same property: the discrepancy bound follows from our communication lower bound, while the depth-3 circuit is easy to construct.

1.2 Polynomials and decision trees

For the setting of polynomials it will be convenient to switch from $0/1$-variables to ± 1-variables. An n-variate polynomial p (over the reals) *sign-represents* a function $f : \{\pm 1\}^n \to \{\pm 1\}$ if it has the same sign for all inputs x: $p(x) > 0$ if $f(x) = 1$ and $p(x) < 0$ if $f(x) = -1$. Such polynomials are also known as "threshold functions". Since $x_i^2 = 1$ for $x_i \in \{\pm 1\}$, we can without loss of generality restrict attention to *multilinear* polynomials. Probably the most important complexity measure for such a poly-

[2] As an aside, the same function can be used to separate the communication complexity class $\mathrm{P}^{\mathrm{NP},cc}$ from PP^{cc} (similar to [7]), and also $\mathrm{P}^{\mathrm{NP},cc}$ from $\mathrm{P}^{\mathrm{NP}^{\|},cc}$. It is not hard to see that our function sits in $\mathrm{P}^{\mathrm{NP},cc}$. On the other hand, using techniques from [8, 12, 1] one can show that $\mathrm{P}^{\mathrm{NP}^{\|},cc} \subseteq \mathrm{PP}^{cc}$. As we show here, the latter class does not contain our function. We omit the rather technical definitions and proofs. One can also define the communication analogue of Aaronson's class PostBQP [1], and show $\mathrm{PP}^{cc} \subsetneq \mathrm{PostBQP}^{cc} \subseteq \mathrm{UPP}^{cc}$.

nomial is its *degree*, which is the size of its largest monomial. Define the *sign-degree* of f as the minimal degree $sdeg(f)$ among all polynomials p that sign-represent f.[3] Functions with low sign-degree have found various applications in complexity theory, for instance in the proof by Beigel et al. [8] that PP is closed under intersection, and in a number of oracle results [7, 5]. They are also closely related to threshold circuits and neural networks.

Once the degree of p has been fixed to $sdeg(f)$, one may ask *how well* p approximates f. We formalize this as follows. Suppose p sign-represents f and p is *normalized* in the sense that $|p(x)| \leq 1$ for all $x \in \{\pm 1\}^n$. Then define the (worst-case) *bias* of p as $\min_x |p(x)|$. This measures how far away from the crossover point 0 the polynomial is. Note that the normalization condition is needed to avoid increasing the bias by just multiplying the polynomial by a large number. Now we ask: what is the best-achievable (i.e. maximal) bias among such polynomials?[4]

Another question is to ask how large the weights (coefficients) need to be in *integer-coefficient* sign-representing polynomials for f. Clearly, these two questions are closely related: if we need large integer weights then the maximal bias will be small, and vice versa. We state this relation between bias and weights more precisely in Section 2.2; for the purposes of this introduction we will treat these two problems as basically equivalent.

It has been known for a long time that for *linear* threshold functions (those of sign-degree at most 1), weights of size $2^{O(n \log n)}$ suffice [24]. Håstad [15] exhibited a function where weights of that size are also necessary. Equivalently, the best bias among normalized degree-1 polynomials for Håstad's function is $2^{-\Theta(n \log n)}$.[5]

Very little seems to be known about the best bias obtainable for functions having $sdeg(f) > 1$. We present two results about this. First, we show that the best-achievable bias is at least double-exponentially small: every total function f has a sign-representing polynomial of degree $sdeg(f)$ with worst-case bias at least $1/N \cdot N!$, where $N = \sum_{i=0}^{sdeg(f)} \binom{n}{i}$. This lower bound on the bias is roughly $n^{-n^{sdeg(f)}}$. That does not look very impressive, but Håstad's example shows that this is actually essentially tight for $sdeg(f) = 1$. After a first version of this paper appeared, Podolski [27] showed our bound is in fact essentially tight for all values of $sdeg(f)$: for each d he exhibits a family of n-bit Boolean functions f with $sdeg(f) = d$, such that any degree-d normalized polynomial that sign-represents f has worst-case bias at most $n^{-\Omega(n^d)}$ (the constant in the Ω depends on d).

Second, we also study the *average* bias obtainable, where the average is taken under the uniform distribution on all inputs. We show that every total function f has a sign-representing polynomial of degree $sdeg(f)$ with average-case bias at least $1/\sum_{i=0}^{sdeg(f)} \binom{n}{i} \approx 1/n^{sdeg(f)}$. Hence there is an exponential gap between worst-case and average-case bias. In addition, we exhibit a family of functions where our lower bound on the achievable average-case bias is close to optimal.

Finally, to further motivate the study of sign-representing polynomials and bias, let us mention the close relation between sign-representing polynomials for f and randomized decision trees. On the one hand, the acceptance probability of a depth-d randomized decision tree can be written as a polynomial p of degree at most d. If the decision tree computes some function f with success probability at least $1/2 + \beta$ on all inputs, then the polynomial $p - 1/2$ will sign-represent f with bias β. On the other hand, if we have a degree-d polynomial that sign-represents f, we can obtain from this a randomized decision tree of depth at most d that computes f with bias roughly $\beta/\sqrt{n^d}$ (see Section 4.2.1). Accordingly, up to relatively moderate changes in the bias, degree of sign-representing polynomial is equivalent to depth of decision trees.

2 Preliminaries

2.1 Communication complexity

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Alice gets input x, Bob gets input y, and together they want to compute $f(x,y)$ with minimal communication between them. We assume familiarity with deterministic and probabilistic two-party communication protocols [19].

A protocol P computes f with *bias* $\beta \geq 0$ if its acceptance probability is at least $1/2 + \beta$ for every input $(x,y) \in f^{-1}(1)$ and at most $1/2 - \beta$ for $(x,y) \in f^{-1}(0)$. We use $\beta(P)$ for P's bias. The *cost* $C(P)$ of a protocol P is its worst-case communication. Let $\mathrm{UPP}(f)$ denote the minimal cost $C(P)$ among all protocols P that compute f with positive bias. Let $\mathrm{PP}(f)$ denote the minimum of $C(P) + \log(1/\beta(P))$ among all protocols P that compute f with positive bias. Note that the bias is lower bounded by $2^{-\mathrm{PP}(f)} \geq 2^{-n-1}$ for such protocols. In contrast, for UPP-protocols the bias is unrestricted (whence the 'U').

Obviously $\mathrm{UPP}(f) \leq \mathrm{PP}(f)$ for all f. We list some of the main results that are known about these complexity measures:

[3] Note that we do not allow $p(x) = 0$ for any x. The literature, for instance [5, 25], also contains a notion of "weakly sign-represents", which requires that p's sign equals $f(x)$ whenever $p(x) \neq 0$, and that $p(x) \neq 0$ for at least one input x. We will not consider this alternative definition here.

[4] The restriction to polynomials of degree $sdeg(f)$ is natural but also somewhat limiting: it could be that polynomials of degree slightly larger than $sdeg(f)$ can achieve much better bias.

[5] If one only wants the sign of the degree-1 polynomial p to equal f for *most* instead of all inputs, then the situation changes dramatically: weights of size roughly $\sqrt{n}$ already suffice [30]. We will not study such "low-weight approximators" here.

- Almost all f have UPP$(f) \geq n - O(1)$ [4].
- The inner product function $f(x,y) = \sum_{i=1}^{n} x_i y_i \bmod 2$ has UPP$(f) \geq n/2$ [11].
- Let $srank(f)$ be the *sign-rank* of f (minimal rank among all $2^n \times 2^n$ matrices M having $M_{xy} > 0$ if $f(x,y) = 1$, and $M_{xy} < 0$ if $f(x,y) = 0$). Then UPP(f) equals $\log srank(f)$ up to a bit [4].
- PP-complexity is essentially determined by *discrepancy*. Let $\mu : \{0,1\}^n \times \{0,1\}^n \to [0,1]$ be an input distribution. Then the *discrepancy* of f w.r.t. μ is

  $$disc_\mu(f) = \max_R |\mu(R \cap f^{-1}(1)) - \mu(R \cap f^{-1}(0))|,$$

  where the maximum is taken over all rectangles $R = S \times T \subseteq \{0,1\}^n \times \{0,1\}^n$. We have PP$(f) = \Theta(\log(1/\min_\mu disc_\mu(f)) + \log n)$ [17].
- Two-way UPP-protocols are not more powerful than one-way UPP-protocols [26], and the same holds for PP-protocols [17].

2.2 Sign-representing polynomials

Our polynomials will always be over the real numbers. When talking about sign-representing polynomials, it is convenient to switch from $0/1$-variables to ± 1-variables.

Let $[n] = \{1, \ldots, n\}$. An n-variate multilinear polynomial (often just called a polynomial) is a function

$$p(x) = \sum_S \hat{p}(S) x_S,$$

where $x = (x_1, \ldots, x_n) \in \{\pm 1\}^n$, the sum goes over all sets $S \subseteq [n]$ of indices of variables, the $\hat{p}(S)$ are reals (known as the Fourier coefficients of p), and the *monomial* x_S is a function of x given by $x_S = \prod_{i \in S} x_i$ (i.e. the parity of the variables in S). If $S = \emptyset$, then x_S is the constant-1 function. The *degree* of p is $deg(p) = \max\{|S| \mid \hat{p}(S) \neq 0\}$.

We define an inner product between functions $f, g : \{\pm 1\}^n \to \mathbb{R}$ by

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

It easy to see that the set of all monomials x_S forms an orthonormal set with respect to this inner product, and the Fourier coefficients of p can be expressed as $\hat{p}(S) = \langle p, x_S \rangle$. Parseval's identity says

$$\frac{1}{2^n} \sum_x p(x)^2 = \sum_S \hat{p}(S)^2.$$

We say that p *sign-represents* a function $f : \{\pm 1\}^n \to \{\pm 1\}$ if it has the same signs: $p(x) > 0$ whenever $f(x) = 1$ and $p(x) < 0$ whenever $f(x) = -1$. The *sign-degree* of f is $sdeg(f) = \min\{deg(p) \mid p \text{ sign-represents } f\}$. O'Donnell and Servedio [25] have shown that almost all f have $sdeg(f) \approx n/2$.

In order to be able to define the bias of p, we assume $|p(x)| \leq 1$ for all inputs x. We call such p *normalized*. The *worst-case* bias of p is

$$\beta = \min_x |p(x)|$$

and the *average-case* bias is

$$\overline{\beta} = \frac{1}{2^n} \sum_x |p(x)|.$$

Much of the literature on sign-representations considers sign-representing polynomials q with integer coefficients (a.k.a. *weights*) and focuses on the magnitude of the largest weight, while our work considers sign-representing polynomials p satisfying $\max_x |p(x)| \leq 1$ and focuses on the bias of p away from 0. Here we will relate these two approaches to each other: roughly, small bias for p corresponds to large weight for q.

Let $N = \sum_{i=0}^{d} \binom{n}{i}$. First, suppose we have a degree-d polynomial q with integer coefficients. Let $q_{max} = \max_S |\hat{q}(S)|$ be its largest weight. Note that $\max_x |q(x)| \leq \sum_S |\hat{q}(S)| \leq N q_{max}$. Define $p = q/\max_x |q(x)|$, then clearly $|p(x)| \leq 1$ for all x. We have the following lower bound on the worst-case bias β of p:

$$\beta = \min_x |p(x)| = \frac{\min_x |q(x)|}{\max_x |q(x)|} \geq \frac{1}{N q_{max}}.$$

Conversely, suppose we have a degree-d polynomial p satisfying $\beta \leq |p(x)| \leq 1$ for all x. Now define $\tilde{q} = p \cdot N/\beta$ and define q by rounding positive coefficients of $\tilde{q}$ down and rounding negative coefficients up to obtain integer coefficients. We have $|\tilde{q}(x)| \geq N$ and $|q(x) - \tilde{q}(x)| < N$ for every x. Accordingly, the polynomials p, $\tilde{q}$, and q all have the same sign for every x. Moreover, the magnitude of the largest coefficient of q is

$$q_{max} \leq \tilde{q}_{max} \leq \max_x |p(x)| N/\beta \leq N/\beta.$$

Summarizing:

Corollary 1. *Let $N = \sum_{i=0}^{d} \binom{n}{i}$. For every integer-coefficient polynomial q of degree d with maximal weight q_{max}, there is a normalized polynomial p of degree at most d with bias $\beta \geq 1/(N q_{max})$ that sign-represents the same function. For every normalized polynomial p of degree d with bias β, there is an integer-coefficient polynomial q of degree at most d with maximal weight $q_{max} \leq N/\beta$ that sign-represents the same function.*

3 Separating PP^{cc} and UPP^{cc}: The communication version of ODD-MAX-BIT

In this section we state our main result about communication complexity: a function that is in UPP^{cc} but not in PP^{cc}. We use a distributed version of the ODD-MAX-BIT function of Beigel [7]. Let $x, y \in \{0,1\}^n$, and $k = \max\{i \in [n] \mid x_i = y_i = 1\}$ be the rightmost position where x and y both have a 1 (set $k = 0$ if there is no such position). Define $f(x, y)$ to be the least significant bit of k, i.e. whether this k is odd or even. We will show here that $\text{UPP}(f) = O(\log n)$ while $\text{PP}(f) = \Omega(n^{1/3})$.

3.1 UPP-upper bound

For $i \in [n] = \{1, \ldots, n\}$, define probabilities $p_i = c2^i$, where $c = 1/\sum_{i=1}^{n} 2^i$ is a normalizing constant. Consider the following protocol. Alice picks a number $i \in [n]$ with probability p_i and sends over i, x_i. If $x_i = y_i = 1$ then Bob outputs the least significant bit of i, otherwise he outputs a fair coin flip. This computes f with positive—though exponentially small—bias. Hence

$$\text{UPP}(f) \leq \lceil \log n \rceil + 1.$$

3.2 Quantum lower bound

We will actually prove the lower bound for *quantum* protocols (without prior entanglement). Let

$$\text{QPP}(f) = \min_P \left(C(P) + \log(1/\beta(P)) \right)$$

be the PP-type *quantum* communication complexity of f, which is the minimum over all *quantum* protocols P that compute f with positive bias. It is known that $\text{QPP}(f) = \Theta(\text{PP}(f))$ [17], hence lower bounding $\text{PP}(f)$ is equivalent to lower bounding $\text{QPP}(f)$. It won't be necessary to precisely define quantum protocols here, since the only property we use is the following result by Razborov. This was first proved in [28], and made more explicit in [18, Section 5]. It allows us to translate a quantum protocol to a polynomial:

Lemma 1 (Razborov)**.** *Consider a q-qubit quantum communication protocol on m-bit inputs x and y, with outputs 0 and 1, and acceptance probabilities denoted by $P(x, y)$. For $i \in \{0, \ldots, m/4\}$, define*

$$P(i) = \text{Exp}_{|x|=|y|=m/4, |x \wedge y|=i}[P(x, y)],$$

where the expectation is taken uniformly over all $x, y \in \{0,1\}^m$ that each have weight $m/4$ and that have intersection size i. For every $d \leq m/4$ there exists a single-variate degree-d polynomial p (over the reals) such that $|P(i) - p(i)| \leq 2^{-d/4+2q}$ for all $i \in \{0, \ldots, m/8\}$.

Note that if we pick $d = 8q + 4\log(1/\varepsilon)$, then p approximates P to within an additive ε for all $i \in \{0, \ldots, m/8\}$.

We also use the following special case of a result due to Ehlich and Zeller [10] and Rivlin and Cheney [29]:

Lemma 2 (Ehlich & Zeller; Rivlin & Cheney)**.** *Let r be a single-variate degree-d polynomial such that $r(0) \leq -1$ and $r(i) \in [0, 2]$ for all $i \in [k]$. Then $d \geq \sqrt{k/4}$.*

Consider a quantum protocol with q qubits of communication that computes f with bias $\beta > 0$. Let $\beta(x, y) = P(x, y) - 1/2$. Then $\beta(x, y) \geq \beta$ if $f(x, y) = 1$, and $\beta(x, y) \leq -\beta$ if $f(x, y) = 0$. Our goal is to lower bound $q + \log(1/\beta)$.

Define $d = \lceil 8q + 4\log(2/\beta) \rceil$ and $m = 32d^2 + 1$. Assume for simplicity that $2m$ divides n. We will partition $[n]$ into $n/2m$ consecutive intervals, each of length $2m$. In the first interval (from the left), fix x_i and y_i to 0 for even i; in the second, fix x_i and y_i to 0 for odd i; in the third, fix x_i and y_i to 0 for even i, etc. In the jth interval there are m unfixed positions left. Let $x^{(j)}$ and $y^{(j)}$ denote the corresponding m-bit strings in x and y, respectively.

We will define successively, for all $j = 1, 2, \ldots, n/2m$, particular strings $x^{(j)}$ and $y^{(j)}$ so that the following holds. Let X^j and Y^j denote n-bit strings where the first j blocks are set to $x^{(1)}, \ldots, x^{(j)}$ and $y^{(1)}, \ldots, y^{(j)}$, respectively, and all the other blocks are set to 0. In particular, X^0 and Y^0 are all zeros. We will define $x^{(j)}$ and $y^{(j)}$ so that

$$\beta(X^j, Y^j) \geq 2^j\beta \qquad \text{or} \qquad \beta(X^j, Y^j) \leq -2^j\beta$$

depending on whether j is odd or even. Note that this holds automatically for $j = 0$.

Assume that $x^{(1)}, \ldots, x^{(j-1)}$ and $y^{(1)}, \ldots, y^{(j-1)}$ are defined on previous steps. On the current step, we have to define $x^{(j)}$ and $y^{(j)}$. Without loss of generality assume that j is odd, thus we have $\beta(X^{j-1}, Y^{j-1}) \leq -2^{j-1}\beta$. Consider some $i = 0, 1, \ldots, m/4$. Run the protocol on the following distribution: $x^{(j)}$ and $y^{(j)}$ are chosen randomly subject to each having weight $m/4$, and having intersection size i, the blocks with indexes smaller than j are fixed (on previous steps), the blocks with indexes larger than j are set to zero. Let $P(i)$ denote the expected value of $\beta(x, y)$ as a function of i. Note that for $i = 0$ we have $P(i) = \beta(X^{j-1}, Y^{j-1}) \leq -2^{j-1}\beta$. On the other hand, for each $i > 0$ the expectation is taken over x, y with $f(x, y) = 1$, because the rightmost intersecting point is in the jth interval and hence odd (the even indices in the jth interval have all been fixed to 0). Thus $P(i) \geq \beta$ for those i. Now assume, by way of contradiction, that $\beta(X^j, Y^j) \leq 2^j\beta$ for all $x^{(j)}, y^{(j)}$ and hence $P(i) \leq 2^j\beta$ for all such i. By Lemma 1, for our choice of d, we can approximate $P(i)$ to within additive difference of $\beta/2$ by a polynomial p of degree d. (We do this by applying Razborov's lemma to the protocol obtained from the original protocol by fixing all

bits outside the jth block.) Let r be the degree-d polynomial

$$\frac{p - \beta/2}{2^{j-1}\beta}.$$

From the properties of P and the fact that p approximates P up to $\beta/2$, we see that $r(0) \leq -1$ and $r(i) \in [0,2]$ for all $i \in [m/8]$. But then by Lemma 2, the degree of r is at least $\sqrt{(m/8)/4} = \sqrt{d^2 + 1/32} > d$, which is a contradiction. Hence there exists an intersection size $i \in [m/8]$ where $P(i) \geq 2^j\beta$. Thus there are particular $x^{(j)}, y^{(j)}$ with $\beta(X^j, Y^j) \geq 2^j\beta$.

For $j = n/2m$ we obtain $|\beta(X^j, Y^j)| \geq 2^{n/2m}\beta$. But for every x, y we have $|\beta(x,y)| \leq 1/2$, hence

$$1/2 \geq 2^{n/2m}\beta.$$

This implies

$$2m \log(1/\beta) \geq n,$$

hence

$$\begin{aligned}(q + \log(1/\beta))^3 &\geq (q + \log(1/\beta))^2 \log(1/\beta) \\ &= \Omega(m \log(1/\beta)) \\ &= \Omega(n).\end{aligned}$$

Since this holds for every quantum protocol computing f with q qubits of communication and bias $\beta > 0$, we have

$$\mathrm{QPP}(f) = \Omega(n^{1/3}).$$

4 The bias of sign-representing polynomials

In this section we study the bias of polynomials that sign-represent Boolean functions.

4.1 Lower bound on the worst-case bias

First we give a lower bound on the *worst-case* bias.

Theorem 1. *Let $N = \sum_{i=0}^{d} \binom{n}{i}$. If there is a degree-$d$ polynomial that sign-represents $f : \{\pm 1\}^n \to \{\pm 1\}$, then there is a normalized degree-d polynomial that sign-represents f with worst-case bias $\beta \geq \frac{1}{N \cdot N!}$.*

Proof. Let $m_1, \ldots, m_N$ be all the monomials of degree at most d in the n variables $x_1, \ldots, x_n$. Any degree-d polynomial $p(x_1, \ldots, x_n)$ is a linear combination $p = \Sigma_{j=1}^{N} p_j m_j$ of those monomials. Let a be an assignment of ± 1-values to the variables $x_1, \ldots, x_n$ and let $m_i(a) \in \{\pm 1\}$ stand for the value of monomial m_i on a. We are given that the following system of 2^n linear inequalities (in N variables p_j) is consistent:

$$\{ \quad f(a) \sum_{j=1}^{N} m_j(a) p_j > 0 \quad | \quad a \in \{\pm 1\}^n \}. \qquad (1)$$

We can multiply any solution of (1) by a large number, so the following system is also consistent:

$$\{ \quad f(a) \sum_{j=1}^{N} m_j(a) p_j \geq 1 \quad | \quad a \in \{\pm 1\}^n \}. \qquad (2)$$

We claim that system (2) has a solution where $f(a) \sum_{j=1}^{N} m_j(a) p_j \leq N \cdot N!$ for all a. To show this, pick a solution $\tilde{p}_1, \ldots, \tilde{p}_N$ to (2) and for each $j = 1, \ldots, N$ add to the system (2) the inequality $p_j \geq 0$ if $\tilde{p}_j \geq 0$, and the inequality $p_j \leq 0$ otherwise. Let

$$\{ \quad \sum_{j=1}^{N} b_{ij} p_j \geq c_i \quad | \quad i = 1, \ldots, N + 2^n \quad \} \qquad (3)$$

be the resulting system.

We need to introduce some terminology about linear programming. The set of all solutions to a system of linear inequalities is called a *polyhedron*. A point A of a polyhedron is called its *vertex* if there is no line segment that is entirely included in the polyhedron and that has A as inner point. Let a polyhedron P be defined by a system of linear inequalities $\sum_{j=1}^{N} u_{ij} p_j \geq v_i$. Let $\tilde{p}$ be a point in P. Consider all the inequalities from the system that hold with equality for $p = \tilde{p}$. Let $S_{\tilde{p}}$ stand for the system consisting of such equalities $\sum_{j=1}^{N} u_{ij} p_j = v_i$. Then one can prove the following: $\tilde{p}$ is a vertex of P iff the rank of $S_{\tilde{p}}$ (that is, the rank of its matrix) is equal to N.

An (affine) line is a subset of $\mathbb{R}^N$ of the form $r + L$ where $r \in \mathbb{R}^N$ and L is a one-dimensional linear subspace of $\mathbb{R}^N$. System (3) has the following property: no affine line is entirely included in the polyhedron P of solutions to (3) (every line crosses a hyperplane $p_j = 0$ for some j). This implies that P has a vertex. Indeed, start at any point $\tilde{p}$ in P. If the rank of $S_{\tilde{p}}$ is equal to N, we are done. Otherwise, the set of solutions to $S_{\tilde{p}}$ contains an affine line passing through $\tilde{p}$. As this line is not entirely included in P, there is a point $\hat{p}$ on the line where the line first gets out of P. In other words, there is an inequality $\sum_{j=1}^{N} u_{ij} p_j \geq v_i$ that is an equality for $p = \hat{p}$ and that is false for points of the line lying further from $\tilde{p}$ than $\hat{p}$. This equality cannot be a linear combination of those in $S_{\tilde{p}}$ (that would mean that all the points on the line satisfy that equality). Thus replacing $\tilde{p}$ by $\hat{p}$ we can increase the rank of $S_{\tilde{p}}$ and repeat the argument.

Now pick any solution $\tilde{p}_1, \ldots, \tilde{p}_N$ to (3) such that the rank of the system $S_{\tilde{p}}$ is N. Write this system in matrix form: $Mp = c$. Without loss of generality we may assume that the size of matrix M is $N \times N$. By Cramer's rule, every $\tilde{p}_k$ has the form A_k/B, where B is the determinant of M and A_k is the determinant of the matrix obtained from M by replacing its kth column by column vector c. Note that $m_j(a) \in \{\pm 1\}$ for all j, a, therefore all b_{ij}, c_i are equal to 0, 1 or -1. Hence $|B| \geq 1$ and $|A_k| \leq N!$.

Thus we obtain the bound $|\tilde{p}_k| \leq N!$ and

$$1 \leq f(a) \sum_{j=1}^{N} m_j(a) \tilde{p}_j \leq N \cdot N!,$$

for all $a \in \{\pm 1\}^n$, so the normalized degree-d polynomial

$$\sum_{j=1}^{N} \tilde{p}_j m_j / (N \cdot N!)$$

sign-represents f with bias at least $1/(N \cdot N!)$. □

As mentioned in the introduction, Håstad [15] showed that this bound is essentially tight for $d = 1$, and Podolski [27] recently showed this for all d.

4.2 Bounds on the average-case bias

In this section we analyze the average-case bias.

4.2.1 Lower bound

We first show that a sign-representing polynomial can be converted into a probability distribution on parities (and their negations).

Lemma 3. *Let $N = \sum_{i=0}^{d} \binom{n}{i}$. Suppose degree-$d$ normalized polynomial p sign-represents $f : \{\pm 1\}^n \to \{\pm 1\}$ with bias β. Then there exists a degree-d normalized polynomial q that sign-represents f with bias at least $\beta/\sqrt{N}$, and whose coefficients (in absolute value) form a probability distribution.*

Proof. Let $p(x) = \sum_S \hat{p}(S) x_S$ be the Fourier representation of p. Define

$$\begin{aligned} P &= \sum_S |\hat{p}(S)| \\ &\leq \sqrt{N} \sqrt{\sum_S |\hat{p}(S)|^2} \\ &= \sqrt{N} \sqrt{\sum_x p(x)^2 / 2^n} \\ &\leq \sqrt{N}. \end{aligned}$$

Here the first inequality is Cauchy-Schwarz, the last equality is Parseval's identity, and the last inequality is because p is normalized. We just define $q = p/P$. Then q sign-represents f with bias β/P, and it is normalized because $p(x) \leq P$ for all x. Clearly

$$\sum_S |\hat{q}(S)| = \sum_S |\hat{p}(S)|/P = 1,$$

so the $|\hat{q}(S)|$ form a probability distribution. □

Note that the polynomial q constructed in the above lemma can be viewed as a randomized decision tree of depth d: pick set S with probability $|\hat{q}(S)|$, query its variables, and output $sign(\hat{q}(S)) x_S$. This will compute f with success probability at least $1/2 + 1/2\sqrt{N}$.

The worst-case bias $\min_x |q(x)|$ of q could be as low as $\beta/\sqrt{N}$. However, its *average-case* bias can be lower bounded as follows:

$$\begin{aligned} \overline{\beta} &= \frac{1}{2^n} \sum_x |q(x)| \\ &\geq \frac{1}{2^n} \sum_x q(x)^2 \\ &= \sum_S |\hat{q}(S)|^2 \\ &\geq \frac{(\sum_S |\hat{q}(S)|)^2}{N} \\ &= \frac{1}{N}. \end{aligned}$$

Here the first inequality is because q is normalized, the second equality is Parseval's identity, and the last inequality is Cauchy-Schwarz. Note that the lower bound is independent of the worst-case bias β of the initial polynomial p. For instance, even if the initial β is double-exponentially small, we can construct from this a polynomial (and randomized decision tree) whose average-case bias is at worst exponentially small in $sdeg(f)$.

Corollary 2. *Every $f : \{\pm 1\}^n \to \{\pm 1\}$ can be sign-represented by a normalized polynomial q of degree $sdeg(f)$ with average-case bias at least $1/\sum_{i=0}^{sdeg(f)} \binom{n}{i}$.*

4.2.2 Tightness

We now show that this general lower bound is at most about quadratically far from optimal. We will need the m-bit majority function $\mathrm{MAJ}_m : \{\pm 1\}^m \to \{\pm 1\}$, defined as the sign of the sum of its m inputs.

Theorem 2. *Let $n = dm$ for odd m, and consider a function $f : \{\pm 1\}^n \to \{\pm 1\}$ that is the parity of d independent m-bit majorities. Then $sdeg(f) = d$, and there is a degree-d normalized polynomial sign-representing f with average-case bias $1/\Theta(m)^{d/2}$. Conversely, every degree-d normalized polynomial that sign-represents f, has average-case bias at most $1/\Theta(m)^{d/2}$.*

Before we prove this, note that $1/\Theta(m)^{d/2}$ is roughly $1/\sqrt{\binom{n}{d}}$, matching our general lower bound up to a square. In fact, reformulated as a bound on the average *squared* bias, our results are essentially tight.

Proof. Write the input as $x = x_1 \dots x_d$ with $x_i = x_{i1} \dots x_{im} \in \{\pm 1\}^m$, so $f(x) = \prod_{i=1}^{d} \mathrm{MAJ}_m(x_i)$.

The degree-1 normalized polynomial $\sum_{j=1}^{m} x_{ij}/m$ sign-represents majority on the ith input block (because m is odd, the polynomial is never 0). Hence the following is a degree-d normalized polynomial that sign-represents f:

$$\prod_{i=1}^{d}\left(\sum_{j=1}^{m} x_{ij}/m\right).$$

We can embed a d-bit parity in this function: in each block, fix $(m-1)/2$ input variables to 1 and $(m-1)/2$ to -1, leaving one variable to determine the majority value of that block. Since parity needs maximal sign-degree, it follows that $sdeg(f) \geq d$ and hence $sdeg(f) = d$.

The worst-case bias of our polynomial is $1/m^d$, since each of the d factors can be as small as $1/m$. It is well known that the sum of m uniformly distributed ± 1-variables has expectation $\Theta(\sqrt{m})$ (in fact, the theory of random walks on the line says this expectation goes to $\sqrt{2m/\pi}$ for large m). Hence for a uniformly random input, each $|\sum_j x_{ij}/m|$ has expectation $1/\Theta(\sqrt{m})$. Since the expectation of the product of independent random variables is the product of the expectations, the average-case bias of our polynomial is

$$\prod_{i=1}^{d}\frac{1}{\Theta(\sqrt{m})} = \frac{1}{\Theta(m)^{d/2}}.$$

It remains to upper bound the average-case bias of degree-d sign-representing polynomials for f. Let $p = \sum_S \hat{p}(S)x_S$ be such a polynomial, with average-case bias

$$\overline{\beta} = \frac{1}{2^n}\sum_x |p(x)| = \frac{1}{2^n}\sum_x f(x)p(x) = \sum_S \hat{p}(S)\langle f, x_S\rangle. \tag{4}$$

Let U be the collection of all m^d sets of variables containing exactly one variable from each of the d blocks. We can partition any set S of variables as $S = S_1 \cup \cdots \cup S_d$, where S_i are the variables from block i. If $|S| \leq d$ and $S \notin U$, then at least one S_i will be empty, and we have $\langle \text{MAJ}_m, x_{S_i}\rangle = \frac{1}{2^m}\sum_{x_i \in \{\pm 1\}^m} \text{MAJ}_m(x_i)\cdot 1 = 0$ because majority on an odd number of bits has equally many $+1$-inputs as -1-inputs. Hence for such S we have:

$$\langle f, x_S\rangle = \prod_{i=1}^{d}\langle \text{MAJ}_m, x_{S_i}\rangle = 0.$$

On the other hand, if $S \in U$ then $|S_i| = 1$ for all i. The inner product of MAJ_m with any one of its variables (say the first one) is

$$\begin{aligned}
\langle \text{MAJ}_m, x_{\{1\}}\rangle &= \frac{1}{2^m}\sum_{z\in\{0,1\}^m} \text{MAJ}_m(z)z_1 \\
&= \frac{1}{2^m}\sum_{z:|z_2\ldots z_m|=(m-1)/2} \text{MAJ}_m(z)z_1 + \\
&\quad \frac{1}{2^m}\sum_{z:|z_2\ldots z_m|\neq(m-1)/2} \text{MAJ}_m(z)z_1 \\
&= \frac{1}{2^m}\sum_{z:|z_2\ldots z_m|=(m-1)/2} 1 \\
&= \frac{2}{2^m}\binom{m-1}{(m-1)/2} \\
&= \Theta(1/\sqrt{m}).
\end{aligned}$$

The third equality holds because if $|z_2\ldots z_m| = (m-1)/2$ then $\text{MAJ}_m(z) = z_1$, while if $|z_2\ldots z_m| \neq (m-1)/2$ then $\text{MAJ}_m(z)$ is independent of z_1. Hence for $S \in U$ we have

$$\langle f, x_S\rangle = \prod_{i=1}^{d}\langle \text{MAJ}_m, x_{S_i}\rangle = \frac{1}{\Theta(m)^{d/2}}.$$

Equation (4) thus becomes

$$\overline{\beta} = \frac{1}{\Theta(m)^{d/2}}\sum_{S\in U}\hat{p}(S). \tag{5}$$

It remains to bound $\sum_{S\in U}\hat{p}(S)$. To that end, define a d-variate multilinear polynomial q by

$$q(y_1,\ldots,y_d) = p(y_1^m,\ldots,y_d^m).$$

That is, we substitute the variable y_i for each of the m variables x_{ij}. Note that if a monomial in p contains some x_{ij} and $x_{ij'}$, then the degree of this monomial will decrease under this substitution (both variables will be replaced by y_j, and $y_j^2 = 1$). Hence the only degree-d monomials of p whose degree does not decrease under this substitution, are the ones containing exactly one variable from each of the d blocks, i.e. the monomials x_S with $S \in U$. The substitution maps all such x_S to the same degree-d monomial $y_1\cdots y_d$. Accordingly, the coefficient $\hat{q}([d])$ of that monomial in q will be $\sum_{S\in U}\hat{p}(S)$. Because p is normalized, q is normalized as well, and we have

$$\begin{aligned}
\left(\sum_{S\in U}\hat{p}(S)\right)^2 &= \hat{q}([d])^2 \\
&\leq \sum_{T\subseteq[d]}\hat{q}(T)^2 \\
&= \frac{1}{2^d}\sum_{y\in\{\pm1\}^d} q(y)^2 \\
&\leq 1,
\end{aligned}$$

where the last equality is Parseval's identity. Combining this with Eqn (5) proves the last part of the theorem. □

5 Future work

We mention the following open problems:

- Another communication complexity class question that has been open since it was first stated by Babai et al. [6], is to separate Σ_2 and Π_2 (and other classes in PH). Could our techniques help there?
- How does the tradeoff between degree and bias change if one allows degrees higher than $sdeg(f)$?

Acknowledgments. We thank Hartmut Klauck for answering a question about PP^{cc} vs UPP^{cc}, Adi Shraibman for sending a version of [22], and Alexander Sherstov for comments.

References

[1] S. Aaronson. Quantum computing, postselection, and probabilistic polynomial-time. In *Proceedings of the Royal Society*, volume A461(2063), pages 3473–3482, 2005.

[2] A. Aho, J. Ullman, and M. Yannakakis. Notions of information transfer in VLSI circuits. In *Proceedings of 15th ACM STOC*, pages 133–139, 1983.

[3] E. Allender. A note on the power of threshold circuits. In *Proceedings of 30th IEEE FOCS*, pages 580–584, 1989.

[4] N. Alon, P. Frankl, and V. Rödl. Geometrical realization of set systems and probabilistic communication complexity. In *Proceedings of 26th FOCS*, pages 277–280, 1985.

[5] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):1–14, 1994.

[6] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proceedings of 27th IEEE FOCS*, pages 337–347, 1986.

[7] R. Beigel. Perceptrons, PP, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[8] R. Beigel, N. Reingold, and D. Spielman. PP is closed under intersection. *Journal of Computer and Systems Sciences*, 50(2):191–202, 1995.

[9] C. Damm, M. Krause, C. Meinel, and S. Waack. On relations between counting communication complexity classes. *Journal of Computer and Systems Sciences*, 69(2):259–280, 2004.

[10] H. Ehlich and K. Zeller. Schwankung von Polynomen zwischen Gitterpunkten. *Mathematische Zeitschrift*, 86:41–44, 1964.

[11] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *Proceedings of 16th IEEE Conference on Computational Complexity*, pages 100–106, 2001.

[12] L. Fortnow and N. Reingold. PP is closed under truth-table reductions. *Information and Computation*, 124(1):1–6, 1996.

[13] B. Halstenberg and R. Reischuk. Relations between communication complexity classes. *Journal of Computer and Systems Sciences*, 41(3):402–429, 1990.

[14] B. Halstenberg and R. Reischuk. Different modes of communication. *SIAM Journal on Computing*, 22(5):913–934, 1993.

[15] J. Håstad. On the size of weights for threshold gates. *SIAM Journal on Discrete Mathematics*, 7(3):484–492, 1994.

[16] M. Karchmer, I. Newman, M. Saks, and A. Wigderson. Non-deterministic communication complexity with few witnesses. *Journal of Computer and Systems Sciences*, 49(2):247–257, 1994. Earlier version in Structures'92.

[17] H. Klauck. Lower bounds for quantum communication complexity. In *Proceedings of 42nd IEEE FOCS*, pages 288–297, 2001.

[18] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. In *Proceedings of 45th IEEE FOCS*, pages 12–21, 2004.

[19] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[20] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 2006. To appear.

[21] N. Linial and A. Shraibman. Learning complexity vs. communication complexity. Manuscript available at Linial's homepage, 2006.

[22] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of 39th ACM STOC*, 2007.

[23] S. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. *Journal of Computer and Systems Sciences*, 63(3):449–473, 2001. Earlier version in FOCS'95.

[24] S. Muroga. *Threshold logic and its applications*. Wiley-Interscience, 1971.

[25] R. O'Donnell and R. Servedio. Extremal properties of polynomial threshold functions. In *Proceedings of 18th IEEE Conference on Computational Complexity*, pages 3–12, 2003.

[26] R. Paturi and J. Simon. Probabilistic communication complexity. *Journal of Computer and Systems Sciences*, 33(1):106–123, 1986. Earlier version in FOCS'84.

[27] V. Podolski. Personal communication, unpublished manuscript. 2007.

[28] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya of the Russian Academy of Sciences, mathematics*, 67(1):159–176, 2003.

[29] T. J. Rivlin and E. W. Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on Numerical Analysis*, 3(2):311–320, 1966.

[30] R. Servedio. Every linear threshold function has a low-weight approximator. In *Proceedings of 21st IEEE Conference on Computational Complexity*, pages 18–32, 2006.

[31] A. Sherstov. Halfspace matrices. In *Proceedings of 22nd IEEE Conference on Computational Complexity*, 2007.

[32] A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of 39th ACM STOC*, 2007.

[33] A. C.-C. Yao. Some complexity questions related to distributive computing. In *Proceedings of 11th ACM STOC*, pages 209–213, 1979.