Rectangle Size Bounds and Threshold Covers in Communication Complexity

Hartmut Klauck *
School of Mathematics
Institute for Advanced Study
Princeton, NJ08540, USA
klauck@ias.edu

Abstract

We investigate the power of the most important lower bound technique in randomized communication complexity, which is based on an evaluation of the maximal size of approximately monochromatic rectangles, with respect to arbitrary distributions on the inputs. While it is known that the 0-error version of this bound is polynomially tight for deterministic communication, nothing in this direction is known for constant error and randomized communication complexity. We first study a one-sided version of this bound and obtain that its value lies between the MA- and AM-complexities of the considered function. Hence the lower bound actually works for a (communication) complexity class between $MA \cap co-MA$ and $AM \cap co-AM$, and allows to show that the MA-complexity of the disjointness problem is $\Omega(\sqrt{n})$. Following this we consider the conjecture that the lower bound method is polynomially tight for randomized communication complexity. First we disprove a distributional version of this conjecture. Then we give a combinatorial characterization of the value of the lower bound method, in which the optimization over all distributions is absent. This characterization is done by what we call a bounded error uniform threshold cover, and reduces showing tightness of the bound to the construction of an efficient protocol for a specific communication problem. We then study relaxations of bounded error uniform threshold covers, namely approximate majority covers and majority covers, and exhibit exponential separations between them. Each of these covers captures a lower bound method previously used for randomized communication complexity.

*Supported by NSF grant CCR-9987845. Work partially done while at CWI and supported by the EU 5th framework program QAIP IST-1999-11234 and by NWO grant 612.055.001.

1 Introduction

Communication complexity has grown into a central area in theoretical computer science since the seminal article by Yao [Y79], finding more and more applications that range from VLSI resource-tradeoffs (e.g. [T79]) to data-stream computations (e.g. [BJKS02]), see the excellent monograph [KN97] for pre-1997 applications. While communication complexity has been helpful by inspiring upper bounds in other models, its main importance lies in the lower bounds it provides. Often also variations of the techniques first devised for communication complexity become important in other areas, e.g. the field of branching programs (see [W00]).

Lower bounds for deterministic communication complexity are usually not very hard to prove, but they are often not strong enough in applications. Considering randomized communication complexity is frequently necessary. The lower bound on the monotone circuit depth of the matching function given in [RW92] is an example where randomized communication complexity is used to prove a lower bound for some resource in a deterministic model. Furthermore, since randomized algorithms are considered standard today and any problem for which we can describe an efficient randomized algorithm is considered tractable, lower bounds on the randomized communication complexity are necessary to show that a communication problem is hard. Also, communication complexity is an interesting scenario to study the power of randomization.

Basically all known lower bounds on randomized communication complexity can be derived by considering properties of rectangles in the communication matrix. Following Yao [Y83] these proofs are usually done in two steps. First, so-called distributional communication complexity is considered. The distributional deterministic communication complexity with error ϵ under a distribution μ on the inputs is the minimal complexity of a deterministic protocol computing a function while erring with probability at most ϵ under μ. According to the Yao-principle the randomized complexity of a problem equals the maximum over all distributions of the distributional deterministic complexity. Hence this first step can always be done without loss of generality (or degradation of the bounds).

After choosing an appropriate distribution on the inputs one is left to analyze the deterministic distributional communication complexity. The 2^c message sequences used by a communication c protocol partition the communication matrix into 2^c rectangles labelled with the output of the protocol on that message sequence. Proving a lower bound on the number of rectangles needed in such a partition is then done by showing that all $1-\epsilon$-correct 1- or 0-rectangles are small. This approach or variants of it have been used by Yao [Y83], Babai et al. [BFS86], Razborov [R92], and adapted to partial functions also by Raz [R99], so that virtually all important lower bound proofs (except [KS92] and [HS96]) for randomized communication complexity follow the described pattern.

More precisely, the lower bound method goes as follows: First one fixes a distribution that puts roughly as much weight on the 1-inputs as on the 0-inputs of a function f. One decides whether one proves a bound on the size of rectangles containing predominantly 1-inputs or 0-inputs. Then one shows that all rectangles of the desired type with size larger than $1/2^k$ must contain an ϵ-fraction of wrongly classified inputs. As the consequence the randomized communication complexity of f is $\Omega(k)$.

The main question motivating this paper is whether this lower bound method is tight, i.e., whether we may *always* prove lower bounds at most polynomially smaller than the actual randomized communication complexity using this method. This question is stated as Open Problem 3.23 in [KN97]. We believe this is one of the most important open questions regarding the foundations of communication complexity. Answering the question in the affirmative demands showing an upper bound on the randomized communication complexity in terms of the maximal value obtained by the lower bound method.

It is well known that this is possible in the case $\epsilon = 0$, i.e., the corresponding lower bound method for deterministic protocols based on the size of monochromatic rectangles always yields results being at most quadratically smaller that the deterministic communication complexity. This result can be proved in two steps: first the 0-error rectangle bound is characterized via nondeterministic communication complexity (see Theorem 2.16 in [KN97]). Then the deterministic communication complexity is upper bounded by the product of the nondeterministic and co-nondeterministic communication complexities [AUY83], Theorem 2.11 in [KN97]. In this paper we consider the analogous question in the situation when the error probability is larger than 0. The first 5 sections of this paper can be read as an extended abstract. Section 7 contains the definitions concerning the communication complexity model, all proofs are provided in the sections 8–11.

2 Power of the rectangle bound

First let us fix some notation and give a formal definition of the main lower bound method investigated in this paper.

Definition 1 *Let μ be a distribution on $\{0,1\}^n \times \{0,1\}^n$ and $\alpha \le 1/2$. Then μ is α-balanced for $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, if*

$$\alpha \le \mu(f^{-1}(1)), \mu(f^{-1}(0)) \le 1-\alpha.$$

$1/2$-balanced distributions are called strictly balanced, $1/4$-balanced distributions are called balanced.

Definition 2 *The communication matrix of a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a matrix M_f with rows and columns each corresponding to $\{0,1\}^n$, and with $M_f(x,y) = f(x,y)$.*

A rectangle is a product set in $\{0,1\}^n \times \{0,1\}^n$. Rectangles are usually labelled, a v-rectangle being labelled with $v \in \{0,1\}$. $v(R)$ gives the label of R.

The size of a rectangle (or any other set) R with regard to some distribution μ on $\{0,1\}^n \times \{0,1\}^n$ is $\mu(R) = \sum_{x,y\in R} \mu(x,y)$. Let $err(R,\mu,v) = \mu(f^{-1}(1-v)|R)$ denote the error of a v-rectangle R.

We consider both one-sided and two-sided versions of the rectangle bound.

Definition 3

$$size(\mu,\epsilon,f,v) = \max\{\mu(R) : err(R,\mu,v) \le \epsilon\},$$

where R runs over all rectangles in M_f.

$$bound^{(1)}_\epsilon(f) = \max_\mu \log(1/size(\mu,\epsilon,f,1)),$$

where μ runs over all balanced distribution on $\{0,1\}^n \times \{0,1\}^n$. Furthermore

$$bound_\epsilon(f) = \max\{bound^{(1)}_\epsilon(f), bound^{(1)}_\epsilon(\neg f)\}.$$

We use the conventions $bound(f) = bound_{1/4}(f)$ and $bound^{(1)}(f) = bound^{(1)}_{1/4}(f)$.

Let us first note a fundamental property of the rectangle bound, namely error reducibility.

Lemma 1 *Let $\epsilon \le 1/2 - \Omega(1)$. Assume $bound^{(1)}_\epsilon(f) = k$. Then $bound^{(1)}_{\epsilon^l}(f) \le O(lk)$.*

Assume $bound_\epsilon(f) = k$. Then $bound_{\epsilon^l}(f) \le O(lk)$.

Also note that the definition is almost invariant with respect to the "balance" of the underlying distribution.

Lemma 2 *Let* $\max_\mu \log(1/size(\mu, \epsilon, f, 1)) = k$ *where* μ *runs over all* α*-balanced distribution on* $\{0,1\}^n \times \{0,1\}^n$ *for some constant* $\alpha \geq 4\epsilon$*. Then*

$$k = \Theta(bound_\epsilon^{(1)}(f)).$$

For definitions of the different communication complexity modes considered in this paper see section 7. Note that all randomized modes of communication complexity are defined to have a public coin there.

It is well known that $bound(f)$ yields a lower bound on the randomized communication complexity $R_\epsilon(f)$, see [KN97].

Method 1 (ϵ-error complexity)

1. *Pick a balanced distribution on* $\{0,1\}^n \times \{0,1\}^n$.
2. *Pick* $v \in \{0,1\}$.
3. *Show that all* $1-\epsilon$*-correct* v*-rectangles in* M_f *have size* $< 2^{-b}$.
4. *Then* $R_\epsilon(f) \geq \Omega(b)$.

As an example we give the following fact due to Razborov [R92], which will be used several times in this paper. Let $DISJ(x,y) = \wedge_{i=1}^n (\neg x_i \vee \neg y_i)$ be the set disjointness problem.

Fact 1 *For* $DISJ$ *there is an balanced distribution* μ *on* $\{0,1\}^n \times \{0,1\}^n$*, so that every 0-rectangle* R *in* $\{0,1\}^n \times \{0,1\}^n$ *either satisfies* $\mu(R) \leq 2^{-\beta n}$ *or* $err(R, \mu, 1) \geq \epsilon$ *for some constants* $\beta, \epsilon > 0$.

In other words (using Lemma 1), $R(DISJ) \geq bound^{(1)}(DISJ) = \Omega(n)$.

We now try to determine exactly for which class of problems the lower bound method works. We show that $bound^{(1)}(f)$ lies between the MA- and AM-communication complexities of f, see section 7 for definitions.

Theorem 1 *1. For* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *and* $\epsilon \in [1/2^{MA(f)}, 1/2 - \Omega(1)]$:

$$MA(f) \geq \Omega\left(\sqrt{bound_\epsilon^{(1)}(f)}\right).$$

2. For all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *and* $\epsilon \leq 1/2 - \Omega(1)$*:*

$$AM_\epsilon(f) \leq O(bound_\epsilon^{(1)}(f) + \log(1/\epsilon)).$$

With this theorem and Fact 1 we can conclude a new lower bound in communication complexity.

Corollary 1

$$MA(DISJ) = \Omega(\sqrt{n}),$$

while

$$N(\neg DISJ) = O(\log n).$$

It seems unlikely that, but remains unknown whether $DISJ$ has efficient AM-protocols. Actually no separation between larger classes than $MA \neq co{-}MA$ (shown above) is known within the communication complexity version of the polynomial hierarchy (see [BFS86], polynomial time is replaced by polylogarithmic communication in this definition). Note that it is still open whether the polynomial hierarchy in communication complexity is strict.

We can conclude the following relations between Arthur Merlin communication and the lower bound method.

Corollary 2

$$\max\{MA(f), MA(\neg f)\} \geq \Omega\left(\sqrt{bound(f)}\right),$$

$$bound(f) \geq \Omega(\max\{AM(f), AM(\neg f)\}).$$

If we could show that any function with both small AM- and $co-AM$-complexity also has small randomized complexity we could show that lower bound method 1 is always polynomially tight. If, on the other hand, $R(f) \geq g(bound(f))$ for some superpolynomial g and some f, then there is a superpolynomial separation between $\max\{AM(f), AM(\neg f)\}$ and $R(f)$.

The first attempt to prove tightness of $bound(f)$ one would try uses the Yao-principle and switches to (non-) deterministic distributional complexity with error in the hope to employ similar techniques as in previous combinatorial results [AUY83], where $D(f) \leq O(N(f) \cdot N(\neg f))$ is shown for all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$.

It is well known (see Theorem 3.20 in [KN97]) that

Fact 2

$$R_\epsilon(f) = \max_\mu D_\epsilon^\mu(f).$$

We observe that by the same proof

Lemma 3

$$AM_\epsilon(f) = \max_\mu N_\epsilon^\mu(f).$$

Hence if we could relate the ϵ-error distributional nondeterministic complexity $N_\epsilon^\mu(f) + N_\epsilon^\mu(\neg f)$ to the ϵ-error distributional deterministic complexity $D_\epsilon^\mu(f)$ for all distributions μ we would have shown that the rectangle bound is always polynomially tight. But the approach does not work as shown in the next result.

Theorem 2 *There is a function $WHICH : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, a balanced distribution μ on $\{0,1\}^n \times \{0,1\}^n$, and a constant $\epsilon > 0$ so that*

$$N_0^\mu(WHICH), N_0^\mu(\neg WHICH) = O(\log n),$$

$$D_\epsilon^\mu(WHICH) = \Omega(n).$$

So this first attempt to prove that the rectangle bound is tight, fails. Proving $R(f) \leq poly(AM(f) + AM(\neg f))$ requires an argument not considering the distributional complexity for all distributions separately. Also note that the distributions maximizing $N_\epsilon^\mu(f), N_\epsilon^\mu(\neg f), D_\epsilon^\mu(f)$ are in general not the same.

The proof of Theorem 2 establishes a lower bound on the number of rectangles needed to partition (with small error) the communication matrix into rectangles, while errorfree covers (with overlapping rectangles), and hence large errorfree 1- and 0-rectangles exist.

3 The rectangle bound and bounded error uniform threshold covers

In this section we start another approach to prove that the lower bound method is tight. Instead of considering Arthur Merlin complexity we characterize the lower bound method itself combinatorially.

Definition 4 *A uniform threshold cover with parameters s, t for a communication problem $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is a set of rectangles with labels from $\{0,1\}$ in the communication matrix of f, so that for each input x, y at least t of the adjacent rectangles bear the correct label $f(x,y)$ and at most s of the adjacent rectangles bear the wrong label $1 - f(x,y)$.*

A one-sided uniform threshold cover is as above, but only 1-labelled rectangles are used, and inputs with $f(x,y) = 1$ lie in at least t rectangles, while inputs with $f(x,y) = 0$ lie in at most s rectangles.

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a communication problem. Let P denote the minimal size of a one-sided uniform threshold cover with parameters s, t for f. Then $UT_{s,t}^{(1)}(f) = \lceil \log P \rceil$ is called the one-sided uniform threshold complexity of f.

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a communication problem. Then $UT_{s,t}(f) = \max\{UT_{s,t}^{(1)}(f), UT_{s,t}^{(1)}(\neg f)\}$ is called the uniform threshold complexity of f, and equals (within ± 1) the logarithm of the minimal size of a uniform threshold cover for f.

We will say that a uniform threshold cover with parameters s, t has bounded error, if $t \geq 2s$.

The main features of bounded error uniform threshold covers are first, that the acceptance threshold is the same for *all* inputs, and secondly, the bounded error.

Remark 1 *Given a bounded error uniform threshold cover for f of size 2^k with the parameters $s, 2s$, we can form all possible $l = \log(1/\epsilon)$ tuples of 1-rectangles, and take the intersections of the rectangles in such tuples into a new cover, labelled as 1-rectangles. Then we proceed analogously with the 0-rectangles. Clearly each input is in $(2s)^l = (1/\epsilon) \cdot s^l$ correctly labelled rectangles, and in at most s^l incorrectly labelled rectangles. Hence, there is a value $s' = (1/\epsilon)s^l$ with $UT_{\epsilon s', s'}(f) \leq O(k \cdot \log(1/\epsilon))$.*

We now characterize the lower bound method in terms of bounded error uniform threshold covers.

Theorem 3

1. (a) *bound*$^{(1)}(f) \leq O(UT_{s,2s}^{(1)}(f))$.

   (b) *bound*$(f) \leq O(UT_{s,2s}(f))$.

2. (a) $UT_{n,n^2}^{(1)}(f) \leq O($*bound*$^{(1)}(f) \cdot \log n)$.

   (b) $UT_{n,n^2}(f) \leq O($*bound*$(f) \cdot \log n)$.

Note that $UT_{s,2s}^{(1)}(f) \leq O(bound(f))$ is not always true, as we will show after Theorem 4 in Remark 3.

For upper bounding $UT_{n,n^2}^{(1)}(f)$ in terms of $bound^{(1)}(f) \cdot \log n$ we describe an algorithm that constructs a uniform threshold cover. The algorithm starts with the distribution that is uniform on 1-inputs and uniform on 0-inputs with probability 1/2 each, then iteratively picks large almost 1-chromatic rectangles into the cover and adjusts the distribution by increasing the weights of wrongly covered 0-inputs and decreasing the weights of correctly covered 1-inputs. This can be done in a way so that the underlying distribution stays balanced all through the algorithm. Since there are much more 1-inputs in the rectangles than 0-inputs, the weight of a 0-input doubles each time it is covered, and hence this cannot happen too often. On the other hand when the algorithm terminates all 1-inputs are covered often enough.

So we have a quite natural version of covers that captures the technique used in most of the lower bounds for randomized communication complexity. Showing that $R(f) \leq poly(UT_{n,n^2}(f))$ for all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ would immediately show tightness of $bound(f)$. We have been unable to prove such a result so far. Nevertheless, Theorem 3 turns the problem of showing tightness of the lower bound method into a combinatorial one, not involving a maximization over distributions. Alternatively we may reformulate the problem as follows.

Corollary 3 *Let $UT[R] : R \to \{0,1\}$ for a rectangle $R \subseteq \{0,1\}^N \times \{0,1\}^N$ be the following communication problem* *(for some value t depending on R):*

$$UT[R](x,y) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N/2} x_i \wedge y_i \geq t \\ & \text{and } \sum_{i=N/2+1}^{N} x_i \wedge y_i \leq \sqrt{t} \\ 0 & \text{if } \sum_{i=1}^{N/2} x_i \wedge y_i \leq \sqrt{t} \\ & \text{and } \sum_{i=N/2+1}^{N} x_i \wedge y_i \geq t \\ \textit{undef} & \text{else.} \end{cases}$$

A protocol for computing $UT[R]$ works under the promise that R contains only defined inputs. A randomized protocol computes UT, iff for each R (that contains only defined inputs) the players (knowing R, t) compute $UT[R]$ correctly on each input with high probability. Then

$$R(UT) \leq poly(\log N)$$
$$\iff \forall f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\} :$$
$$R(f) \leq poly(bound(f) \log n).$$

Note that for the communication problem UT, the threshold t is usually much smaller than N.

4 Comparing different notions of threshold covers

4.1 Majority Covers

We now consider variations of the notion of threshold covers. The most immediate and general is a majority cover.

Definition 5 *A majority cover for a function f is a set of labelled rectangles so that for each input the majority of the adjacent rectangles bears the correct label. Ties are broken in favor of $f(x,y) = 1$.*

Let $PP(f)$ denote the logarithm of the size of a smallest majority cover for a function f.

The above notion of majority covers corresponds to majority nondeterministic protocols, which accept an input, whenever there are more nondeterministic computations leading to acceptance than to rejection: each computation in a nondeterministic protocol corresponds to a rectangle. Majority covers are also equivalent to randomized protocols with error "moderately" bounded away from 1/2 as shown in [HR90]:

Fact 3 *There is a majority cover of size 2^k for $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, iff there are ϵ, c with $c + \log(1/\epsilon) = \Theta(k)$, as well as a randomized protocol with error $1/2 - \epsilon > 0$ and communication c computing f (in the protocol the players are allowed to use a private source of randomness only).*

Note that a randomized protocol can be viewed as a probability distribution on deterministic protocols, and that each deterministic protocol induces a partition of the communication matrix into rectangles labelled with the function value. Then the union of all these rectangles is a uniform threshold cover for f, though not one with bounded error. The number of rectangles used is this cover is upper bounded by the number of message sequences used in the randomized protocol (here we use the fact that the randomized protocol can access private random sources only). Thus:

Corollary 4 $PP(f) \leq O(k) \iff \exists s : UT_{s,s+1}(f) \leq O(k)$.

Hence, if we drop the bounded error feature from uniform threshold covers, we can as well drop the uniformity feature.

It is shown in [Kl01] that majority covers have a strong connection to a lower bound method in communication complexity based on discrepancy.

Definition 6 *Let μ be any distribution on $\{0,1\}^n \times \{0,1\}^n$ and f be any function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Then let*

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

where R runs over all rectangles in M_f. Denote $disc(f) = \min_\mu disc_\mu(f)$.

Fact 4 *For all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$:*

$$\log \frac{1}{disc(f)} \leq PP(f) \leq O(\log \frac{1}{disc(f)} + \log n).$$

Note that discrepancy $1/2^k$ under some distribution essentially means that all rectangles with size at least $1/2^{k/2}$ have error at least $1/2 - 1/2^{k/2}$. We can prove lower bounds for the PP-complexity in the following way.

Method 2 (PP-complexity, discrepancy)

1. *Pick a distribution μ on $\{0,1\}^n \times \{0,1\}^n$.*
2. *Show that all rectangles R in M_f satisfy* $|\mu(R \cap f^{-1}(0)) - \mu(R \cap f^{-1}(1))| < 2^{-b}$.
3. *Then $PP(f) \geq \Omega(b)$.*

Consider the function $MAJ : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with

$$MAJ(x,y) = 1 \iff \sum_{i=1}^{n} (x_i \wedge y_i) \geq n/2.$$

Obviously MAJ has a majority cover of size $O(n)$. It is easy to see that $DISJ$ and its complement can both be reduced to MAJ. Hence we can easily separate majority covers from one-sided bounded error uniform threshold covers using Fact 1: $bound^{(1)}(DISJ) = \Omega(n)$ which implies with Theorem 3 that $UT^{(1)}_{s,2s}(DISJ) = \Omega(n)$.

Corollary 5 *$PP(MAJ) = O(\log n)$.*
$UT^{(1)}_{s,2s}(MAJ) = \Omega(n)$ for all s.

So a majority cover is in fact much stronger than even a one-sided uniform threshold cover with bounded error.

4.2 Approximate Majority Covers

Let us now consider a relaxation of majority covers that has bounded error in some sense. Compared to bounded error uniform threshold covers we now drop the uniformity constraint on the threshold.

Definition 7 *An approximate majority cover is a majority cover in which for each input at least $3/4$ of the adjacent rectangles bear the correct label. Let $APP(f)$ denote the logarithm of the size of a minimal approximate majority cover for f.*

Remark 2 *Note that the parameters $1/4, 3/4$ can be improved to arbitrary constant $\epsilon, 1-\epsilon$ by forming k-tuples of rectangles and taking their intersections as the new approximate majority cover with $k = \log(1/\epsilon)$.*

The definition of approximate majority covers is similar to threshold computations on Turing machines in a class named BPP_{path} as considered in [HHT97]. We prefer our naming to BPP_{path}, since the class has little similarity to BPP and is not defined in terms of paths here. It is shown in [HHT97] that BPP_{path} contains $MA \cup co - MA$ and is hence probably much more powerful than BPP. We immediately get a similar result for communication complexity using Theorems 1 and 3.

Theorem 4 *$APP(f) \leq O(UT^{(1)}_{s,2s}(f))$ for all s.*
$APP(f) \leq O(UT^{(1)}_{s,2s}(\neg f))$ for all s.
$APP(f) \leq \min\{O(MA(f)^2), O(MA(\neg f)^2)\}$.

Remark 3 *It is easy to see that $PP(EQ) = \Theta(\log n)$ for the equality function $EQ(x,y) = 1 \iff x = y$, since $PP(f) \geq \log D(f)$. Hence $UT^{(1)}_{s,2s}(EQ) = \Omega(\log n)$. On the other hand $bound(EQ) = \Theta(1)$, since $R(EQ) = \Theta(1)$ (see Example 3.13 in [KN97] and note that randomized protocols are defined to have a public coin here). Hence the relation $UT^{(1)}_{s,2s}(f) \leq O(bound(f)\log n)$ from Theorem 3 cannot be improved to $\leq O(bound(f))$, but possibly to $\leq O(bound(f) + \log n)$.*

APP-complexity has an interesting connection to a lower bound method as follows.

Theorem 5 *If $APP(f) = k$ then for all balanced distributions μ there is a rectangle of size $1/2^{O(k)}$ with error 1/4.*

If for all balanced distributions μ there is a rectangle of size $1/2^k$ and error $1/4$ then $APP(f) \leq O(k) + \log n$.

The proof of the second statement first applies Yao's principle (extended to two quality parameters) to switch to a construct called a "randomized rectangle", and from there to an approximate majority cover using a discretization.

Thus given that $APP(f)$ is small, there is a large rectangle with small error for each distribution, sometimes a 1-rectangle, sometimes a 0-rectangle. We are lead to the following lower bound method.

Method 3 (APP-complexity)

1. *Pick a balanced distribution on $\{0,1\}^n \times \{0,1\}^n$.*
2. *Show that all $1-\epsilon$-correct rectangles in M_f have size $< 2^{-b}$.*
3. *Then $APP(f) \geq \Omega(b)$.*

Actually it has been shown by Yao in [Y83] that for some f and some balanced distribution neither large $1-\epsilon$-correct 0- nor 1-rectangles exist, hence he demonstrated that this function has linear APP complexity, which is a much stronger result than his conclusion that the function has linear randomized bounded error communication complexity.

4.3 Separations

We give separation results between the different types of covers. First we state that approximate majority covers are actually much more powerful than one-sided bounded error uniform threshold covers and hence also than MA-protocols.

Theorem 6 *There is a function $BOTH : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ so that*

$$APP(BOTH) = O(\log n),$$

$$bound^{(1)}(BOTH) \geq \Omega(n),$$

$$bound^{(1)}(\neg BOTH) \geq \Omega(n).$$

Hence also

$$UT^{(1)}_{s,2s}(BOTH), UT^{(1)}_{s,2s}(\neg BOTH) \geq \Omega(n),$$

$$MA(BOTH), MA(\neg BOTH) \geq \Omega(\sqrt{n}).$$

So approximate majority covers are exponentially more powerful than one-sided bounded error uniform threshold covers for $BOTH$ and its complement. In terms of the lower bound methods this means that for $BOTH$ it is true that for every balanced distribution there is a rectangle of size $1/poly(n)$ with constant error, but there exists a balanced distribution, where all 1-rectangles either have error $1/2 - o(1)$ or size $1/2^{\Omega(n)}$, and there exists a balanced distribution, where all 0-rectangles either have error $1/2-o(1)$ or size $1/2^{\Omega(n)}$.

To complete the picture we compare the power of APP and PP covers.

Theorem 7

$$PP(MAJ) = O(\log n),$$

$$APP(MAJ) = \Omega(n).$$

Note that the above result can be read as saying that for MAJ for all distributions there exists a rectangle with discrepancy $1/poly(n)$ (having hence size $1/poly(n)$ and error $1/2 - 1/poly(n)$), while there is a balanced distribution μ where any rectangle with *constant* error has size $1/2^{\Omega(n)}$.

5 Conclusions

Virtually all[1] known lower bounds on randomized communication complexity in the literature can be seen as instances of Method 1, 2, or 3. We have shown that these three methods have exponential differences in power. It remains open whether Method 1 is polynomially tight for randomized communication complexity. A way to show this is proving that the logarithm of the size of bounded error uniform threshold covers is polynomially related to the randomized communication complexity. This avoids arguing with the optimum (over all balanced distributions) of the lower bound parameter. We have shown in Theorem 2 that arguing for all distributions separately does not yield the desired result.

Methods 2 and 3 have been characterized as more powerful versions of threshold covers. It is interesting that the rectangle based lower bound proofs can be understood in terms of these combinatorial objects that are only in the case of Method 2 known to be directly related to standard communication complexity modes.

The papers [KNSW94] and [HS96] in a sense also deal with threshold covers. In the former nondeterministic protocols are considered, in which each 1-input is covered at least once and at most k times (0-inputs are not allowed to be covered). In the latter paper protocols with limited nondeterminism are considered, i.e., errorless covers of the communication matrix with few deterministic *protocols*, in particular these also yield covers in which an upper bound on the number of times 1-inputs are covered exists.

We now note some further observations and open problems. The most significant open problem related to this paper is whether $UT_{s,2s}(f)$ and $R(f)$ are polynomially related, resp. whether $R(UT) \leq poly(\log N)$.

It can be shown with techniques as in [HHT97] that every f with $APP(f) = poly(\log n)$ is in the polylog-communication complexity polynomial hierarchy, see [BFS86] for a definition of the latter. It seems improbable, however, that the same holds for all functions with $PP(f) = poly(\log n)$, since then the communication complexity version of the polynomial hierarchy would collapse. Note that the separation of the polynomial hierarchy for communication complexity is open. Hence Method 3 allows to show that MAJ is not contained in the class of problems with $APP(f) = poly(\log n)$, the largest class of problems inside the polynomial hierarchy for which such a lower bound is known. Showing that this class is a proper subset of the hierarchy is open. It is also open, which methods might be applied to separate this hierarchy or even to prove superlogarithmic lower bounds for Σ_2-protocols. The fact that the discussed lower bound methods really analyze threshold covers maybe sheds some light on why lower bounds for higher classes in the polynomial hierarchy (starting with AM) are unknown.

Regarding the fine-structure of the relations between the discussed complexity measures there are several open problems. Is it possible to separate AM- from MA-complexity? This could be done by using the rectangle method to separate $AM(f)$ from $UT^{(1)}_{s,2s}(f)$. But it is also possible that $UT^{(1)}_{s,2s}(f)$ is always polynomially related to $AM(f)$. Furthermore *any* lower bounds for AM communication complexity are desirable as progress on the problem of showing lower bounds for higher classes in the polynomial hierarchy. Also a separation of $MA(f)$ from $UT^{(1)}_{s,2s}(f)$ would be interesting.

Another interesting issue is the role of interaction. For nondeterministic communication protocols with only 1 round are optimal, not so for randomized, deterministic, and even communication with limited nondeterminism [KN97, Kl98]. Clearly 1-round AM-protocols are also optimal, but this seems unlikely for MA-protocols. A candidate problem to establish this conjecture is the majority of the outcomes of pointer jumping on $\sqrt{n}$ paths of length k, with the promise that $3/4 \cdot \sqrt{n}$ paths lead to the same output. A randomized protocol with k rounds and $O(k \log n)$ communication can solve this problem, but MA-protocols of complexity $o(\sqrt{n})$ using $k-1$ rounds possibly not.

[1] A recent exception is a $\Omega(n)$ lower bound on the information which must be exchanged in computing $DISJ$ [BJKS02] (and hence on the communication complexity). For product distributions this quantity can also be lower bounded by the square root of the rectangle bound, see [Kl02].

6 Organization of the rest of the paper

In section 7 we formally define the different modes of communication complexity considered in this paper, section 8 provides proofs of elementary properties of the rectangle bound. In section 9 we give the proofs concerning the comparison of the rectangle bound with MA- and AM-communication complexity, and the proof of Theorem 2. Section 10 shows the equivalence between $bound(f)$ and bounded error uniform threshold covers for f. Section 11 shows equivalence between Method 3 and approximate majority covers, and gives separations between the three lower bounds methods resp. the three types of threshold covers.

7 Definitions

We employ the following definitions concerning communication complexity.

Definition 8 *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. In a communication protocol players Alice and Bob receive inputs x resp. y from $\{0,1\}^n$. Their goal is to compute $f(x,y)$. To this end the players exchange binary encoded messages. The communication complexity of a protocol is the worst case number of bits exchanged.*

The deterministic communication complexity $D(f)$ of a function f is the complexity of an optimal protocol computing f.

In a nondeterministic protocol for a Boolean function f the players are allowed to guess some bits and communicate according to a different deterministic strategy for each guess string. An input is accepted iff at least one computation accepts. The nondeterministic guesses are private and accessible to the guessing player only. The nondeterministic communication complexity $N(f)$ is the complexity of an optimal nondeterministic protocol computing f.

In a randomized protocol for a Boolean function f the players can access a public source of random bits. They can communicate according to a different deterministic strategy for each value of the random bits. It is required that for each input the correct output is produced with probability $1-\epsilon$ for some $\epsilon < 1/2$. The randomized communication complexity $R_\epsilon(f)$ is the complexity of an optimal randomized protocol computing f with error probability ϵ.

Arthur Merlin computations have been introduced in [B85, BM88]. In an Arthur Merlin (AM) protocol the players may first access a public source of random bits and read an arbitrarily long random string. After this phase they start a nondeterministic protocol. A function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is computed if for all x, y with $f(x,y) = 1$ with probability $1-\epsilon$ over the random bits the nondeterministic protocol accepts (i.e., there is a guess string makes players accept), while for all x, y with $f(x,y) = 0$ with probability $1-\epsilon$ over the random bits the nondeterministic protocol does not accept (i.e., there is no guess that makes the players accept). The communication complexity of an Arthur Merlin protocol is the maximum (over the random bits) of the complexities of the nondeterministic protocols. Let $AM_\epsilon(f)$ denote the complexity of an optimal Arthur Merlin protocol for f with error ϵ.

In a Merlin Arthur protocol for a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ the players first make a nondeterministic guess of some length k known to both players. Then the players perform a randomized protocol. It is required that for all x, y with $f(x,y) = 1$ there is a value of the guess, so that the protocol accepts with probability $1-\epsilon$, while for all x, y with $f(x,y) = 0$ there is no value of the guess, so that the protocol accepts with probability larger than ϵ. The complexity of a Merlin Arthur protocol is given by the maximum complexity of the communication (over the guesses and the coin tosses) plus k. Let $MA_\epsilon(f)$ denote the complexity of an optimal Merlin Arthur protocol for f with error ϵ.

In case the subscript fixing the error is dropped we set the error to $1/4$.

Note that an Arthur Merlin protocol is an interactive proof system with verification performed by a communication protocol. Arthur challenges Merlin to provide proof for $f(x,y) = 1$, this proof is verified by Alice and Bob.

A Merlin Arthur protocol uses a randomized protocol to check a fixed proof, whose length is included in the communication cost. The Merlin Arthur model would be ill-defined, if we would simply require the nondeterministic guess to be private and coming without cost. In this case Alice could simply guess Bob's input nondeterministically, and then use a randomized protocol for the equality function with $O(1)$ communication to test if her guess was right (see Example 3.13 in [KN97], note that public coin in the randomized protocol). If so, she can compute any function on x, y and announce the result. Hence any function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ would have Merlin Arthur communication complexity $O(1)$ under such a definition.

In lower bound proofs for randomized complexity one often applies the Yao-principle that states a relation between the complexity in the randomized setting (with small error probability for every input) and the complexity in the deterministic setting where correctness is only demanded with high probability over some distribution on the inputs.

Definition 9 *A deterministic protocol has error ϵ under some distribution μ on the inputs, if the probability that the protocol errs is ϵ.*

The distributional deterministic complexity of $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is $D^\mu_\epsilon(f)$, the minimal complexity of any deterministic protocol with error ϵ under μ.

A nondeterministic protocol has error at most ϵ under some distribution μ on the inputs, if the probability (under μ) of the set of accepted 0-inputs and nonaccepted 1-inputs is at most ϵ.

The distributional nondeterministic complexity of $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is $N^{\mu}_{\epsilon}(f)$, the minimal complexity of any nondeterministic protocol with error ϵ under μ.

Nondeterministic communication complexity is related to a specific type of covers [KN97].

Fact 5 *Let $Cov^{(1)}(f)$ denote the minimum number of monochromatic (0-error) 1-rectangles in a set $\{R_1, \ldots, R_c\}$, so that $f^{-1}(1) = \cup_{i=1}^{c} R_i \neq \emptyset$.*

Then $N(f) = \lceil \log Cov^{(1)}(f) \rceil$.

8 Properties of the rectangle bound

Proof of Lemma 1. Assume $bound^{(1)}_{\epsilon}(f) = k$. Hence for all balanced distributions there is a $1-\epsilon$-correct 1-rectangle of size $1/2^k$ at least. Fix any balanced distribution μ. We construct a rectangle of error ϵ^l and size $2^{-O(kl)}$ for μ inductively.

Let $\mu_0 = \mu$. First we take a rectangle R_1 with error $\leq \epsilon$ and size $s_1 \geq 1/2^k$ guaranteed by our assumption for μ_0. In case R_1 has no error at all we are done. Otherwise we construct a new distribution μ_1 as follows: For all $x, y \notin R_1$ we set $\mu_1(x,y) = 0$. μ_1 is then normalized to a strictly balanced distribution by multiplying $\mu_0(x,y)$ by some factor p_0 when $f(x,y) = 0$ and multiplying $\mu_0(x,y)$ by a factor p_1 when $f(x,y) = 1$.

Note that since $\mu(f^{-1}(0)|R) \leq \epsilon$ we have $p_0 \geq 1/(2\epsilon)$ and $1/(2(1-\epsilon)) \geq p_1 \geq 1/2$. Then we can pick a rectangle R_2 with error ϵ and size $s_2 \geq 1/2^k$ according to μ_1.

Now we compute the size of R_2 according to μ and also its error on μ. By concentrating μ_1 on R_1 we have increased the weights of $x, y \in R_1$ uniformly by a factor of $(1/s_1) \leq 2^k$. Then we have balanced the distribution by multiplying 1-inputs' weights with p_1 and 0-inputs' weights by p_0.

So the weight of 0-inputs in R_2 according to μ is at most

$$s_1 \cdot (1/p_0) \cdot \epsilon \cdot s_2 \leq s_1 \cdot 2\epsilon^2 \cdot s_2.$$

The weight of the 1-inputs is at least

$$s_1 \cdot (1/p_1) \cdot (1-\epsilon) \cdot s_2 \geq s_1 \cdot 2(1-\epsilon)^2 \cdot s_2.$$

The size of R_2 is at least $s_1 \cdot s_2 \geq 1/2^{2k}$.

Assume that $\epsilon \geq 1/4$, then $\epsilon = 1/2 - \delta$ for some $\delta \leq 1/4$. In this case $err(R_2, \mu, 1) \leq$

$$\frac{\epsilon^2}{(1-\epsilon)^2 + \epsilon^2} = \frac{1}{2} - \frac{\delta}{1/2 + 2\delta^2} < 1/2 - (3/2)\delta.$$

Repeating this $O(1/\delta) = O(1)$ times reduces the error to less than $1/4$.

If $\epsilon \leq 1/4$, then R_2 has error at most $2\epsilon^2$. Iterating the construction $O(l)$ times yields the first part of the lemma. Arguing analogously for 0-rectangles yields the second part. □

Proof of Lemma 2. Let the bound be k when μ runs over all α-balanced distributions on the inputs.

First assume that $\alpha < 1/4$. Clearly k is an upper bound on $k' = bound^{(1)}(f)$ in this case. We have to show that also $k = O(k')$.

Let μ be an α-balanced distribution on the inputs. We can balance the weights of 1-inputs and 0-inputs by multiplying the weights of the 0-inputs by some p_0 and the weights of the 1-inputs by some p_1 so that a strictly balanced distribution μ' is obtained. We take a rectangle R of error ϵ and size $1/2^{k'}$ according to μ'. Assume that $\alpha \leq \mu(f^{-1}(1)) \leq 1/2$. Then $p_1 > 1$. Consequently the size of R is slightly smaller according to μ than to μ', and the error is possibly slightly smaller, too. In the other case $\alpha \leq \mu(f^{-1}(0)) \leq 1/2$ the opposite occurs, namely the rectangle is possibly slightly larger, and the error is larger. But in any case the size and the error are changed by constant factors $\Theta(\alpha)$ resp. $\Theta(1/\alpha)$ only. If the error of R is too large we can use the previous lemma to reduce the error probability while decreasing the size.

The case $1/4 \leq \alpha \leq 1/2$ is handled similarly. □

9 The rectangle bound versus Arthur Merlin communication complexity

We now prove the relations between Arthur Merlin communication complexity and the lower bound method based on the size of approximately correct rectangles.

Proof of Theorem 1, part 1. We are given a Merlin Arthur protocol with complexity c and error $1/4$ for the function f. We show that under this condition we can find a large $1-\epsilon$-correct 1-rectangle. Recall that the complexity of the Merlin Arthur protocol includes the "communication" done by Merlin (who is guessing nondeterministically) plus the communication by the players Alice and Bob. Let c_M be the length of the longest guess given by Merlin over all inputs. We will call such a guess string a *proof*. Let c_P denote the length of the longest communication between Alice and Bob occurring during any run of the protocol. Clearly $c_M, c_P \leq c$.

It is possible to reduce the error probability of the protocol to $1/2^{2c}$ by repeating the probabilistic part of the protocol $O(c)$ times independently and taking the majority output, for any fixed proof of Merlin. Let $P(x,y,z)$ denote the (random) output of the protocol for inputs x, y and Merlin's proof z. The acceptance properties of the protocol are then:

If $f(x,y)=1$ then there is a proof z so that $P(x,y,z)$ accepts with probability $1-1/2^{2c}$.

If $f(x,y)=0$ then for all proofs z, $P(x,y,z)$ accepts with probability at most $1/2^{2c}$.

Note that the communication among the players in the new protocol is bounded by $k=O(c\cdot c_P)=O(c^2)$.

There are at most 2^{c_M} different proofs. If we fix such a proof z there is a set s_z of 1-inputs that is accepted on this proof, i.e., for which the protocol accepts with high probability on this proof. In this way the set of 1-inputs is covered by 2^{c_M} subsets $s_1,\ldots,s_{2^{c_M}}$.

Let μ be any balanced distribution over the inputs. For each such distribution we can find at least one proof z so that $\mu(s_z)\geq 1/2^{c_M}$. We fix such a proof. This turns the Merlin Arthur protocol into a randomized protocol so that a subset s_z of 1-inputs of weight $1/2^{c_M}$ is accepted with probability $1-1/2^{2c}$ each and no 0-input is accepted with probability larger than $1/2^{2c}$. The other 1-inputs are accepted with uncertain probability.

We restrict μ to the inputs in $s_z\cup f^{-1}(0)$, by setting the weight of all other inputs to 0 and normalizing to a distribution μ'. Clearly the error of the protocol under μ' is at most $1/2^{2c}$. Furthermore note that either

$$\mu'(x,y)=0 \text{ or } \mu'(x,y)=\Theta(\mu(x,y)), \tag{1}$$

since μ is balanced.

Given such a randomized protocol we may also fix its random choices and get a deterministic protocol (like in the easy direction of the Yao-principle). This yields a deterministic protocol which on expectation has error $1/2^{2c}$ (under μ'). Consequently there exists a deterministic protocol with error $1/2^{2c}$ under μ' and communication $k=O(c^2)$.

A deterministic protocol with communication k easily leads to a set R of $P=2^{O(k)}$ pairwise disjoint rectangles labelled with the protocol output that partition the communication matrix. We show that there exists a large 1-rectangle with small error.

Assume that all rectangles which are larger than $1/(2^{c+1}\cdot 2P)$ have error larger than $1/2^c$. Then the success probability of the protocol on μ' is upper bounded as follows. The small rectangles contribute at most success $P\cdot 1/(2^{c+1}\cdot 2P)\cdot 1\leq 1/2^{c+2}$, the large rectangles all have success at most $1-1/2^c$ and so the overall success probability is at most $1-1/2^c+1/2^{c+2}$, too small in comparison to the maximum error $1/2^{2c}$. Hence there is a 1-rectangle of size $1/(2^{c+1}\cdot 2P)\geq 2^{-\Omega(c^2)}$ with error at most $1/2^c\leq\epsilon$ according to μ'. If we switch from μ' to μ, then the size of a rectangle cannot decrease (compared to μ) by more than a constant factor due to (1). It is also easy to see that the error of the rectangle cannot increase when switching from μ' to μ. Consequently $bound_\epsilon^{(1)}(f)\leq O(c^2)$. □

Now we relate $bound^{(1)}(f)$ to $AM(f)$.

Proof of Theorem 1, part 2. Assume that $bound_\epsilon^{(1)}(f)=c$. Then $bound_{\epsilon^4/8}^{(1)}\leq O(c)$. For *all balanced* distributions μ there is a 1-rectangle R_μ with error at most $\epsilon^4/8$ and size at least $s\geq 1/2^{O(c)}$. Also recall that $AM_\epsilon(f)=\max_\mu N_\epsilon^\mu(f)$, where μ runs over *all* distribution on the inputs, due to Lemma 3. We use a greedy algorithm to construct a cover of the 1-inputs to f with error ϵ and at most $2^{O(c)}\cdot(1/\epsilon^2)\cdot\log(1/\epsilon)$ rectangles for any μ.

So let μ be some distribution on $\{0,1\}^n\times\{0,1\}^n$. We distinguish three cases. First consider the case that $\mu(f^{-1}(1))\leq\epsilon^2$. In this case clearly $N_{\epsilon^2}^\mu(f)=0$ by a protocol that never accepts.

Next consider the case that $\mu(f^{-1}(0))\leq\epsilon^2$. In this case $N_{\epsilon^2}^\mu(f)=0$ by a protocol that always accepts.

Now consider the case that $\mu(f^{-1}(1))\geq\epsilon^2$ and $\mu(f^{-1}(0))\geq\epsilon^2$. Then we can still find a good cover as follows. We first show that for each such distribution a relatively large 1-rectangle with small error exists. Then we use a greedy approach to find a cover.

First we show how to find good rectangles. We (strictly) balance the distribution by multiplying the weights of 1-inputs by some value p_1 and multiplying the weights of 0-inputs by some value p_0. Clearly

$$2(1-\epsilon^2)\geq\frac{1}{p_1},\frac{1}{p_0}\geq 2\epsilon^2.$$

For the resulting strictly balanced distribution μ' there is a 1-rectangle $R_{\mu'}$ of size s having error $\epsilon^4/8$ at most. Then the μ-weight of 1-inputs in $R_{\mu'}$ is at least

$$s\cdot(1-\epsilon^4/8)\cdot(1/p_1)\geq s\cdot(1-\epsilon^4/8)\cdot 2\epsilon^2\geq s\epsilon^2.$$

Furthermore the μ-weight of 0-inputs in $R_{\mu'}$ is at most

$$s\cdot(\epsilon^4/8)\cdot(1/p_0)\leq s\cdot\epsilon^4\cdot(1-\epsilon^2)/4\leq s\epsilon^4/4.$$

We next construct for μ a cover using a greedy approach.

1. Let $\mu_0=\mu$.
2. For μ_i find a rectangle $R_i=R_{\mu_i}$ that contains 1-inputs of weight at least $s\epsilon^2$ and 0-inputs of weight at most $s\epsilon^4/4$.
3. Put R_i into the cover.
4. Remove the weight from all 1-inputs in R_i and uniformly increase the weights of the remaining 1-inputs by some appropriate factor $q(i)$. [Note that this does not affect the balance of the distribution.] Let μ_{i+1} denote the resulting distribution.
5. Stop, when the set of remaining 1-inputs not covered so far has weight $\leq\epsilon/2$ according to μ.
6. Otherwise continue with 2. setting $i:=i+1$.

Clearly the algorithm finds a set of rectangles so that all but a set of weight $\epsilon/2$ of the 1-inputs is covered. In the worst case the weight of any 1-input is increased by a factor of $q(1)\cdots q(i) \leq (3/4)/(\epsilon/2) = (3/2)/\epsilon$ during the course of the algorithm. Hence the weight of 1-inputs in R_i according to μ is at least $s\epsilon^2 \cdot \epsilon \cdot (2/3)$, while the weight of 0-inputs is at most $s\epsilon^4/4$. The error of R_i is thus at most $(3/8)\epsilon$. Since this holds for all rectangles, the weight of 0-inputs in the cover is at most a fraction of $(3/8)\epsilon$ of the weight of all 1-inputs covered, which is at most $(3/4)\cdot(3/8)\epsilon < \epsilon/2$. The weight of 1-inputs not covered is $\epsilon/2$. So the obtained cover has error ϵ.

Now we have to analyze the size of the obtained cover. Each step covers at least a $s\epsilon^2$ fraction of the remaining 1-inputs. Hence the proportion of not yet covered 1-inputs according to μ after k steps is

$$w_k \leq (1-s\epsilon^2)^k.$$

The algorithm stops if this is smaller than $O(\epsilon)$, hence $k \leq O(1/(s\epsilon^2)\cdot\log(1/\epsilon))$, and $N^\mu_\epsilon(f) \leq O(c+\log(1/\epsilon))$.

So indeed for all μ we have $N^\mu_\epsilon(f) \leq O(bound^{(1)}_\epsilon(f) + \log(1/\epsilon))$, and hence $AM_\epsilon \leq O(bound^{(1)}_\epsilon(f) + \log(1/\epsilon))$. $\square$

Proof of Theorem 2. We first define the function that has both small nondeterministic and co-nondeterministic complexity with 0 error under some distribution, but large deterministic complexity for some constant error under the same distribution.

Let $WHICH((x_1,x_2),(y_1,y_2)) =$

$$\begin{cases} 1 \text{ if } \neg DISJ(x_1,y_1)=1 \wedge \neg DISJ(x_2,y_2)=0 \\ 0 \text{ if } \neg DISJ(x_1,y_1)=0 \wedge \neg DISJ(x_2,y_2)=1 \\ 0 \text{ otherwise.} \end{cases}$$

We employ the following more specific and optimized version of Fact 1, which follows from some fine-tuning of the result in [R92].

Fact 6 *Let ν_a be the distribution on $\{0,1\}^n\times\{0,1\}^n$ which is uniform on $\{(x,y) : |x|=|y|=n/4, |x\cap y|=0\}$ (and 0 elsewhere), and let ν_r be the distribution on $\{0,1\}^n\times\{0,1\}^n$ which is uniform on $\{(x,y): |x|=|y|=n/4, |x\cap y|=1\}$ (and 0 elsewhere). Let ν be the distribution on $\{0,1\}^n\times\{0,1\}^n$, which is defined by $\nu(x,y)=(3/4)\cdot\nu_a(x,y)+(1/4)\cdot\nu_r(x,y)$.*

Then for any constant $\delta>0$ there is a constant $\beta(\delta)>0$, so that any rectangle R either has size $2^{-\beta(\delta)n}$, or $\nu(x,y : x\cap y\neq\emptyset \,|\, R)\geq 1/4-\delta$, i.e.,

$$\nu(\{(x,y): x\cap y\neq\emptyset\}\cap R) \geq (\frac{1}{4}-\delta)\cdot\nu(R) - 2^{-\beta(\delta)n}.$$

Now to the definition of the distribution on the inputs. In the distribution $\nu\times\nu$ two instances (x_1,y_1) and (x_2,y_2) are chosen independently from ν.

For the hard distribution μ on inputs we pick inputs as in $\nu\times\nu$, but inputs with $DISJ(x_1,y_1)=DISJ(x_2,y_2)$ are removed, so that the function value on the two instances differs with probability 1 (μ is normalized to a distribution after this removal). On μ the task of a protocol is to determine on which of the two set pairs $\neg DISJ$ is true. Note that either $\mu(x_1,x_2,y_1,y_2) = \Theta(\nu\times\nu(x_1,x_2,y_1,y_2)) = \Theta(\nu(x_1,y_1)\cdot\nu(x_2,y_2))$ or $\mu(x_1,x_2,y_1,y_2)=0$, for all inputs (x_1,x_2,y_1,y_2) from $\{0,1\}^{4n}$, since $Prob_\nu(DISJ(x_1,y_1)=1)=3/4$. Also note that μ is strictly balanced.

There is a simple nondeterministic protocol for $WHICH$ making no error under the distribution μ. One can simply use the trivial protocol for $\neg DISJ$ on the first instance. This covers all 1-inputs of $WHICH$, but accepts no 0-input with weight larger than 0. Analogously we can find a protocol for $\neg WHICH$ under μ. So $N^\mu_0(WHICH), N^\mu_0(\neg WHICH)\leq \log n+1$.

Now we turn to the complexity of a deterministic protocol with error. Such a protocol with communication c immediately yields a set of $P=2^{O(c)}$ pairwise *disjoint* rectangles labelled with values 0,1, so that a $1-\epsilon$ fraction of all inputs according to μ are in correctly labelled rectangles. Call the 1-rectangles $R_1,\ldots,R_P$, the 0-rectangles $S_1,\ldots,S_P$.

We show that such a partition can only exist if $c=\Omega(n)$. So for the sake of contradiction assume that $c\leq\gamma n$ for some arbitrarily small constant γ we can choose later.

Note that the difficulty for a deterministic protocol is that the corresponding cover consists of disjoint rectangles, even on the inputs with $\mu(x_1,x_2,y_1,y_2)=0$. In case the reader would prefer to loosen this restriction and require a protocol to be deterministic only on those inputs with $\mu(x_1,x_2,y_1,y_2)>0$ we can still give those inputs some small probability, so that the above nondeterministic protocols would have small error, while the following lower bound on deterministic protocols would be unchanged.

We use the following notation. An input (x_1,x_2,y_1,y_2) is in *quadrant* A, if $\neg DISJ(x_1,y_1)=0$ and $\neg DISJ(x_2,y_2)=0$, in quadrant B, if $\neg DISJ(x_1,y_1)=1$ and $\neg DISJ(x_2,y_2)=0$, in quadrant C, if $\neg DISJ(x_1,y_1)=0$ and $\neg DISJ(x_2,y_2)=1$, and in quadrant D, if $\neg DISJ(x_1,y_1)=1$ and $\neg DISJ(x_2,y_2)=1$. Note that inputs in quadrants A and D have probability 0 under μ. Under $\nu\times\nu$ quadrants B and C have weight $3/16$, quadrant A has weight $9/16$ and quadrant D has weight $1/16$.

Let R_i be a 1-rectangle and $(x_1,x_2,y_1,y_2)\in R_i$ with $\neg DISJ(x_1,y_1)=1$. Then the set $R_i(x_1,y_1)=\{x_2,y_2 : x_1,x_2,y_1,y_2\in R_i\}$ is a rectangle in $\{0,1\}^n\times\{0,1\}^n$. Let

$$\begin{aligned} & \mu(R_i(x_1,y_1)\,|\,x_1,y_1) \\ = \; & \frac{\mu(\{(x_1,y_1)\}\times R_i(x_1,y_1))}{\mu(\{(x_1,y_1)\}\times\{0,1\}^n\times\{0,1\}^n)} \end{aligned}$$

denote the weight of $R_i(x_1, y_1)$ relative to the inputs with fixed x_1, y_1. Note that

$$\mu(\{(x_1, y_1)\} \times \{0,1\}^n \times \{0,1\}^n) = \Theta(\nu(x_1, y_1)).$$

Also $\mu(\{(x_1, y_1)\} \times R_i(x_1, y_1)) \leq$

$$O(\nu(x_1, y_1) \cdot \nu(R_i(x_1, y_1))),$$

since all inputs x_1, x_2, y_1, y_2 with $\neg DISJ(x_2, y_2) = 1$ have weight 0 in μ and all x_1, x_2, y_1, y_2 with $\neg DISJ(x_2, y_2) = 0$ have weight $\mu(x_1, x_2, y_1, y_2) = \Theta(\nu(x_1, y_1) \cdot \nu(x_2, y_2))$. Then

$$\nu(R_i(x_1, y_1)) = \Omega(\mu(R_i(x_1, y_1) \,|\, x_1, y_1)). \qquad (2)$$

We will show that each large $R_i(x_1, y_1)$ must contain many inputs x_2, y_2 with $\neg DISJ(x_2, y_2) = 1$. While this does not create any error in the rectangle R_i, this shows that R_i occupies a significant portion of quadrant D. If this is true for many rectangles R_i then a situation is reached in which the majority of quadrant D is occupied. Since a symmetric argument applies to the 0-rectangles S_i we are lead into a contradiction, since the rectangles are not allowed to intersect nontrivially.

We set $\epsilon = 2\delta = 1/16$ fixing the protocol's error and the constant from Fact 6, and choose $\gamma < \beta(\delta)/2$. Let x_1, y_1 be an input with $\neg DISJ(x_1, y_1) = 1$. If $\mu(R_i(x_1, y_1)|x_1, y_1) \geq \Omega(2^{-\gamma n})$, then with (2) $\nu(R_i(x_1, y_1)) \geq 2^{-\beta(\delta)n}$, and hence the fraction of $x_2, y_2 \in R_i(x_1, y_1)$ with $\neg DISJ(x_2, y_2) = 1$ is at least $1/4 - \delta$ according to ν. In other words, the proportion of 0-inputs to 1-inputs of $\neg DISJ(x_2, y_2)$ in $R_i(x_1, y_1)$ is $3/4 + \delta$ to $1/4 - \delta$.

Hence $\{(x_1, y_1)\} \times R_i(x_1, y_1)$ occupies at least a $(1/4 - \delta)/(3/4 + \delta) \geq 1/3 - 2\delta$ fraction of the weight it covers in quadrant B (according to $\nu \times \nu$) also on the inputs in quadrant D (according to $\nu \times \nu$). Recall that on $\nu \times \nu$ quadrant D has weight $1/16$ and quadrant B has weight $3/16$.

We next show that a fraction of $1 - 2\epsilon$ of all 1-inputs of $WHICH$ (resp. of quadrant B) is covered by $R_i(x_1, y_1)$ with $\mu(R_i(x_1, y_1)|x_1, y_1) \geq \Omega(\epsilon \cdot 2^{-\gamma n})$ and hence $\nu(R_i(x_1, y_1)) \geq 2^{-\beta(\delta)n}$ (using (2)). Then in quadrant D weight at least $(1 - 2\epsilon) \cdot (1/3 - 2\delta) \cdot 3/16 > 1/32$ is occupied by the rectangles R_i (according to $\nu \times \nu$). Since the same holds for the rectangles S_i we get at least one input in quadrant D that is covered twice, a contradiction to the requirements on those rectangles. Therefore $\gamma \geq \beta(\delta)/2 = \Theta(1)$ and hence $c = \Omega(n)$.

The weight of 1-inputs (according to μ) that are in rectangles $R_i(x_1, y_1)$ with $\mu(R_i(x_i, y_i) \,|\, x_1, y_1) \geq \epsilon/const \cdot 2^{-\gamma n}$ (and $\neg DISJ(x_1, y_1) = 1$) is at least $1 - 2\epsilon$, since at most an ϵ fraction of 1-inputs is not covered and the fraction of 1-inputs covered by smaller rectangles can be bounded as follows. Each small rectangle $R_i(x_1, y_1)$ can cover 1-inputs of weight at most $\epsilon/const \cdot 2^{-\gamma n} \cdot \mu(\{(x_1, y_1)\} \times \{0,1\}^n \times \{0,1\}^n)$, and there are at most $2^{\gamma n} \cdot S$ different $R_i(x_1, y_1)$ with $\neg DISJ(x_1, y_1) = 1$, where $S = |\{(x, y) : |x| = |y| = n/4, |x \cap y| = 1\}|$. Note that $\mu(\{(x_1, y_1)\} \times \{0,1\}^n \times \{0,1\}^n) = \Theta(1/S)$. So the 1-inputs covered by small rectangles have weight at most

$$\begin{aligned} & \epsilon/const \cdot 2^{-\gamma n} \\ \cdot \quad & \mu(\{(x_1, y_1)\} \times \{0,1\}^n \times \{0,1\}^n) \cdot 2^{\gamma n} \cdot S \\ \leq \quad & \epsilon. \square \end{aligned}$$

10 Uniform threshold covers and the rectangle bound

Proof of Theorem 3, part 1. Assume that $UT^{(1)}_{s,2s}(f) \leq k$ for some s. Then following Remark 1, $UT^{(1)}_{\epsilon s', s'}(f) \leq O(k)$ for an arbitrarily small constant ϵ and some $s' = O(s)$.

Given a one-sided bounded error uniform threshold cover with $P \leq 2^{O(k)}$ 1-rectangles $S = \{R_1, \ldots, R_P\}$ let $h(x, y)$ denote the number of rectangles R_i the input x, y is included in.

We know that for each x, y with $f(x, y) = 1$ there are at least s' 1-rectangles it is included in, so $h(x, y) \geq s'$. Each x, y with $f(x, y) = 0$ is in at most $\epsilon s'$ 1-rectangles, hence $h(x, y) \leq \epsilon s'$.

Let μ be any balanced distribution on the inputs. We define a probability distribution ν on the 1-rectangles in S as follows. Each rectangle $R \in S$ receives the weight $\sum_{x,y \in R} \mu(x, y)$. We then normalize these weights to a distribution on 1-rectangles in S. The probability of some rectangle S is then

$$\begin{aligned} & \sum_{x,y \in R} \frac{\mu(x, y)}{\sum_{R'} \sum_{x',y' \in R'} \mu(x', y')} \\ = \quad & \sum_{x,y \in R} \frac{\mu(x, y)}{\sum_{x',y' \in \{0,1\}^n \times \{0,1\}^n} \mu(x', y') \cdot h(x', y')}. \end{aligned}$$

If we first pick a rectangle according to ν and then on that rectangle an input (according to μ restricted to R), we get some input x, y with probability

$$\frac{\mu(x, y) \cdot h(x, y)}{\sum_{x',y' \in \{0,1\}^n \times \{0,1\}^n} \mu(x', y') \cdot h(x', y')}.$$

So the weight of x, y in this experiment is proportional to $\mu(x, y) \cdot h(x, y)$. Hence the probability of picking a 0-input in this way is at most

$$\frac{\mu(f^{-1}(0)) \cdot \epsilon s'}{\mu(f^{-1}(1)) \cdot s'} \leq \frac{3/4 \cdot \epsilon s'}{1/4 \cdot s'} \leq 3\epsilon. \qquad (3)$$

Assume that all rectangles that are larger than ϵ^2/P according to μ have error larger than 4ϵ. Then, if we first pick a rectangle R and then an input $x, y \in R$ the probability that $f(x, y) = 1$ can be bounded as follows. The small rectangles contribute at most $P\cdot(\epsilon^2/P)\cdot 1 \le \epsilon^2$ to this probability. All larger rectangles have error 4ϵ at least, and hence when picking one of them the probability of getting a 1-input is at most $1 - 4\epsilon$, so the overall probability of getting a 1-input is at most $1 - 4\epsilon + \epsilon^2$, a contradiction to (3).

Hence there exists a 1-rectangle of size at least $\epsilon^2/P = \Omega(1/P)$ having error at most $O(\epsilon)$ according to μ.

When given a bounded error uniform threshold cover we can do the same construction for the 0-inputs, and hence $UT_{s,2s}(f) = k$ allows us to find both a 1-rectangle and a 0-rectangle with the desired properties for any balanced μ. $\square$

Proof of Theorem 3, part 2.

Assume $bound^{(1)}_{1/4}(f) = k$ for some $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Then $bound^{(1)}_{1/n^5}(f) \le O(k \log n)$ using Lemma 1. In other words, for each balanced distribution μ on $\{0,1\}^n \times \{0,1\}^n$ there exists a $1 - 1/n^5$-correct 1-rectangle of size at least $s = 2^{-O(k \log n)}$.

We show how to construct a one-sided uniform threshold cover with parameters n, n^2. The cover is produced by an algorithm. Let μ_1 be the distribution which is uniform on the 1-inputs of f with probability 1/2 and uniform on the 0-inputs of f with probability 1/2.

1. Set $l = 1$, $Cov_0 = \emptyset$.
2. Find a $1 - 1/n^5$-correct rectangle R_l of size at least s according to μ_l and let $Cov_l = Cov_{l-1} \cup \{R_l\}$.
3. Let $I_0(l)$ denote the set of 0-inputs in R_l, let $I_1(l)$ denote the set of 1-inputs in R_l.
4. Construct μ_{l+1} be as follows:
   - the weight of all inputs in $I_1(l)$ is reduced by a factor of $1 - 1/n^4$. The obtained "free" weight is used to increase the weights of all inputs in $I_0(l)$ by a fixed factor.
   - Any input in $I_1(l)$ that is covered more than n^2 times receives weight 0. Its weight is used to increase the weight of all 1-inputs by a fixed factor.
5. STOP if all 1-inputs are covered at least n^2 times.

Note that each 0-input that is covered in some iteration l at least doubles its weight at that point. Namely, since a rectangle R_l has error $1/n^5$, weight $\mu_l(R_l)\cdot(1-1/n^5)\cdot 1/n^4$ is distributed to inputs in $I_0(l)$ of weight $\mu_l(R_l)\cdot 1/n^5$, more than doubling the weight of each such input. Since no 0-input has weight more than 1 or less than $1/2^{2n}$ this implies that no 0-input is ever covered more than $2n$ times.

We have to show that the distributions μ_l are all balanced for step 2. to work. The second part of step 4. does not change the balance of the distribution. In the first part of step 4. some weight is shifted from 1-inputs to 0-inputs. The inputs in $I_1(l)$ are reduced in weight by a factor of $(1 - 1/n^4)$. But as soon as an input is covered n^2 times this reductions stops.

Let us assume for the moment the following lemma justifying step 2., whose proof will be provided at the end of this section.

Lemma 4 $1/2 \ge \mu_l(f^{-1}(1)) \ge 1/2 - O(1/n)$ *for all* l.

The lemma clearly implies that the distributions μ_l are all balanced, hence step 2. is applied correctly. We use Lemma 4 only to ensure that the obtained cover is good on the 0-inputs, but not to ensure that it is good on the 1-inputs. To analyze the number of iterations of the algorithm we consider the following modification of step 2.

2.' First strictly balance μ_l by uniformly increasing the weights of 1-inputs and decreasing the weights of 0-inputs by fixed factors. Then pick a size s rectangle R_l with error $1/n^5$ according to that distribution.

The size of R_l on μ_l is at least $s \cdot d_l$, when $d_l = \mu_l(f^{-1}(1))/(1/2)$ denotes the distortion of the balance of μ_l compared to μ_1. Note that $\mu_l(f^{-1}(1)) \le 1/2$, so $d_l \le 1$.

We will show that the modified algorithm terminates and produces a size $O(n^3/s)$ cover. This immediately implies that each 1-input is covered at least n^2 times. Then this is also true for the original algorithm: The original algorithm uses larger rectangles in step 2. and hence terminates faster. Furthermore we will show that Lemma 4 holds for both the modified and the original algorithm. Note that, however, only the original algorithm guarantees that the cover is good on the 0-inputs.

Let S_l denote the set of 1-inputs not covered n^2 times before iteration l, and let $N_l = |S_l|$. Since the weight of each input in S_l is reduced in the first part of step 4. by a factor of $(1 - 1/n^4)$ each time it is covered, this decreases the weight of such an input by $(1 - 1/n^4)^{n^2-1} \ge (1 - O(1/n^2))$. In the second part of step 4. the weights of all 1-inputs are increased by some fixed factor. Furthermore $\mu_1(x, y) = 1/(2N_1)$ for all 1-inputs x, y. Hence for all $x, y \in S_l$ and $x', y' \in S_l$:

$$\begin{aligned} & (1 - O(1/n^2)) \cdot \mu_l(x', y') \\ \le\ & \mu_l(x, y) \le (1 + O(1/n^2)) \cdot \mu_l(x', y'). \end{aligned}$$

The average weight of an input in S_l is $(1/(2N_l))\cdot d_l$. Hence the for *all* $x, y \in S_l$

$$\begin{aligned} & (1/(2N_l))d_l(1 - O(1/n^2)) \\ \le\ & \mu_l(x, y) \le (1/(2N_l))d_l(1 + O(1/n^2)). \end{aligned} \tag{4}$$

Let $h_l(x,y)$ denote the number of times input x, y is covered by Cov_{l-1}. There are N_1 1-inputs to f. Then let

$$h(l) = \sum_{(x,y)\in f^{-1}(1)} \frac{1}{N_1} \cdot pos(n^2 - h_l(x,y)),$$

where $pos(x) = x$ if $x \geq 0$ and $pos(x) = 0$ otherwise. $h(l)$ denotes the average number of times 1-inputs still have to be covered. Clearly $n^2 \geq h(l) \geq 0$, and if $h(l) > 0$, then $h(l) \geq \frac{1}{N_1}$.

In each step 1-inputs of weight $s(1 - 1/n^5)d_l$ according to μ_l are covered. Let C_l denote the set of 1-inputs $x, y \in R_l$ with $h_l(x,y) < n^2$. Due to (4): $\mu_1(C_l) \geq sd_l(1-1/n^5)\cdot(1-O(1/n^2))\cdot(1/d_l)N_l/N_1 \geq s/2\cdot N_l/N_1$. Including R_l in the cover reduces $h(l)$ by at least sN_l/N_1 hence. Then

$$\begin{aligned}
& h(l+1) \\
\leq \;& h(l) - sN_l/N_1 \\
= \;& \sum_{(x,y)\in f^{-1}(1)} \frac{1}{N_1} \cdot pos(n^2 - h_l(x,y)) - \frac{N_l}{N_1}s \\
= \;& \sum_{(x,y)\in S_l} \frac{1}{N_1} \cdot (n^2 - h_l(x,y) - s) \\
\leq \;& \sum_{(x,y)\in S_l} \frac{1}{N_1} \cdot (n^2 - h_l(x,y)) \cdot (1 - s/n^2) \\
\leq \;& h(l) \cdot (1 - s/n^2).
\end{aligned}$$

For some $l = O(n^3/s)$ iterations $h(l) = 0$. Hence the constructed cover contains no more than $O(n^3/s)$ rectangles. Since the algorithm with the original step 2. terminates at least as fast we have $UT^{(1)}_{n,n^2}(f) \leq O(k \log n)$.

Given that $UT_{s,t}(f) = \max\{UT^{(1)}_{s,t}(f), UT^{(1)}_{s,t}(\neg f)\}$ we can simply do the same construction for the 0-inputs and get the desired result for $bound(f)$. $\square$

Proof of Lemma 4. First let us look at the algorithm with the modified step 2'. Let $s_l d_l$ denote the weight of R_l in μ_l. As argued before, the algorithm stops as soon as $\prod(1 - s_l/n^2) < 1/(n^2 \cdot N_1)$. So there is a sequence $s_1, \ldots, s_k$ so that $\prod_{l=1}^{k}(1 - s_l/n^2) < 1/(n^2 \cdot N_1)$ and all $s_l \geq s$, and k is minimal with this property. The weight transferred to the 0-inputs is then at most $\sum_{l=1}^{k-1} s_l d_l/n^4 \leq \sum_{l=1}^{k-1} s_l/n^4$, since no weight is transferred from R_k. We may hence adjust s_k so that $\prod_{l=1}^{k}(1-s_l/n^2) = 1/(n^2\cdot N_1)$.

For each k it is true that $\sum_{l=1}^{k} s_l$ is maximized for $s_1 = \cdots = s_k =: \bar{s}$ because: let $s'_l = (1 - s_l/n^2)$. Then $\prod s'_l = 1/(n^2 N_1)$ and we want to maximize $\sum(1 - s'_l n^2) = k - n^2 \sum s'_l$ or equivalently minimize $\sum s'_l$. This is achieved when $s_1 = \cdots = s_k$.

Then $k = O(n^3/\bar{s})$ and the transferred weight is at most $k\cdot\bar{s}/n^4 = O(1/n)$. Note that this implies $d_l \geq 1 - O(1/n)$.

In case the original step 2. of the algorithm is applied $\mu_l(R_l) = s_l \geq s$, and potentially a larger weight is transferred to the 0-inputs. But this also makes the algorithm terminate quicker.

The algorithm stops at least when $\prod(1 - s_l/(d_l n^2)) < 1/(n^2 \cdot N_1)$. We may substitute $s''_l = s_l/d_l$ and are left with the problem of finding the maximum of $\sum s''_l d_l$ under the constraint that $\prod(1 - s''_l/n^2) = 1/(n^2 \cdot N_1)$ and $s''_l = s_l/d_l \geq s/d_l \geq s$. This is the problem we have just analyzed. $\square$

11 Comparing the power of different threshold covers

First let us show that one-sided bounded error uniform threshold covers for some function f can easily be converted into approximate majority covers. The same also holds for the complement of f.

Proof of Theorem 4. Assume that $UT^{(1)}_{t,9t}(f) = k$. Then there exist 2^k rectangles so that each 1-input is in at least $9t$ rectangles and each 0-input is in at most t rectangles. Now label all the rectangles as 1-rectangles and add $3t$ times the 0-labelled rectangle covering all inputs. This is clearly an approximate majority cover, hence $APP(f) \leq UT^{(1)}_{t,9t}(f) \leq O(UT^{(1)}_{t,2t}(f))$. $\square$

Now we relate $APP(f)$ to a version of the rectangle size bound.

Proof of Theorem 5. Assume that $APP(f) = k$, then we can find 2^k labelled rectangles making up an approximate majority cover for f. We first have to show that in this case for each balanced distribution μ there exists a $3/4$-correct rectangle of size $1/2^{O(k)}$ at least. The proof is analogous to the proof of Theorem 3.1.a, but this time we are guaranteed to find a large rectangle with small error, not a large 1-rectangle with small error. To adapt the proof one has to replace the uniform threshold values $s', \epsilon s'$ by the expected correct height $(1-\epsilon)\cdot E[h(x,y)]$ and incorrect height $\epsilon \cdot E[h(x,y)]$.

Now we show the opposite direction, namely, given that for each balanced distribution μ we can find a $3/4$-correct rectangle of size $1/2^k$, then we can construct an approximate majority cover.

First notice that in fact we can find a rectangle of size $1/2^k$ and error at most $1/4$ for *all* distributions on the inputs, since on unbalanced distributions we may simply take $\{0,1\}^n \times \{0,1\}^n$ as a rectangle with error $1/4$ when choosing the appropriate label for that rectangle.

To construct the approximate majority cover we first consider the fact that $\min_\mu \max_v size(\mu, \epsilon, f, v) \geq 1/2^k$ in a somewhat different light. Let

$$para(\mu, R) = \epsilon/(3\mu(R)) + err(R, \mu, v(R)) \cdot 2^k.$$

This parameter controls the quality of a rectangle. Yao's application of the minimax-principle to randomized algorithms (see [KN97]) provides us with the following statement.

Lemma 5 *The following two statements are equivalent for all f.*

1. *For all distributions μ there is a rectangle R_μ with parameter α.*

2. *There is a probability distribution D on rectangles so that for all distributions μ on inputs the expected parameter of a rectangle is α.*

The latter could be named a "randomized rectangle" because it resembles a randomized algorithm. We know that for all distributions μ on $\{0,1\}^n \times \{0,1\}^n$ there is a rectangle R_μ with size $1/2^k$ and error $\epsilon = 1/4$, hence $para(\mu, R_\mu) \leq (1/4)/(3/2^k) + (1/4) \cdot 2^k = (1/3) \cdot 2^k$. The randomized rectangle then offers a distribution on rectangles with expected parameter $(1/3) \cdot 2^k$ at most. Hence the expected rectangle size satisfies $(1/4)/(3E[\mu(R)]) \leq (1/3) \cdot 2^k \Rightarrow E[\mu(R)] \geq 1/2^{k+2}$. The expected error satisfies $E[err(R, \mu, v(R))] \cdot 2^k \leq (1/3)2^k \Rightarrow E[err(R, \mu, v(R))] \leq 1/3$.

A randomized rectangle immediately gives us an approximate majority cover for f, though not of the desired size. To see this note that if we consider a distribution $\mu_{x,y}$ concentrated on some fixed input x, y, then $Prob_D(v(R) \neq f(x,y) | x, y \in R) = E[err(R, \mu_{x,y}, v(R))] \leq 1/3$.

From the direct application of the Yao-principle we do not get a bound on the number of rectangles with nonvanishing probabilities used in D. We use the following discretization.

Lemma 6 *Assume there is a randomized rectangle for f with expected size s and expected (constant) error $\epsilon \leq 1/3$. Then there is an approximate majority cover for f having size $O(s \cdot n)$.*

The lemma clearly implies $APP(f) \leq O(k) + \log n$. So let us prove the lemma. We independently pick $t = c \cdot (1/s) \cdot n$ rectangles $R_1, \ldots, R_t$ from the distribution D for some large enough constant c. Our claim is that this yields the desired approximate majority cover. Let $w(x,y)$ denote the random variable counting the number of R_i with label $v(R) \neq f(x,y)$ and $x, y \in R_i$. Let $h(x,y)$ denote the number of all R_i with $x, y \in R_i$.

Consider the distribution $\mu_{x,y}$ concentrated on x, y. We know that the expected size of a rectangle picked from D is at least s. Since $\mu_{x,y}(R) \in \{0,1\}$, with probability at least s a chosen rectangle contains x, y. So $E[h(x,y)] \geq s \cdot c \cdot (1/s) \cdot n = cn$.

We know $E[w(x,y)] \leq \epsilon E[h(x,y)]$, and want to bound $Prob(w(x,y) \geq 1.1 \cdot \epsilon E[h(x,y)])$, which is maximized if $E[w(x,y)]$ is as large as possible, hence we assume $E[w(x,y)] = \epsilon E[h(x,y]$. Using the Chernov bound

$$\begin{aligned} & Prob(w(x,y) \geq 1.1 \cdot \epsilon E[h(x,y)]) \\ = \; & Prob(w(x,y) \geq 1.1 \cdot E[w(x,y)]) \\ \leq \; & e^{-E[w(x,y)]/300} = e^{-\epsilon E[h(x,y)]/300} \\ \leq \; & e^{-\epsilon cn/300}. \end{aligned}$$

Let $c = O(1/\epsilon)$ be large enough, so that the above probability is at most 2^{-2n-1}. Then the probability that there exists one of the 2^{2n} inputs x, y with $w(x,y) \geq 1.1 \cdot \epsilon h(x,y)$ is smaller than 1. Consequently there exists a choice of t rectangles so that for all x, y: $w(x,y) \leq 1.1 \cdot \epsilon h(x,y)$.

By Remark 2 we get an approximate majority cover. $\square$

We now show an exponential gap between (even one-sided) bounded error uniform threshold covers and approximate majority covers.

Proof of Theorem 6. Consider the function $BOTH$: $(\{0,1\}^{2n} \times \{0,1\}) \times (\{0,1\}^{2n}) \to \{0,1\}$ defined as follows:

$$\begin{aligned} & BOTH((x_1, x_2, a), (y_1, y_2)) \\ = \; & (DISJ(x_1, y_1) \wedge a) \vee (\neg DISJ(x_2, y_2) \wedge \neg a). \end{aligned}$$

Hence depending on a the function either computes $DISJ$ on the first pair of inputs or $\neg DISJ$ on the second pair.

First we show that $APP(BOTH) = O(\log n)$. Note $APP(\neg DISJ) = O(\log n)$, since $N(\neg DISJ) = O(\log n)$ and $APP(f) \leq N(f)$. Hence also $APP(DISJ) = O(\log n)$, since $APP(f) = APP(\neg f)$ for all f. To find an approximate majority cover for $BOTH$ we take the approximate majority cover for $DISJ$ and intersect all its rectangles with the rectangle defined by $a = 1$. We also take the approximate majority cover for $\neg DISJ$ and intersect all its rectangles with the rectangle defined by $a = 0$. The union of these sets of rectangles is an approximate majority cover for $BOTH$. So $APP(BOTH) = O(\log n)$.

Now we consider $UT^{(1)}_{s,2s}(BOTH)$. We consider a distribution on inputs in which $a = 1$ with probability 1. In this case with probability 1, $BOTH((x_1, x_2, a), (y_1, y_2)) = DISJ(x_1, y_1)$. Since there is a balanced distribution on inputs so that each 1-rectangle either has size $1/2^{\Omega(n)}$ or error at least ϵ for some constant $\epsilon > 0$ (see Fact 1), we can choose this distribution on the input positions x_1, x_2, and fix x_2, y_2 arbitrarily. In this way we get a balanced distribution with the same properties for $BOTH$ and hence $UT^{(1)}_{s,2s}(BOTH) = \Omega(n)$ (using Theorem 3).

Now we consider $UT^{(1)}_{s,2s}(\neg BOTH)$. We may proceed as above, by fixing $a = 0$ and considering the quality of 1-

rectangles for $\neg(\neg DISJ)$. So we get $UT^{(1)}_{s,2s}(\neg BOTH) = \Omega(n)$. □

Proof of Theorem 7. It is easy to construct a majority cover for MAJ. The cover contains n 1-rectangles defined by $x_i \wedge y_i$ plus $\lceil n/2 \rceil$ 0-rectangles covering $\{0,1\}^n \times \{0,1\}^n$. If we have $MAJ(x,y) = 1$, then at least $\lceil n/2 \rceil$ 1-rectangles $x_i \wedge y_i$ contain x, y, else at most $\lceil n/2 \rceil - 1$ 1-rectangles contain x, y.

For the lower bound we have to argue that there is a balanced distribution for which all $1-\epsilon$-correct rectangles have size at most $1/2^{\Omega(n)}$. Let $n' = 6k+2$ be the input length for some k satisfying $k \equiv 1 \bmod 2$ and $k \equiv 1 \bmod 3$.

First we fix $2k$ variables $x_i, y_i = 1$. There are $n = 4k+2$ remaining variables. $MAJ(x,y) = 1 \iff \sum_{i=1}^n x_i \wedge y_i \geq k+1$ under this fixing. We pretend in the following that there are n variables.

Let us define the distribution. Let μ_r be the uniform distribution on $\{(x,y) : |x| = |y| = n/2, |x \cap y| = k\}$, and let μ_a be the uniform distribution on $\{(x,y) : |x| = |y| = n/2, |x \cap y| = k+1\}$. Then let μ be defined by $\mu(x,y) = (3/4) \cdot \mu_r(x,y) + (1/4) \cdot \mu_a(x,y)$. The distribution is obviously balanced.

We have to show that there are no large rectangles with small error, neither 1-rectangles nor 0-rectangles. This is handled in the following way.

Claim 1 *If there is a 1-rectangle of size $\Omega(s)$ and error $O(\delta)$ according to μ and MAJ, then there is a 0-rectangle of size $\Omega(s)$ and error $O(\delta)$ according to μ and MAJ, and vice versa.*

Proof of the claim. Assume there is a size s 1-rectangle $R = A \times B$ with $(1-\delta)s$ 1-inputs and δs 0-inputs on μ. Let $\neg A = \{x : \overline{x} \in A\}$. We claim that $\neg A \times B$ is an $O(\delta)$-error size $\Omega(s)$ 0-rectangle. Note that $f(x,y) = 0 \iff |x \cap y| = k$ and $f(x,y) = 1 \iff |x \cap y| = k+1$ under μ, and that $|x|, |y| = 2k+1$. Hence $|\overline{x} \cap y| = |y| - |x \cap y| = 2k+1-k-f(x,y)$, and so $f(x,y) \neq f(\overline{x},y)$ with probability 1. So the rectangle $\neg A \times B$ has entries with reversed function value compared to $A \times B$. The claim follows with $\mu(x,y) = \Theta(\mu(\overline{x},y))$. □

We are going to show that each 0-rectangle has size at most $2^{-\Omega(n)}$ or has error ϵ for some constant ϵ. Then $APP(MAJ) = \Omega(n)$.

We consider the following way to choose inputs according to μ: First we choose a *frame*, namely a partition of $\{1,\ldots,n\}$ into sets z_k of size k, z_x, z_y of size $(4k+2-k-1)/2 = \lceil 3k/2 \rceil$, and $\{i\}$ of size 1, uniformly under all such partitions. Then x is chosen to contain all of z_k and with probability 1/2 also $\{i\}$. x is filled up to a size $2k+1$ set by choosing uniformly elements of z_x. y is chosen similarly, only with the filling up done from the set z_y. Note that this produces the distribution μ.

Now we fix z_k arbitrarily. Let μ_{z_k} denote the corresponding distribution on inputs. Ignoring the variables in z_k the players choose sets with an intersection size in $\{0,1\}$, i.e., they solve the (complement of the) disjointness problem on a specific distribution.

We employ Fact 6 at this point. A technical problem for the application of this fact is that for μ_{z_k} subsets of size $k+1$ are chosen from a size $n-k = 3k+2$ universe. To overcome this we may fix arbitrary disjoint subsets $s_x, s_y \subseteq \{1,\ldots,n\} - z_k$ of size $l = k/3 + 2/3$ each. The variables in s_x are set to 1 in x and the variables in s_y are set to 1 in y. After fixing z_k, s_x, s_y an input is chosen as follows. First z_x and z_y are chosen, under the condition that they include s_x resp. s_y, hence the remaining size of these is $\lceil 3k/2 \rceil - \lceil k/3 \rceil$ each. Then $\{i\}$ is chosen and the frame is complete. Afterwards an input is chosen as before. Call the resulting distribution μ_{z_k,s_x,s_y}.

The number of remaining nonfixed variables when choosing according to μ_{z_k,s_x,s_y} is $n'' = 3k+2-l$. Disregarding the fixed l elements the size of x and of y is $k+1-l = (2/3)k + 1/3 = n''/4$. So disregarding the fixed inputs we have reached the distribution ν of Fact 6.

Under μ the weight of any input x, y can be expressed as the expectation over all possibilities to fix z_k and to fix s_x, s_y of the weight of the input under this fixing. Namely,

$$\mu(x,y) = E_{z_k,s_x,s_y}[\mu_{z_k,s_x,s_y}(x,y)].$$

We know from Fact 6 that for all z_k, s_x, s_y:

$$\mu_{z_k,s_x,s_y}(MAJ^{-1}(1) \cap R) \geq 1/5 \cdot \mu_{z_k,s_x,s_y}(R) - 2^{-\Omega(n)}.$$

Hence also

$$\begin{aligned} & \mu(R \cap MAJ^{-1}(1)) \\ = \; & E_{z_k,s_x,s_y}[\mu_{z_k,s_x,s_y}(MAJ^{-1}(1) \cap R)] \\ \geq \; & (1/5) \cdot E_{z_k,s_x,s_y}[\mu_{z_k,s_x,s_y}(R)] - 2^{-\Omega(n)} \\ = \; & (1/5) \cdot \mu(R) - 2^{-\Omega(n)}. \end{aligned}$$

Hence any 0-rectangle for MAJ under μ either has size $2^{-\Omega(n)}$, or error $1/5$. Due to our previous claim within constant factors the same holds for 1-rectangles. So the lower bound $APP(MAJ) = \Omega(n)$ follows. □

References

[AUY83] A.V. Aho, J.D. Ullman, M. Yannakakis: On Notions of Information Transfer in VLSI Circuits. *15th ACM Symp. on Theory of Computing*, pp. 133–139, 1983.

[B85] L. Babai. Trading Group Theory for Randomness. *17th ACM Symp. on Theory of Computing*, pp. 421–429, 1985.

[BFS86] L. Babai, P. Frankl, J. Simon. Complexity classes in communication complexity theory. *27th IEEE Symp. Found. of Comp. Science*, pp. 303–312, 1986.

[BM88] L. Babai, S. Moran. Arthur-Merlin Games: A Randomized Proof System, and a Hierarchy of Complexity Classes. *Journal of Computer and System Sciences*, vol. 36, pp. 254–276, 1988.

[BJKS02] Z. Bar-Yossef, T.S. Jayram, R. Kumar, D. Sivakumar. An Information Statistics Approach to Data Stream and Communication Complexity. *43rd IEEE Symp. Found. of Comp. Science*, pp. 209–218, 2002.

[HR90] B. Halstenberg, R. Reischuk. Relations between communication complexity classes. *Journal of Computer and System Sciences*, vol. 41, pp. 402–429, 1990.

[HHT97] Y. Han, L.A. Hemaspaandra, T. Thierauf. Threshold Computation and Cryptographic Security. *SIAM Journal on Computing,* vol.26, pp. 59–78, 1997.

[HS96] J. Hromkovič, G. Schnitger. Nondeterministic Communication with a Limited Number of Advice Bits. *28th ACM Symp. on Theory of Computing*, pp. 551–560, 1996.

[KS92] B. Kalyanasundaram, G. Schnitger. The Probabilistic Communication Complexity of Set Intersection, *SIAM Journal on Discrete Mathematics,* vol.5, pp. 545–557, 1992.

[KNSW94] M. Karchmer, I. Newman, M. Saks, A. Wigderson. Non-deterministic Communication Complexity with Few Witnesses. *Journal of Computer and System Sciences*, vol. 49, pp. 247–275, 1994.

[Kl98] H. Klauck. Lower bounds for computation with limited nondeterminism. *13th IEEE Conference on Computational Complexity,* pp. 141–153, 1998.

[Kl01] H. Klauck. Lower Bounds for Quantum Communication Complexity. *42nd IEEE Symp. Foundations of Computer Science*, pp. 288–297, 2001.

[Kl02] H. Klauck. On Quantum and Approximate Privacy. *19th Symposium on Theoretical Aspects of Computer Science*, pp. 335–346, 2002.

[KN97] E. Kushilevitz, N. Nisan. Communication Complexity. Cambridge University Press, 1997.

[R99] R. Raz. Exponential Separation of Quantum and Classical Communication Complexity. *31st ACM Symp. on Theory of Computing*, pp. 358–367, 1999.

[RW92] R. Raz, A. Wigderson. Monotone Circuits for Matching Require Linear Depth. *Journal of the ACM*, vol.39, pp. 736–744, 1992.

[R92] A. Razborov. On the Distributional Complexity of Disjointness. *Theoretical Computer Science*, vol. 106, pp. 385–390, 1992.

[T79] C.D. Thompson. Area-Time Complexity for VLSI. *11th ACM Symp. on Theory of Computing*, pp. 81–88, 1979.

[W00] I. Wegener. Branching Programs and Binary Decision Diagrams. SIAM monographs, 2000.

[Y79] A.C.C. Yao. Some Complexity Questions Related to Distributed Computing, *11th ACM Symp. on Theory of Computing*, pp. 209–213, 1979.

[Y83] A.C.C. Yao. Lower Bounds by Probabilistic Arguments. *24th IEEE Symp. Foundations of Computer Science*, pp. 420–428, 1983.