Characterizing Non-Deterministic Circuit Size
(*Four variations on one theme*)

Mauricio Karchmer[*]
Avi Wigderson[†]

Abstract

Consider the following simple communication problem. Fix a universe U and a family Ω of subsets of U. Players I and II receive, respectively, an element $a \in U$ and a subset $A \in \Omega$. Their task is to find a subset B of U such that $|A \cap B|$ is even and $a \in B$. With every Boolean function f we associate a collection Ω_f of subsets of $U = f^{-1}(0)$, and prove that its (one round) communication complexity completely determines the *size* of the smallest nondeterministic circuit for f.

We propose a linear algebraic variant to the general approximation method of Razborov, which has exponentially smaller description. We use it to derive four different combinatorial problems (like the one above) that characterize NP. These are tight, in the sense that they can be used to prove super-linear circuit size lower bounds. Combined with Razborov's method, they present a purely combinatorial framework in which to study the P vs. NP vs. $co-NP$ question.

[*] Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139. Partially supported by NSF grant CCR-9212184.

[†] Department of Computer Science, Hebrew University, Jerusalem, Israel 91904.

Introduction

Implicit in the concept of computation, both parallel and sequential, hides a notion of evolution, of time. Each operation processes the outcome of previous computations. This idea of computation suggests two approaches to proving lower bounds: a *Bottom-up* approach in which progress is measured as the computation progresses from the inputs towards the output, and a *Top-down* approach in which progress is measured as the computation reverses itself from the output down to the inputs. Both approaches require, in one form or another, measuring progress as it is made node by node.

Both approaches are present in almost every lower bound proof known to date, with some exceptions. The approximation method (see Andreev [3] , Razborov [14, 15] or Alon & Boppana [2]) can be regarded as a Bottom-up approach. Similarly, proofs which use random restrictions are Bottom-up (see Ajtai [1], Furst, et.al. [5], Yao [20] or Hastad [7]). Depth lower bounds which use communication complexity are Top-down (see Karchmer & Wigderson [10] or Goldmann & Hastad [6]). Two exceptions, both of which use communication complexity though in a more "global" manner, are the lower bounds of Raz & Wigderson [13] and of Razborov [16] for monotone circuit depth .

It is our opinion (as well as that of others) that neither of these approaches works for proving lower bounds for general circuits. In fact, it may be that our failure in proving non-trivial general lower bounds stems from the fact that it is hard to depart from these very intuitive approaches. Alas, one crucial matter remains: What substitute?

One idea that is starting to emerge is the following: Locally, computations for zeroes and ones of a hard function f look alike. Hence, to show that a "small" circuit does not compute f correctly one can *combine* rejecting computations for the zeroes of f to get a rejecting computation for a one of f.

This idea is implicit in Sipser [19] where a new proof is given for the fact that analytic sets are not closed under complement. It is also implicit in Razborov [17] where a generalized version of the approximation method is given which, in principle, provides tight lower bounds for circuit size. Karchmer [9] elicits the idea of Razborov [17] and, in particular, the idea of combining zeroes is made explicit. In addition, Karchmer [9] presents Razborov's monotone lower bound for clique within this framework. Again, Razborov [18] implicitly uses this idea to give a super-linear lower bound for the complexity of majority on non-deterministic branching programs.

All of these papers look at a circuit as a collection of gates stripped away from any explicit structure connecting them to one another. In particular, they don't study the circuit in either a Bottom-up nor a Top-down fashion.

In this paper we explore the possibility of combining computations using addition modulo 2 (this may be considered a concrete implementation of Sipser's notion of a "finite limit"). The main ideas is as follows: Assume that C is a small circuit which allegedly computes a hard function f and such that it rejects all zeroes of f. Pick a subset S of the zeroes of f. To each node g of C, assign the parity of the values that the vectors in S assign to g. Show that for some subset S, suitably chosen, the assigned values define the correct rejecting computation of a one of f.

As it turns out, this approach can be used not only to prove lower bounds for circuit size, but also to lower bound non-deterministic circuit size. In fact, this approach can be used to *characterize* non-deterministic circuit size.

We present 4 variations on this theme and get 4 characterizations of NP. The first characterization uses a covering problem in much the same way as Razborov's generalization of the approximation method [17]. The second characterization uses the notion of hypergraph embeddings and is reminiscent of a similar characterization of NC^1 in terms of graph embeddings (see Razborov [16] and Pudlák & Rödl [12]). The third characterization uses the communication complexity of relations in the spirit of Karchmer & Wigderson [10] where a characterization of circuit depth is given. Finally, the fourth characterization provides an algebraic representation of NP languages using matrices over $GF(2)$.

It is important to stress that all four characterizations are tight. Hence, we can use any of them to get super-linear lower bounds for non-deterministic circuit size. For example, if one uses the third characterization in terms of Communication Complexity (mentioned in the abstract), then it is enough to prove a Communication Complexity lower bound of the form $\log n + \omega(1)$ to get a super-linear size lower bound for the function in question.

Our first characterization of NP in terms of a covering problem closely follows Razborov's generalized approximation method. By way of comparison a few words are in order. The generalized approximation method entails the introduction of a system consisting of a set $\mathcal{M}$ of objects (*Monotone Functionals* over $f^{-1}(0)$) and a family of subsets of $\mathcal{M}$ such that the circuit complexity of f is polynomialy related to the minimum cover number of the system. We proceed in a similar manner with the set $\mathcal{L}$ of *Linear Functionals* over $f^{-1}(0)$ as our basic objects. One important difference is the fact that the cardinality of $\mathcal{L}$ is exponentially smaller than the cardinality of $\mathcal{M}$ making our system much simpler. A second difference and perhaps more important is the algebraic nature of our system which suggests the use of algebraic techniques to study its cover number.

It is interesting to note here that although the approach of Razborov which is based on monotone functionals captures *deterministic* circuit size, our approach captures *non-deterministic* circuit size. In fact, the authors observed (see Karchmer [9]) that an approach based on *self-dual monotone functionals* [1] captures non-deterministic circuit size as well. We do not have a clear understanding of why some classes of functionals capture deterministic size while others capture non-deterministic size. In fact, one could study the P versus NP question in this light. Not only do these approaches provide us with ways of bounding deterministic and non-deterministic circuit size, but they provide us with a framework to compare deterministic and non-deterministic computation.

Our basic idea of combining rejecting computations using addition modulo 2 can be carried out in other models of computation as well. In particular, we use this idea in [11] to study Branching Programs which count the number of accepting paths modulo 2. We use an analog of the first characterization to prove that any such program for *Majority* requires super-linear size. This is similar to a result of Razborov [18] who uses Monotone functionals to prove that non-deterministic Branching Programs for *Majority* require super-linear size.

We choose to characterize $co-NP$ instead of NP. We do so to be consistent with other authors. Of course, all of our characterizations have their respective duals and, hence, can be used to characterize NP as well.

[1] A monotone function f is *self-dual* if for every vector x, $f(\neg x_1, \cdots, \neg x_n) = \neg f(x_1, \cdots, x_n)$.

1 Background

We refer the reader to the excellent survey of Circuit Complexity written by Boppana & Sipser [4]. Here, we will concentrate on the definitions that we use.

Let B_n denote the set of all boolean functions $f : \{0,1\}^n \mapsto \{0,1\}$. A *Boolean Circuit* over $\{\wedge, \oplus, 1\}$ is a directed acyclic graph with $n+1$ sources, one sink and all nodes with indegree 2. Sources are labeled by literals from $\{x_1, \cdots, x_n\} \cup \{1\}$ and nodes are labeled by one of the two boolean operations (gates) $\{\oplus, \wedge\}$. A boolean circuit computes a function in B_n in a natural way. In the same way, any inner node of C computes a function in B_n. In what follows, an $\wedge$-gate will be viewed as the unordered pair of functions from B_n computed at the nodes feeding the gate.

The size of a circuit C, denoted $s(C)$, is defined as the number of gates of C. Similarly, we write $s_\wedge(C)$ for the number of $\wedge$-gates of C. For a function $f \in B_n$ let $s(f)$ and $s_\wedge(f)$ denote the size and number of $\wedge$-gates of an optimal circuit that computes f.

A *non-deterministic* circuit with m non-deterministic variables is a circuit with $n+m+1$ sources labeled by $\{x_1, \cdots, x_n\} \cup \{y_1, \cdots y_m\} \cup \{1\}$. A non-deterministic circuit computes a function $f \in B_n$ as follows: For $x \in \{0,1\}^n$, $f(x) = 1$ iff there exist a setting of the non-deterministic variables $\{y_1, \cdots y_m\}$ which makes the circuit output 1. For a function f, denote by $ns(f)$ and $ns_\wedge(f)$ the size and number of $\wedge$-gates of an optimal non-deterministic circuit for f.

A *co-non-deterministic* circuit is a non-deterministic circuit with the following accepting criteria: For $x \in \{0,1\}^n$, $f(x) = 0$ iff there exist a setting of the non-deterministic variables $\{y_1, \cdots y_m\}$ which makes the circuit output 0. For a function f, denote by $\bar{n}s(f)$ and $\bar{n}s_\wedge(f)$ the size and number of $\wedge$-gates of an optimal co-non-deterministic circuit for f.

The reason for studying Boolean Circuits is well known. If one could prove that a function in NP requires super-polynomial size then one would get that $P \neq NP$. Similarly, if one could prove that a function in NP requires super-polynomial co-non-deterministic circuit size then one would conclude that $NP \neq co-NP$. It is also well known that, although most functions require exponential non-deterministic circuit size, there is no function in NP known to require super-linear circuit size.

In the sequel, we will present a method to bound $\bar{n}s_\wedge(f)$. The proof of the following lemma goes along the lines of a similar lemma proved by Alon & Boppana [2]. It says that $\bar{n}s_\wedge(f)$ cannot be much smaller than $\bar{n}s(f)$.

Lemma 1 *For any function f, $\bar{n}s(f) = O(\bar{n}s_\wedge(f)^2)$.*

2 A covering problem

Fix $f \in B_n$ and let $U = f^{-1}(0)$ and let $U = \{u_1, \cdots, u_m\}$ be an arbitrary ordering of U. In what follows, a subset $A \subseteq U$ would be associated with its characteristic vector in $GF(2)^m$. Furthermore, we will abuse notation and write A for both the subset and its characteristic vector. Given two distinct vectors $A, B \in GF(2)^m$ we will denote by $H(A,B)$ the set of all vectors S such that $< S \cdot A > = < S \cdot B > = 0$. Obviously, $H(A,B)$ is a subspace of co-dimension 2. We say that a subspace V of $GF(2)^m$ is *spanning* if for every $i \in \{1, ..., m\}$ there exists a vector v in the dual of V such that $v_i = 1$. Equivalently, a subspace is spanning if it does not contain any of the vectors

e_i for $i \in \{1, ..., m\}$. For example, the subspace $H(A, B)$ is spanning iff $A \cup B = U$.

A *valuation* of the nodes of a circuit C is a Boolean mapping whose domain is the set of all sub-functions of C, including the input variables and the constant 1. In particular, the computation of C on a given vector is a valuation.

We are now ready to present the main idea. Fix a co-non-deterministic circuit C such that $s_{\wedge}(C) < s_{\wedge}(f)$. How can we show that C does not compute f correctly? One idea is to assume that C rejects U and show that it also rejects a vector not in U. We will do so by combining rejecting computations for vectors in U to form a rejecting computation for a vector not in U.

Fix one rejecting witness w^u for each $u \in U$ (remember that C rejects U). Having done this, we can associate with any node in C computing a function g the set $[|g|] \rightleftharpoons \{u \in U \mid g(u, w^u) = 1\}$. A subset $S \subseteq U$ defines a valuation of the nodes of C by $\phi_S(g) = < S \cdot [|g|] >$. In other words, we can combine the computations of the vectors in S to form a valuation of C as follows: For a node g of C, assign to g the parity of the values that vectors in S assign to g.

Note that for any $S \subseteq U$, $\phi_S(C) = 0$ so that any such valuation is "rejecting". Also, if $|S| \equiv_2 1$ then $\phi_S(1) = 1$ so that the constant 1 is assigned the correct value. Finally, it is easy to see that for any g, h in C we have $\phi_S(g \oplus h) = \phi_S(g) \oplus \phi_S(h)$. Therefore, if $|S| \equiv_2 1$ then the only source of mistakes will be the $\wedge$-gates of C.

¿From now on we will work only with sets S of odd cardinality. Any such set defines a vector $z \in \{0, 1\}^n$ by $z_i = \phi_S(x_i)$. We will denote the defined vector by $\oplus S$. Alternatively, one can define $\oplus S$ as follows: Consider the $n \times m$ matrix M whose columns are all the vectors from U. It is easy to verify that $\oplus S = MS$.

Assume now that $\oplus S \notin U$. Then, if C computes f then there has to be an $\wedge$-gate (g, h) in C where ϕ_S "makes a mistake", that is, where $\phi_S(g \wedge h) \neq \phi_S(g) \wedge \phi_S(h)$. We say that the valuation ϕ_S *preserves* the $\wedge$-gate (g, h) if $\phi_S(g \wedge h) = \phi_S(g) \wedge \phi_S(h)$. Also, we say that the valuation ϕ_S makes a mistake on an $\wedge$-gate if it does not preserve it.

Proposition 1 The valuation ϕ_S makes a mistake on the $\wedge$-gate (g, h) iff the set S intersects exactly 3 of the sets $[|g \wedge h|], [|g \wedge \neg h|], [|\neg g \wedge h|], [|\neg g \wedge \neg h|]$ in an odd number of places.

Proof: By inspection. ■

Conversely, the valuation ϕ_S preserves the $\wedge$-gate if and only if S intersects exactly one of the sets in an odd number of places. The following proposition further simplifies the conditions:

Proposition 2 Let U be partitioned into the 4 sets A_i for $i = 1, ..., 4$. Let $B_1 = A_1 \cup A_2$ and $B_2 = A_2 \cup A_3 \cup A_4$. Also, let $C_1 = A_3 \cup A_4$ and $C_2 = A_1 \cup A_2 \cup A_3$. Let $S \subseteq U$ be of odd cardinality. Then S intersects exactly 3 of the $A's$ in an odd number of places iff $S \in H(B_1, B_2) \cup H(C_1, C_2)$. Note that both $H(B_1, B_2)$ and $H(C_1, C_2)$ are spanning.

Therefore, any $\wedge$-gate of C defines two spanning subspaces of co-dimension 2. Any subset S of odd cardinality defines a valuation which preserves the $\wedge$-gate iff S is not in any of the two subspaces. This motivates the following covering problem:

Definition 1 *For* $f \in B_n$, *let* $\Omega_f \rightleftharpoons \{S \subseteq f^{-1}(0) \mid |S| \equiv_2 1$ *and* $\oplus S \in f^{-1}(1)\}$.

Definition 2 *For a subset $\Omega \subseteq GF(2)^m$ let $\rho(\Omega)$ be the minimum number of spanning subspaces of $GF(2)^m$ of co-dimension 2 whose union contains Ω.*

Theorem 1 For any $f \in B_n$, $\rho(\Omega_f) \leq 2\bar{n}s_\wedge(f)$.

Proof: Let C be a co-non-deterministic circuit which allegedly computes f using $s = s_\wedge(f)$ $\wedge$-gates and assume that $s < \rho(\Omega_f)/2$. Each $\wedge$-gate of C defines 2 spanning subspaces according to propositions 1 and 2. In total, we get strictly fewer than $\rho(\Omega_f)$ many spanning subspaces. By the definition of $\rho(\Omega_f)$, there is a subset $S \subseteq f^{-1}(0)$ of odd cardinality which defines a vector $\oplus S \in f^{-1}(1)$ and which is not covered by any of the subspaces. Therefore, by propositions 1 and 2, the valuation ϕ_S preserves every $\wedge$-gate of C. This means that ϕ_S defines the correct rejecting computation of the vector $\oplus S$ and, thus, C does not compute f correctly. ■

Therefore, to prove lower bounds for the size of non-deterministic circuits it is enough to prove lower bounds for the more combinatorial quantity $\rho(\Omega_f)$. Our next theorem provides a converse to theorem 1.

Theorem 2 For any $f \in B_n$, $\bar{n}s_\wedge(f) = O(\rho(\Omega_f) + n)$.

Proof: Let $t = \rho(\Omega_f)$ and let $\{H(A^i, B^i)\}$ for $i = 1, ..., t$ be a cover of Ω_f by spanning subspaces of co-dimension 2.

Claim 1 *For any $z \in \{0,1\}^n$, $f(z) = 0$ iff there exists a subset $S \subseteq U$ of odd cardinality such that $\oplus S = z$ and for every $i \leq t$, $S \notin H(A^i, B^i)$.*

Proof:[of claim] *i)* If $f(z) = 0$ then the subset $\{z\}$ satisfies the requirements of the claim.
ii) If $f(z) = 1$ then any subset $S \subseteq U$ of odd cardinality and such that $\oplus S = z$ is in Ω_f and therefore it is covered by one of the subspaces. ■

Therefore, to show that $f(z) = 0$ we have to 'guess' a good S. Clearly, we do not have enough time to guess such a long vector. Instead, we will guess a $(2t + n + 1)$-dimensional vector Y whose entries are $Y_{x_i} = < S \cdot [|x_i|] >$ for $i = 1, ..., n$, $Y_U = < S \cdot U >$ and $Y_{A^i} = < S \cdot A^i >$ and $Y_{B^i} = < S \cdot B^i >$ for $i = 1, ..., t$. Given these guesses, the conditions of the claim can be quickly verified. Hence, it is enough to verify that the vector Y is consistent in that its entries are the right values for some S. In fact, the vector Y will be guessed in a way that guarantees that it is consistent.

Let N be the $(2t + n + 1) \times m$ matrix whose rows correspond to the characteristic vectors of the sets $[|x_i|]$ for $i = 1, ..., n$, U and A_i, B_i for $i = 1, ..., t$. Clearly, $NS = Y$. Therefore, to guess a consistent Y it is enough to guess a vector in the column space of the matrix N. Let N' consist of a maximal set of linearly independent columns of N. Clearly, N' has at most $2t + n + 1$ columns. To guess a consistent vector Y it is enough to guess a linear combination of the columns of N'.

A co-nondeterministic circuit C for f can be built as follows. Hard-wire the columns of N' into the circuit C. Note that the definition of N' did not depend on the vector z. Use $2t + n + 1$ bits to guess a consistent Y and check, using $O(t + n)$ $\wedge$-gates, that the conditions of the claim are satisfied. ■

Note that we have used non-uniformity in several places in the construction of C. First, the covering of Ω_f may be non-uniform. Second, it is not clear how to construct the maximal set of linearly independent columns of N efficiently. As far as we know, this is the only characterization of $co - NP$ where non-uniformity is exploited.

Corollary 1 For most functions $f \in B_n$, $\rho(\Omega_f) \geq 2^{n/2}$.

CHARACTERIZATION 1. A function f is in $co - NP$ iff there exists a polynomial $p(n)$ such that for every n, $\rho(\Omega_f) \leq p(n)$.

REMARK. When proving lower bounds to $\rho(\Omega_f)$, it may be useful to consider a sub-family $\Omega \subseteq \Omega_f$ and show that $\rho(\Omega)$ is large. However, a random spanning subspace of co-dimension 2 covers a fourth of the elements of any family $\Omega \subseteq GF(2)^m$. Therefore, an easy calculation shows that $\rho(\Omega) = O(\log |\Omega|)$. This means that to prove super-polynomial lower bounds for $\bar{n}s_\wedge(f)$ one needs to work with sub-families with more than $2^{n^{\omega(1)}}$ members.

REMARK. The proof of theorem 2 implies that in order to prove that $\rho(\Omega_f) \geq s$, it is enough to work with subsets $S \subseteq U$ of cardinality $2s+n+1$. For example, to prove super-linear lower bounds it is enough to work with subsets of cardinality $\Theta(n^2)$.

We can generalize our covering problem to allow for spanning subspaces of arbitrary dimension.

Definition 3 *For a subset $\Omega \subseteq GF(2)^m$ let $\rho_\infty(\Omega)$ be the minimum number of spanning subspaces of $GF(2)^m$ whose union contains Ω.*

Clearly, for every Ω we have $\rho_\infty(\Omega) \leq \rho(\Omega)$. Hence, for any f, $\bar{n}s_\wedge(f) \geq \rho_\infty(\Omega_f)$. Note however that, in principle, ρ_∞ can be much smaller than ρ.

We finish this section by defining a *universal* collection which can be used to "try out" lower bound arguments. Let Ω^m consists off all odd vectors in $GF(2)^m$ other than the vectors $e_1, \cdots, e_m$. The following theorem provides tight bounds to the complexity of Ω^m.

Theorem 3 $\rho(\Omega^m) = \Theta(m)$.

Proof: The upper bound follows by a probabilistic argument. The lower bound follows by noticing that any polynomial over $GF(2)$ which takes value 1 in Ω^m and 0 in $e_1, \cdots, e_m$ has degree at least $m-2$ while the union of t spanning subspaces of co-dimension 2 can be represented by a polynomial of degree $2t$. ■

As stated before, $\rho_\infty(\Omega_f)$ may be much smaller than $\rho(\Omega_f)$. We finish this section by suggesting the following problem: Show that there exists some function $f \in B_n$ with $\rho_\infty(\Omega_f) = n^{\omega(1)}$. It might be interesting to prove even that $\rho_\infty(\Omega^m) = \omega(\log m)$.

3 A universal hypergraph

Our second variation involves hypergraph embeddings. This characterization is reminiscent of a characterization of NC^1 in terms of graph embeddings [16, 12].

Definition 4 *Given two hypergraphs $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$. An* embedding *of H_1 in H_2 is a mapping $\varphi : V_1 \mapsto V_2$ such that if $\{v_1, \cdots, v_k\} \in E_1$ then $\{\varphi(v_1), \cdots, \varphi(v_k)\} \in E_2$.*

In what follows we are going to view $GF(4)^t$ as the set $\{0, 1, x, 1+x\}^t$. We are also going to identify the sets $\{0,1\}^n$ and $GF(2)^n$. We are going to write $\sum$ for addition in $GF(4)^t$ and $\oplus$ for addition in $GF(2)^n$.

Definition 5 **A Universal Hypergraph.** *Let $\mathcal{H}$ be the family of Hypergraphs $\{H_t = (V_t, E_t) \mid t \in \mathcal{N}\}$, where $V_t = \{1, x, 1+x\}^t \subseteq GF(4)^t$ and $\{v_1, \cdots, v_k\} \in E_t$ iff k is odd and $\sum_i v_i \notin V_t$.*

Note that $\{v_1, \cdots, v_k\} \in E_t$ iff the vector $\sum_i v_i$ contains a zero entry. A universal hypergraph can be used to define a notion of hypergraph complexity.

Definition 6 *For a hypergraph H, define $\varrho(H)$ as the minimum t such that H can be embedded in H_t.*

Remark. It is easy to see that for every hypergraph H, $\varrho(H)$ is finite. This can be shown by allowing t to be much bigger than $\log |E|$ and choosing φ randomly.

Definition 7 *For a function $f \in B_n$ let $H_f = (U, \Omega_f)$ where $U = f^{-1}(0)$ and Ω_f is as before.*

Theorem 4 For any function $f \in B_n$, $\varrho(H_f) = \rho(\Omega_f)$.

Proof: *i)* $\varrho(H_f) \leq \rho(\Omega_f)$. Let $t = \rho(\Omega_f)$ and let $\{H(A^i, B^i)\}$ for $i = 1, ..., t$ be a cover of Ω_f. We define an embedding $\varphi : U \mapsto V_t$ by $\varphi(u)_i = 1, x,$ or $1 + x$ according to whether $u \in A^i \setminus B^i, B^i \setminus A^i$, or $A^i \cap B^i$. Note that these three possibilities are exhaustive as the subspace $H(A^i, B^i)$ is spanning so that $u \in A^i \cup B^i$. We claim that φ is a good embedding. Otherwise, for some $S \in \Omega_f$, $\sum_{u \in S} \varphi(u) \notin V_t$. It is easy to see that such an S is not covered by any of the subspaces.

ii)$\rho(\Omega_f) \leq \varrho(H_f)$. Let φ be an embedding of H_f in H_t. Define the subspaces $\{H(A^i, B^i)\}$ for $i = 1, ..., t$ by $u \in A^i$ iff $\varphi(u)_i \in \{1, 1+x\}$ and $u \in B^i$ iff $\varphi(u)_i \in \{x, 1+x\}$. It is clear that the subspaces are spanning. We claim that they cover Ω_f. Otherwise, it is easy to see that any set $S \in \Omega_f$ which is not covered is not mapped to an edge in E_t. ■

Characterization 2. A function f is in $co-NP$ iff there exists a polynomial $p(n)$ such that for every n, $\varrho(H_f) \leq p(n)$.

When working with hypergraph embeddings, it is helpful to understand when the embedding must be 1-1. The following claim gives a sufficient condition.

Claim 2 *If for every $x \neq 0$ there exists a $u \in U$ such that $x \oplus u \notin U$ then any embedding φ is 1-1.*

Proof: Assume that for some $u_1 \neq u_2$ we have that $\varphi(u_1) = \varphi(u_2)$. Consider the vector $x = u_1 \oplus u_2$. By the conditions of the claim, there is a $u \in U$ such that $x \oplus u \notin U$. Clearly, $\{u, u_1, u_2\} \in E_f$ but $\varphi(u) + \varphi(u_1) + \varphi(u_2) = \varphi(u) \in V_t$. ■

For example, it is easy to check that the problem clique(n,k) that checks whether a graph with n nodes contains a k-clique satisfies the conditions of the lemma, thus any embedding of its corresponding hypergraph must be 1-1.

Claim 2 can be generalized as follows: Let Z_t be the set of all even linear dependencies on V_t. That is, $\{v_1, \cdots, v_k\} \in Z_t$ iff k is even and $\sum_i v_i = 0$.

Claim 3 *If for every $x \neq 0$ there exists a $u \in U$ such that $x \oplus u \notin U$ then any embedding φ of H_f is such that $\varphi(U)$ does not contain any of the subsets in Z_t.*

4 A communication problem

Our third variation involves communication complexity. Our goal is to define, for a function f, a communication problem P_f whose communication complexity is a function of the co-non-deterministic circuit complexity of f. This characterization is similar to the equivalence between the circuit depth of a function and the communication complexity of a related problem [10]. For more information on Communication Complexity and its relationship to circuit depth the reader is referred to [8].

For a communication problem P, we write $C(P)$ for the communication complexity of P. Also, we write $C_1^{II}(P)$ for the number of bits that player II has to communicate in one round in order to make sure that player I knows the answer to the problem P.

Definition 8 *For a function $f \in B_n$ define the problem P_f as follows: Player I gets a vector $u \in U$ while player II gets a subset $S \in \Omega_f$. Their goal is to agree on a subset $A \subseteq U$ such that $u \in A$ and $|A \cap S|$ is even.*

Theorem 5 For every $f \in B_n$, $C_1^{II}(P_f) \leq \log \rho_\infty(\Omega_f)$.

Proof: Let $t = \rho_\infty(\Omega_f)$ and let $\{H_i\}$ for $i = 1, ..., t$ be a cover of Ω_f by spanning subspaces. Given $S \in \Omega_f$, player II sends the first i such that $S \in H_i$. As the subspace is spanning, there exists some vector A in the dual of H_i such that $u \in A$. This subset satisfies the requirements of the problem. ■

Note that after the protocol for P_f, player I may need many bits to communicate the answer to player II. In fact, in the proof of theorem 5, player II may use a covering of Ω_f by spanning subspaces of co-dimension 2 and guarantee that player I can respond with the answer using only 1 bit. Our next theorem provides a converse to theorem 5. Together with theorem 1 it implies that $C(P_f) = \Theta(\log \rho(\Omega_f))$.

Theorem 6 For any $f \in B_n$, $\bar{n}s(f) = 2^{O(C(P_f))}$.

Proof: Follows along the lines of the proof of theorem 2 plus some ideas from the proof of the relationship between circuit depth and communication complexity [10]. ■

Characterization 3. A function f is in $co-NP$ iff $C(P_f) = O(\log n)$. In fact, it is enough that P_f can be solved by a 2-round protocol in which player II sends $O(\log n)$ bits and player I responds with 1 bit.

Remark. It is interesting to note that for any P_f, its 2-round communication complexity where player II sends $O(\log n)$ bits and player I responds with 1 bit is within a constant factor of its 2-way unrestricted communication complexity. It could be interesting to study this phenomenon. In particular, if one wants to use reductions to prove lower bounds for P_f, one would have to use other problems with the same characteristic.

5 A representation for $co-NP$ languages

Our fourth and last variation provides a compact algebraic representation of languages in $co-NP$ in terms of two matrices over $GF(2)$.

We start our discussion with some definitions. Let σ_t be the 2-CNF formula on $2t+1$ variables defined by $y_1 \wedge (y_2 \vee y_3) \wedge (y_3 \vee y_4) \wedge \cdots \wedge (y_{2t} \vee y_{2t+1})$. Let $S_t = \sigma_t^{-1}(1)$. For two subsets $A, B \subseteq GF(2)^m$ define $A \oplus B \rightleftharpoons \{a \oplus b \mid a \in A \text{ and } b \in B\}$. Equivalently, a vector $c \in GF(2)^m$ is in $A \oplus B$ iff there exists a vector $b \in B$ such that $c \oplus b \in A$.

Definition 9 *A function $f \in B_n$ affords a* linear representation *of dimension* $2t+1$ *if there exists a linear mapping $P : GF(2)^n \mapsto GF(2)^{2t+1}$ and a subspace Q of $GF(2)^{2t+1}$ such that for every $z \in GF(2)^n$, $f(z) = 0$ iff $P(z) \in Q \oplus S$.*

Clearly, if a function f affords a linear representation of dimension $2t+1$ then we can construct a co-non-deterministic circuit for f with $O(t)$ many $\wedge$-gates. We now prove the converse.

Theorem 7 Any $f \in B_n$ affords a linear representation of dimension $2t+1$ where $t = \rho(\Omega_f) \leq 2\bar{n}s_\wedge(f)$.

Proof: Consider the $(2t+n+1) \times m$ matrix N defined in the proof of theorem 2 where $t = \rho(\Omega_f)$. Note that the first n rows of N correspond to the matrix M used in the definition of $\oplus S$. In the proof of theorem 2 we were concerned with vectors in the column space of N. Therefore, we are free to manipulate the columns of N as long as we do not change its column space. We will do this in order to make the relationship between S and $\oplus S$ more explicit.

Pick a set B of n columns of N whose first n entries span the column space of the matrix M. To make things simpler, we will assume that the vectors $e_1, \cdots, e_n$ are in U and we are going to pick the columns of N associated with these vectors. Clearly we could use any other set of vectors, but this collection makes things nicer. Next, to every column not in B we add a suitable linear combination of the columns in B so as to put zeroes in its first n entries.

Note that the relationship between S and $\oplus S$ has been made explicit since $\oplus S$ can now be read from the first n entries of S. We let N_1 be the $(2t+1) \times n$ matrix corresponding to the last $2t+1$ rows of the columns in B and let N_2 be the $(2t+1) \times (m-n)$ matrix corresponding to the last $2t+1$ rows of the columns not in B.

Let $P : GF(2)^n \mapsto GF(2)^{2t+1}$ be the linear transformation defined by the transpose of N_1. Also, let Q be the column space of N_2 in $GF(2)^{2t+1}$. Then for a vector $z \in GF(2)^n$, $f(z) = 0$ iff there exist a vector $q \in Q$ such that $P(z) \oplus q \in \sigma_t^{-1}(1)$ iff $P(z) \in Q \oplus S_t$, as required. ■

We have argued that if a function f is in $co-NP$ then it affords a linear representation of polynomial dimension. Note that the transformation P and the subspace Q can be represented by an $n \times (2t+1)$ matrix and a $(2t+1) \times (2t+1)$ matrix respectively. Also note that the definition of σ does not depend in any way on the function f. Therefore, any two matrices of the above dimensions define a $co-NP$ function and, vice versa, any $co-NP$ function can be represented by two such matrices.

Characterization 4. A function f is in $co-NP$ iff there exists a polynomial $p(n)$ such that for every n, f affords a linear representation of dimension at most $p(n)$.

Remark. Note that as a corollary to theorem 7 we get that the problem of deciding whether a given subspace contains a satisfying assignment to a given 2-CNF is NP-complete.

6 Conclusion

We have given 4 algebraic characterizations of NP, all of which have as a common ancestor the idea of combining rejecting computations of zeroes of a hard function in order to get a rejecting computation for a one of the function.

These characterizations provide us with combinatorial and algebraic frameworks in which to prove lower bounds for non-deterministic circuit size. It would be interesting to study these frameworks in a more general setting. In particular, it would be interesting to study $\rho(\Omega)$ for different subsets $\Omega \subseteq GF(2)^m$ and try to get necessary conditions which guarantee that $\rho(\Omega)$ is large.

Also, by comparing the present characterizations to others of a similar flavor but which capture other complexity classes one could get a better understanding of the differences among the different complexity classes. In particular, one should try to compare either the covering problem suggested here based on *Linear functionals*, or the covering problem suggested in Karchmer [9] based on *Self-dual monotone functionals*, both of which characterize NP, with that suggested by Razborov [17] based on *Monotone functionals* and which characterizes P.

Acknowledgments We are grateful to Drew Sutherland for helping us improve the readability of a previous draft.

References

[1] M. Ajtai, "Σ_1^1-Formulae on finite structures", *Annals of Pure and Applied Logic* **24**, pp. 1-48 (1983).

[2] N.Alon, R. Boppana, "The monotone circuit complexity of boolean functions", *Combinatorica* **7**, pp. 1-22 (1987).

[3] A.E. Andreev, "On a method for obtaining lower bounds for the complexity of individual monotone functions", *Sov. Math. Dokl.* **31**, pp. 530-534 (1985).

[4] R. Boppana, M. Sipser, "The Complexity of finite functions", In: *The Handbook of Theoretical Computer Science*, (J. van Leeuwen, ed.), Elsevier Science Publishers B.V., 1990 pp. 759-804.

[5] M. Furst, J.B. Saxe, M. Sipser, "Parity circuits and the polynomial time hierarchy", *Mathematical Systems Theory* **17**, pp. 13-27 (1984).

[6] M. Goldmann, J. Hastad, "A lower bound for monotone clique using a communication game", *Inf. Proc. Letters* **41**, pp. 221-226 (1992).

[7] J. Hastad, "Improved lower bounds for small depth circuits", *Proceedings of* 18th *STOC*, pp. 6-20 (1986).

[8] M. Karchmer, "Communication complexity: A new approach to circuit depth", The MIT Press, (1989).

[9] M. Karchmer, "On proving lower bounds for circuit size" *Manuscript* (1992).

[10] M. Karchmer, A. Wigderson, "Monotone circuits for connectivity require super-logarithmic depth", *SIAM J. of Disc. Math.* **3**, pp. 25-265 (1990).

[11] M. Karchmer, A. Wigderson, "On Span Programs", *Submitted to Structures'93*, (1992).

[12] P. Pudlák and V. Rödl, "A combinatorial approach to complexity", *Combinatorica* **12**, pp. 221-226 (1992).

[13] Raz, R., Wigderson, A.: "Monotone Circuits for Matching Require Linear Depth", *J. of the Assoc. of Comp. Machinery* **39**, pp.736-744, (1992).

[14] A.A. Razborov, "Lower bounds for the monotone complexity of some boolean functions", *Sov. Math. Dokl.* **31**, pp. 354-357 (1985).

[15] A.A. Razborov, "Lower bounds on the size of bounded depth networks over a complete basis with logical addition", *Mathematical Notes of the Academy of Sciences of the USSR* **41**, pp. 333-338 (1987).

[16] A.A. Razborov, "Applications of matrix methods for the theory of lower bounds in computational complexity", *Combinatorica* **10** pp. 81-93 (1990).

[17] A.A. Razborov, "On the method of approximations", *Proceedings of* 21st *STOC*, (1989) pp. 167-176.

[18] A.A. Razborov, "Lower bounds on the size of switching-and-rectifier networks for symmetric Boolean functions", *Math. Notes of the Academy of Sciences of the USSR*, 48(6), pp. 79-91, (1990).

[19] M. Sipser, "A topological view of some problems in complexity theory", *Colloquia Mathematica Societatis János Bolyai* **44**, pp 387-391, (1984).

[20] A. C.-C. Yao, "Separating the polynomial-time hierarchy by oracles", *Proceedings of* 26th *FOCS*, pp. 1-10 (1985).