Lower Bounds for Computation with Limited Nondeterminism

Hartmut Klauck
Johann-Wolfgang-Goethe-Universität Frankfurt
60054 Frankfurt am Main, Germany
klauck@thi.informatik.uni-frankfurt.de

Abstract

We investigate the effect of limiting the number of available nondeterministic bits in different computational models. First we relate formula size to one-way communication complexity and derive lower bounds of $\Omega(n^{2-\epsilon}/\log^{1-\epsilon} n)$ on the size of formulae with $n^{\epsilon}/\log^{\epsilon} n$ nondeterministic bits for $0 < \epsilon \leq 1/2$. Next we prove a rounds-communication hierarchy for communication complexity with limited nondeterminism solving an open problem of [HrS96]. Given a bound s on the number of nondeterministic bits and a number k of rounds we construct a function which can be computed deterministically in k rounds with $O(sk \log n)$ bits communication, but requires $\Omega(n/(s^2k^2 \log n))$ in $k-1$ rounds though s nondeterministic bits are available. We apply this result to show a reversal hierarchy for 2-way automata with limited nondeterminism exhibiting exponential gaps. Furthermore we investigate the effect of limited nondeterminism on monotone circuit depth. All results presented in the paper have the common core that limited nondeterministic communication has high round dependence.

1 Introduction

Some of the fundamental open questions in complexity theory concern the power of nondeterminism over determinism. While for Turing machines almost no results are known which show that nondeterminism actually helps (except [PST83]), the power of nondeterminism is far better understood in the world of communication complexity ([KN96], [Hr97]).

There has been extensive research on the subject of *limited nondeterminism*, i.e., the subject of viewing nondeterministic bits as a resource. This research was focussed mainly on the areas of Turing machine complexity, circuit complexity, and automata, an overview and many references are given in [GLM96], a very general definition of computation with limited nondeterminism is given in [CC97]. Lower bound results derivable in the first two areas so far, however, share the drawback that they rely upon unproven assumptions like $\mathcal{P} \neq \mathcal{NP}$. This paper gives explicit lower bounds on computation with limited nondeterminism in various models of computation.

The investigation of limited nondeterminism in communication complexity has been started recently in [HrS96]. While probabilistic protocols need only a very limited number of probabilistic bits, namely $\log n + O(1)$ (see [Ne91], logarithms are base 2 in the paper), seemingly insignificant restrictions on the number of guess bits may result in drastically increased communication in the nondeterministic model.

A prominent feature of unlimited nondeterminism is that one-round protocols suffice to achieve optimal communication, which stands in sharp contrast to deterministic and randomized protocols (see [DGS87], [HR93], and [NW93] for round-hierarchies in communication complexity). Our central result (theorem 2) shows that limited nondeterministic communication has a strong round dependence. This feature is the core of all results in this paper.

We begin with a result using one-way communication complexity. In [Kl97] we showed that the Nečiporuk lower bound on formula size (see [N66], [W87], or [BS90]) can be rephrased in terms of asymmetrical one-way communication complexity and derived lower bounds on probabilistic formulae via probabilistic communication complexity arguments. Here we derive a lower bound on the size of formulae with limited nondeterminism. We devise a suitable variant of the iterated disjointness problem and prove lower bounds of $\Omega(n^{2-\epsilon}/\log^{1-\epsilon} n)$ when nondeterminism is limited to $n^{\epsilon}/\log^{\epsilon} n$ bits for every $0 < \epsilon \leq 1/2$. Due to inherent restrictions our Nečiporuk method cannot prove better lower bounds than n^2/s for s guess bits, so significantly better results need to use new techniques. In the case of unlimited nondeterminism the asymmetricity of the input partition becomes unimportant and only trivial lower bounds are possible through our method. This is no coincidence, since general nondeterministic formulae are as powerful as nondeterministic circuits with regard to size. The functions for which we prove

the lower bounds can be computed more efficiently when nondeterminism is less restricted: $O(n^\epsilon \log^{1-\epsilon} n)$ nondeterministic bits suffice for size $O(n^{1+\epsilon}/\log^\epsilon n)$.

Then we turn to general communication complexity. A weak rounds-communication hierarchy for limited nondeterminism is claimed in [HrS96] holding only for at most a logarithmic number of guess bits. We show that for any s and k there is a function on n inputs which can be computed deterministically in k rounds with communication $O(sk \log n)$ when A starts, while any k-round protocol with s nondeterministic bits needs communication $\Omega(n/(s^2k^2 \log n))$ when B starts.

We continue with an application of this hierarchy. A k-visit automaton is a 2-way NFA, which is allowed to visit each input bit at most k times, thus the notion of a k-visit automaton is a stronger form of the notion of a k-reversal bounded automaton. Astonishingly there is a hierarchy over k even for unlimited nondeterminism, but the hierarchy for limited nondeterminism is much stronger, having exponential gaps (upper bounds hold for k-reversal automata, so a reversal hierarchy is implicit). Limited nondeterminism for one-way automata has been investigated e.g. in [GKW90], reversal complexity for 2-way NFAs e.g. in [Hr91]. Note that for Turing machines the reversal measure usually includes the reversals on the work tapes, see e.g. in [CY91].

Finally we investigate monotone circuit depth. Due the Karchmer Wigderson result in [KW90] there is an equivalence between communication complexity in a specific game and monotone bounded fan-in circuit depth (this equivalence can also be interpreted in terms of (unbounded fan-in) depth and logarithm of fan-in versus rounds and communication per round, see [NW93] and [K88]). It is easy to show that depth can be reduced to the logarithm of size by unlimited nondeterminism without a significant increase in size. We construct a family of functions on n variables for every d (with $\sqrt{n} \geq d \geq \log n$) that can be computed in monotone depth $\Theta(d)$ with at most $\epsilon n/d$ nondeterministic bits for some constant $\epsilon > 0$. n/d nondeterministic bits suffice for monotone depth $O(\log n)$. The construction builds on the lower bound for the matching problem in [RW92]. A similar result can be derived for functions in monotone $\mathcal{P}$ via a recent result of [RM97]. This separates the monotone $\mathcal{NC}$ hierarchy in the presence of $n/\text{polylog } n$ nondeterministic bits.

s,t-connectivity is an important monotone problem. We show that the monotone depth is $\Omega(\log^2 k)$ when $(n/k) \log n$ nondeterministic bits are available.

Nisan and Wigderson ([NW93]) rederived the separation of monotone $\mathcal{AC}^0$ given in [KPY84] from the deterministic round hierarchy in communication complexity ([DGS87]). We generalize the result to limited nondeterminism.

2 Definitions and Results

We first need definitions on communication complexity and the other computational models considered here.

Definition 1 *Let $R \subseteq X_A \times X_B \times Y$ be a relation. In a communication protocol player A and B receive inputs x_A, x_B from X_A and X_B respectively. Their goal is to agree on $y \in Y$ such that $(x_A, x_B, y) \in R$. The players exchange binary encoded messages. The communication complexity of a protocol is the worst case number of bits exchanged. The communication complexity $C(R)$ of a relation R is the complexity of the optimal protocol solving R. The communication matrix of R is the matrix M with $M(x_A, x_B) = \{y | (x_A, x_B, y) \in R\}$.*

A nondeterministic protocol for a Boolean function f is 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. If we limit the number of nondeterministic bits to some value s, then we assume that these bits are shared and known to both players without communication, otherwise they are private. The nondeterministic communication complexity $NC(f)$ is the complexity of the optimal nondeterministic protocol computing f. $NC_s(f)$ is the complexity of the optimal nondeterministic protocol using at most s nondeterministic bits and computing f.

A protocol has k rounds, if the players exchange k messages (of any length) switching the speaker on very message. Superscripts of the form $C^{A,k}$ denote the complexity, when the protocols are restricted to at most k rounds with player A starting. If $k = 1$ then we will also use the term one-way communication complexity.

Definition 2 *A (Boolean) formula is a Boolean circuit with fan-in 2 and fan-out 1 on Boolean variables $x_1, \ldots, x_n$ and the constants 0,1, that can all be read arbitrarily often. The gate functions may be chosen from an arbitrary basis.*

A nondeterministic formula with s guess bits is a formula which has access to additional input variables $a_1, \ldots, a_s$. The formula accepts an input x if there is an a such that (a, x) is accepted.

The size of a formula is the number of leaves, the size of a circuit is the number of gates. The depth of a formula or circuit is the length of the longest path from a leaf to the top gate.

A monotone circuit is only allowed to use the gate functions $\wedge$ and $\vee$ (we consider fan-in 2 and, if indicated, unbounded fan-in). A nondeterministic monotone circuit has access to nondeterministic bits $a_1, \ldots, a_s, \overline{a_1}, \ldots, \overline{a_s}$ (the guess bits are given negated and unnegated to the circuit, note that guess bits which are given only unnegated can be removed).

Definition 3 *A $k-$visit automaton is a 2-way NFA which visits each letter at most k times. The automaton has bounded nondeterminism $s(n)$, if on each input of length n at most $s(n)$ nondeterministic bits are guessed. The size of the automaton is the number of its states.*

Obviously by definition k-visit is at least as strong as the reversal measure, since a (nonresting) automaton that changes the direction of its head k times (including the start) visits each letter at most k times.
Now we can state the results of this paper formally. We claim lower bounds for exact values s of available guess bits, though we later prove them for values ϵs. This means that the functions in the claims are scaled versions of the equally named functions in the proofs. We begin with the lower bound on formula size.

Theorem 1 *There is a Boolean function $D_{n,s}$ (with $s \leq n$) on $N = O(ns \log n)$ inputs such that any formula with s nondeterministic bits for $D_{n,s}$ has size $\Omega(n^2 s \log n)$. $D_{n,s}$ can be computed by a formula with $O(s \log n)$ nondeterministic bits and size $O(ns^2 \log n)$. In particular for $0 < \epsilon \leq 1/2$ let $s = n^{\frac{\epsilon}{1-\epsilon}}$, then the size is lower bounded by $\Omega(N^{2-\epsilon}/\log^{1-\epsilon} N)$ even with $N^\epsilon/\log^\epsilon N$ guess bits. $O(N^\epsilon \log^{1-\epsilon} N)$ guess bits suffice for size $O(N^{1+\epsilon}/\log^\epsilon N)$.*

Note that nondeterministic formulae with unlimited nondeterminism can simulate nondeterministic circuits within a constant factor increase in size. Our argument develops a variant of the Nečiporuk method. It is easy to see that this method is inherently restricted such that for every function f on N inputs with s guess bits the provable lower bound is smaller than N^2/s.
This is our result on the nondeterministic rounds-communication hierarchy:

Theorem 2 *For every s, k there is a Boolean function f_k^s with input length n and*

- $C^{A,k}(f_k^s) = O(sk \log n)$
- $NC_s^{B,k}(f_k^s) = \Omega(n/(s^2k^2 \log n))$
- $NC_{O(s \log n)}^{B,k}(f_k^s) = O(sk \log n)$.

The function requires to follow a set of $\Theta(s)$ paths in a bipartite graph up to the kth vertices. Note that one can save a round when following a single path by guessing an edge with $\log n$ nondeterministic bits.
Turning our attention to 2-way automata with unlimited nondeterminism we find a weak hierarchy.

Theorem 3 *Any language that can be computed by a k-visit automaton of size q can also be computed by a $k-j$-visit automaton of size $O(q^{j+2})$ (if $j < k$).*
There is a language $L \subseteq \{0,1\}^n$ such that any k-visit automaton for L needs size at least $\Omega(N^{1/k})$ and at most $O(N^{1/k} \log^2 N)$, where $N = \Theta(2^{n/2})$ is the minimal size of a 1-visit automaton for L.

For limited nondeterminism there is a stronger hierarchy:

Theorem 4 *For any s, k there is a language $F_k^s \subseteq \{0,1\}^n$ which can be decided by a k-visit automaton that works deterministic and has size $kn^{O(s)}$. F_k^s requires size $2^{\Omega(n/(s^2k^3 \log n))}$ for any $k-1$-visit automaton with s guess bits.*

The depth of monotone circuits can be reduced by nondeterminism. Even the circuit structure can be destroyed without increasing size much.

Theorem 5 *A monotone nondeterministic circuit with size c can be converted to an equivalent monotone nondeterministic formula with depth $\log c + O(1)$ and size $O(c)$. If unbounded fan-in is allowed then depth 2 and size $O(c)$ suffice.*

Again limiting nondeterminism may lead to maximal depth, as shown in the next result, which uses direct sums and employs the deterministic lower bound for bipartite perfect matching from [RW92].

Theorem 6 *Let d, n such that $\sqrt{n} \geq d \geq \log n$. There is an explicit Boolean function g_n^d on n variables and a constant $\epsilon > 0$ so that g_n^d can be computed by a monotone deterministic circuit of depth $O(d)$, and every nondeterministic monotone circuit with $\epsilon n/d$ guess bits needs depth $\Omega(d)$. g_n^d can be computed in monotone depth $O(\log n)$ with n/d guess bits.*

Recently a separation of the monotone $\mathcal{NC}$ hierarchy was achieved in [RM97]. Employing this result we get the following.

Theorem 7 *There are constants $c, \epsilon > 0$ such that the following holds: Let d, n such that $n^{1/c} \geq d \geq \log n$. There is an explicit Boolean function h_n^d on n*

variables that can be computed by a monotone deterministic circuit of polynomial size in depth $O(d)$, and every nondeterministic monotone circuit with $\epsilon \frac{n \log^{3c} d}{d^{3c-1}}$ guess bits needs depth $\Omega(d)$. h_n^d can be computed in monotone depth $O(\log n)$ if $\frac{n \log^{2c} d}{d^{2c}}$ guess bits are allowed.

For $d = \log^k n$ theorem 7 gives a function computable in polynomial size, but only in depth $\Theta(d)$, although n/polylog n nondeterministic bits are available, and thus separates the monotone $\mathcal{NC}$-hierarchy for limited nondeterminism.
In the s,t-connectivity problem one has to decide for a given graph with distinguished vertices s, t, whether there is a path from s to t. The deterministic monotone depth is $\Theta(\log^2 n)$ (see [KW90]). Intuitively for a drastical depth reduction much information on the path has to be guessed nondeterministically. We confirm this intuition:

Theorem 8 *The depth of a monotone circuit for s,t-connectivity using $(n/k)\log n$ nondeterministic bits $(1 \leq k \leq n)$ is $\Omega(\log^2 k + \log n)$ and $O(\log n \log k)$.*

We conjecture the "tight" bound to be $\Theta(\log n \log k)$. Now consider circuits with unbounded fan-in. For limited nondeterminism the monotone $\mathcal{AC}^0$ hierarchy is strict.

Theorem 9 *For every $k \geq 2$ and $s \geq n$ there is a function q_k^s on $N = O(sn^k)$ inputs that can be computed by a deterministic monotone circuit with fan-in $O(s)$, depth $k+1$, and size $O(N)$. Any monotone circuit with depth k, unbounded fan-in, and s nondeterministic bits for q_k^s has size $2^{\Omega((N/s)^{1/k}/k)}$.*
q_k^s can be computed by a monotone circuit with unbounded fan-in, $O(s \log n)$ nondeterministic bits, depth k, and size $O(N)$ (with the exception of $k = 2$, where the size is $O(N \log N)$).

Nondeterminism allows simplifications of the structure of computation in many models. The conclusion of our results is that this has to be paid by a high consumption of nondeterministic bits. When this resource is restricted, "round phenomena" analogous to the deterministic world appear.
In the following sections the proofs of the theorems are given. The proofs of some propositions are given in an appendix.

3 A lower bound on formula size

First we introduce a variation of the Nečiporuk method in terms of communication:

Definition 4 *Let f be a Boolean function on n inputs and $y_1 \dots y_k$ a disjoint partition of the input variables into k blocks.*
Let player B know the inputs in y_i and A all the other inputs. The nondeterministic one-way communication complexity with s guess bits of f under this partition is called $NC_s^{A,1}(f^i)$. Define the s-nondeterministic Nečiporuk function to be $1/4 \sum_{i=1}^k NC_s^{A,1}(f^i)$.

Proposition 1 *The s-nondeterministic Nečiporuk function is a lower bound on the size of nondeterministic Boolean formulae with s guess bits.*

Definition 5 *$\mathcal{P}_{m,s}$ denotes the set of all size s subsets of a m element universe.*
$D_{n,s}$ denotes the following language (and its characteristic function) for $1 \leq s \leq n$:

$$D_{n,s} = \{(x_1, \dots, x_{n+1}) | \forall i : x_i \in \mathcal{P}_{n^3,s} \\ \wedge \exists i : |\{j | j \neq i; x_i \cap x_j \neq \emptyset\}| = s\}.$$

We consider the $n+1$ partitions of the inputs, where B has set x_i and A has the remaining n sets.

Lemma 1 *Let the partition of the input be as described. There is some constant $\epsilon > 0$ such that for all i*

$$NC_{\epsilon s}^{A,1}(D_{n,s}^i) = \Omega(ns \log n)$$

PROOF: We have to show that all nondeterministic one-way protocols with ϵs guess bits for $D_{n,s}^i$ communicate much. A nondeterministic one-way protocol with ϵs guess bits and communication c is equivalent to a cover of the communication matrix for partition i with $2^{\epsilon s}$ matrices that have at most 2^c different rows and contain no ones which are wrong. We first construct a large submatrix with some nice properties, and show the lemma for this "easier" problem.
Partition the universe into n disjoint sets $U_1, \dots, U_n$ with $|U_i| = n^2 = m$. We choose vectors of n subsets of the universe such that the ith subset belongs to U_i. So all the sets are pairwise disjoint. Now the question the protocol has to answer, is whether the set of B intersects with s sets on the side of A.
We will further restrict the set of inputs. There are $\binom{m}{s}$ subsets of U_i of size s. We select a set of such subsets so that any two of these have no more than $s/2$ common elements. For this start at some subset and remove all subsets in "distance" at most $s/2$. This is continued as long as possible, so a set of subsets of U_i has been chosen which have pairwise "distance" $s/2$. Since in one step at most $\binom{s}{s/2}\binom{m}{s/2}$ subsets are removed the number of selected subsets is at least

$$\frac{\binom{m}{s}}{\binom{s}{s/2}\binom{m}{s/2}} \geq \left(\frac{m}{s}\right)^{s/2} / 2^{3s/2}.$$

Consider the rows belonging to vectors of subsets drawn as described. The columns can be restricted to vectors of elements drawn from $U_1 \times \cdots \times U_n$, where only s positions are not "empty". This submatrix is called M.
Consider any protocol matrix M' which covers at least $1/2^r$ ones of M with $r = \epsilon s$. We will show that any such matrix needs many different rows.
Each row of M corresponds to a vector of sets. Let us say that a set of rows of M has k *difference positions* if there are k positions in the corresponding set vectors such that for each of the positions at least two set vectors differ.
We show that any row with many ones does not agree with many rows of M, i.e., contains ones these do not have. Since the matrices in the cover have one-sided error their rows are thus either sparse or cannot be put on many rows of M. Note that each row of M contains $\binom{n}{s}s^s$ ones.

Proposition 2 *Let z be a row of M' that appears t times in M'. Let the rows of M appearing at the same positions have δn difference positions. Then z has at most $2\binom{n}{s}s^s/2^{\delta s/9}$ ones.*

At least one half of all ones in M' appear in rows with at least $\binom{n}{s}s^s/2^{r+1}$ ones. Due to proposition 2 such a row appears on a set of rows of M which have no more than δn difference positions with $r+1 = \delta s/9 - 1$. So the row can cover at most all ones in $\binom{m}{s}^{\delta n}$ rows of M and thus at most $\binom{m}{s}^{\delta n}\binom{n}{s}s^s$ ones. Since at least $(m/s)^{sn/2}\binom{n}{s}s^s/(2^{3sn/2}2^{r+1})$ ones have to be covered by these rows

$$\begin{aligned}
&\frac{(m/s)^{sn/2}\binom{n}{s}s^s}{\binom{m}{s}^{\delta n}\binom{n}{s}s^s 2^{3sn/2}2^{r+1}} \\
&\geq \frac{(m/s)^{sn/2}}{(em/s)^{9\epsilon sn+18n}2^{3sn/2}2^{\epsilon s+1}} \\
&= 2^{\Omega(sn\log n)}
\end{aligned}$$

rows are needed (for $\epsilon = 1/20$ and $s \geq 400$). □

Theorem 1 *Any nondeterministic formula with s guess bits for $D_{n,20s}$ needs at least size $\Omega(n^2 s\log n)$. $D_{n,s}$ can be computed by a nondeterministic formula of size $O(ns^2\log n)$ when $O(s\log n)$ guess bits are allowed.*

Proof Sketch: The lower bound follows from lemma 1 and proposition 1. For the upper bound consider the following: the formula uses a guess string consisting of a number i with $1 \leq i \leq n$ and pairs $(j_1, w_1), \ldots, (j_s, w_s)$, where $1 \leq j_k \leq n$ and $1 \leq w_k \leq n^3$ for all k. The i indexes a set and the pairs witness that set i and set j_k intersect on element w_k.
The formula checks the following: $j_1 < \cdots < j_s$. This ensures that we have witnesses for s different sets. Also check for all $1 \leq l \leq n$ that if $l = i$, then all the elements in the guess string are in x_l, if $l = j_k$ then $w_k \in x_l$. All these tests can be implemented in size $O(ns^2\log n)$. □

4 A rounds-communication hierarchy

Nisan and Wigderson proved in [NW93] that an explicitly given function has a randomized k round communication complexity of $\Omega(n/k^2 - k\log n)$ if B starts communicating and a deterministic k round communication complexity of $k\log n$ if A starts communicating. This function requires following a path of length k in a bipartite graph with outdegree 1. However, guessing one edge with $\log n$ nondeterministic bits allows B to use the first round productively and finish in k rounds without increased communication. So a "harder" function is needed. We investigate a function which generalizes the function considered in [NW93].

Definition 6 *Let V_A and V_B be disjoint sets of n vertices each.*
Let $F_A = \{f_A | f_A : V_A \to V_B\}$, and $F_B = \{f_B | f_B : V_B \to V_A\}$.
Then let $f(v) = f_A(v)$ (resp. $f_B(v)$) if $v \in V_A$ (resp. $v \in V_B$).
Define that $f^{(0)}(v) = v$ and $f^{(k)}(v) = f(f^{(k-1)}(v))$.
Then $g_k^s : V_A^s \times F_A \times F_B \to (V_A \cup V_B)^s$ is defined by $g_k^s(v^1, \ldots, v^s, f_A, f_B) = (f^{(k)}(v^1), \ldots, f^{(k)}(v^s))$.
The function $f_k^s : V_A^s \times F_A \times F_B \to \{0,1\}$ is the PARITY of all bits in the binary code of the output of g_k^s.

The above definition generalizes the function considered in [NW93], now following s paths in parallel (but on one single graph). Player A receives an element of F_A, player B receives an element of F_B, and both know the s start vertices. Intuitively player B has to guess $s\log n$ bits in order to use the first round productively. In [NW93] a randomized zero-error (and thus nondeterministic with $O(\log n)$ guess bits) protocol is described computing g_k^1 within communication $O((n/k)\log^2 n)$, k rounds with B starting. This protocol does not generalize to g_k^s, so no better upper bound than $O(n\log n)$ is known when only $o(s\log n)$ nondeterministic bits are available. For our lower bound some notation is useful.

Definition 7 *Let Ω be a finite set, $X \subseteq \Omega$, $Pr : X \to [0,1]$ a probability distribution, and $x \in X$ be a random variable distributed with Pr. Subsets of X are events.*

The entropy of X is $H(x) = -\sum_{z \in X} Pr(z) \log Pr(z)$. The entropy of X given an event W of a random variable $y \in Y \subseteq \Omega$ (distributed with $Pr' : Y \to [0,1]$) is $H(x|y \in W) = -\sum_{z \in X} Pr(z|y \in W) \log Pr(z|y \in W)$. The conditional entropy of X given Y is $H(x|y) = E_a[H(x|y = a)]$, the expectation over the elementary events of Y.
The information on X is $I(x) = \log|\Omega| - H(x)$. Conditional information is defined by $I(x|y \in W) = \log|\Omega| - H(x|y \in W)$ and $I(x|y) = \log|\Omega| - H(x|y)$.

Theorem 2 will be derived mainly from the following lemma.

Lemma 2 $NC^{B,k}_{s/3200}(f^s_k) \geq \frac{n}{2s^2k^2} - 3sk \log n$

PROOF: As in [NW93] we consider k round protocols (with B starting) for f^s_k, which must communicate in round $t > 1$ additionally $v^i_{t-1} = f^{(t-1)}(v^i)$ for all i, i.e., follow all paths with one round delay. This costs $sk \log n$ bits extra. Assume that overall $\epsilon/2n - 2sk \log n$ bits for $\epsilon = 1/(s^2k^2)$ are sent. We will show that this communication does not suffice thus proving the claimed lower bound for general protocols.
A nondeterministic protocol with $s/3200$ guess bits consists of $2^{s/3200}$ deterministic protocols which recognize languages that cover f^s_k. A deterministic protocol tree is a tree, where the sons of a node correspond to the messages sent at their father. The nodes z can be labeled with the submatrices $F^z_A \times F^z_B$ of inputs that share the communication on the path from the root to the node. Let $\delta = \sqrt{\epsilon}/20$ and let c_z bits be communicated before the node is reached. Assume that it is A's turn to speak and that z is in depth t. Nisan and Wigderson call z *nice* (for the ith path) if

1. $I(f^z_A) \leq 2c_z$
2. $I(f^z_B) \leq 2c_z$
3. $I(f^z_B(v^i_{t-1})) \leq \delta$

with an analogous definition for the situation when B speaks. The main lemma of [NW93] can be stated as follows.

Fact 3 *If z is nice for i, and w is a random child of z, then $Pr[w \text{ not nice for } i] \leq 22\sqrt{\epsilon} + \frac{1}{n}$, where children are chosen with probability proportional to the number of inputs arriving at them.*
With probability at most $22k\sqrt{\epsilon} + k/n$ a uniformly chosen input is in a leaf which is not nice for i.

Consider a deterministic protocol tree for f^s_k. We show that with probability at least $1 - 2^{-s/3195+1}$ a randomly chosen input reaches a leaf in which the information on the end vertices of $\Omega(s)$ paths, each conditioned on the values of all other paths, is small. This allows to bound the distributional complexity of f^s_k with a large one-sided error and hence the limited nondeterministic complexity.
Consider one of the s paths to be traced. All other paths use at most $(s-1)k$ edges. Let the random variable $y \in Y$ correspond to the possibilities to fix exactly $(s-1)$ paths. Then Y contains all sets of inputs with the property that the inputs in one set have exactly $s-1$ paths in common. Let the random variable y be distributed uniformly. The subsets $W \subseteq Y$ are events, and some of these correspond to fixing $r < s-1$ paths.
Call a node of the protocol tree *very nice* for i if

1. $I(f^z_A|y) \leq 2c_z + 4sk \log n$
2. $I(f^z_B|y) \leq 2c_z + 4sk \log n$
3. $I(f^z_B(v^i_{t-1})|y) \leq \delta$

again with an analogous definition for the situation when B speaks. The following can be proved in a similar fashion as fact 3.

Proposition 4 *Let $W \subseteq Y$ be an event that does not contain an elementary event fixing path i. An input chosen randomly under the condition W reaches a leaf which is not very nice for i with probability at most $22k\sqrt{\epsilon} + k/n$.*

Consider the experiment of choosing a random input uniformly. Let the random variable x_i be 1 if the input reaches a leaf which is not very nice for i. Then let $x = \sum x_i$ under the single experiment of picking an input uniformly. In order to analyze the probability that x gets large we use martingales (see e.g. [MR95]).

Definition 8 *Let x be an integer valued random variable. The expected value of x is $E[x] = \sum_a a \cdot Pr(x = a)$, the conditional expectation is $E[x|y = u] = \sum_a a \cdot Pr(x = a|y = u)$, and the random variable $E[x|y]$ is $f(u) = E[x|y = u]$. For a σ-field $(\Omega, \mathcal{F})$ the expectation $E[x|\mathcal{F}]$ is defined by $E[x|y]$ for some random variable y that takes distinct values on the elementary events of $\mathcal{F}$.*

The definition of $E[x|\mathcal{F}]$ does not depend on the specific values of y on elementary events.

Definition 9 *Given the σ-field $(\Omega, \mathcal{F})$ with $\mathcal{F} = 2^\Omega$ any sequence $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_m$ of subsets of 2^Ω is called a* filter *if*

1. $\mathcal{F}_0 = \{\emptyset, \Omega\}$
2. $\mathcal{F}_m = 2^\Omega$
3. $(\Omega, \mathcal{F}_i)$ *is a σ-field for all i.*

Instead of defining martingales formally we only state a description how to create one.

Fact 5 *Let $(\Omega, \mathcal{F}, Pr)$ be a probability space with a filter $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_m$.*
Let x be a random variable over this space and let $Z_i = E[x|\mathcal{F}_i]$. Then the sequence $Z_0, Z_1, \ldots, Z_m$ is a martingale.

This sequence is usually referred to as the Doob Martingale of x. A filter can be generated by defining a sequence of partitions, where the partition i corresponds to the set of elementary events in $\mathcal{F}_i$. Consider the filter generated by the following partitions of the inputs: $\mathcal{F}_0$ is trivial, i.e., contains the set of all inputs and the empty set. For $\mathcal{F}_1$ partition the set of all inputs according to the possibilities to fix path 1 including its starting vertex. For $\mathcal{F}_i$ partition the previous partition according to the possibilities to fix path i including its starting vertex, i.e., partition all inputs according to the possibilities to fix paths 1 to i. $\mathcal{F}_{s+1}$ equals 2^Ω.
Recall that the random variable of interest is $x = \sum x_i$, the number of paths, for which a random leaf is not very nice. Let $Z_0 = E[x]$ and $Z_i = E[x|\mathcal{F}_i]$, i.e., the expected value of the sum and the random variable of the expected value of the sum depending on how the first i paths are fixed. We want to bound the probability that $|Z_s - Z_0|$, i.e., the difference between x and its expectation is large with Azuma's Inequality.

Fact 6 *Let $Z_0, Z_1, \ldots$ be a martingale sequence such that*

$$|Z_i - Z_{i-1}| \leq c_i.$$

Then for all $t > 0$ and $\lambda > 0$,

$$Pr(|Z_t - Z_0| \geq \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^t c_i^2}\right).$$

What happens to the expected value of the sum, if one more path is fixed? The last vertex of that path may reach large information, so we must assume that $x_i = 1$. Other changes in $E[x] = \sum E[x_i]$ can be bounded as follows:

$$E[\sum_{j>i} x_j|\mathcal{F}_i] \leq \sum_{j>i} E[x_j|W_j]$$

for events $W_j \subseteq Y$ and so we have that

$$\begin{aligned}
&|Z_i - Z_{i-1}| \\
&\leq 1 + E[\sum_{j>i} x_j|\mathcal{F}_i] - E[\sum_{j>i} x_j|\mathcal{F}_{i-1}] \\
&\leq 1 + \sum_{j>i} E[x_j|W_j] \\
&\leq 1 + s(22k\sqrt{\epsilon} + k/n) \leq 24
\end{aligned}$$

with proposition 4. With

$$Z_r = E[x] = \sum E[x_i] \leq s \cdot (22k\sqrt{\epsilon} + k/n) \leq 23$$

we have

$$\begin{aligned}
Pr(\sum x_i > s/2 + 23) &\leq Pr(|Z_r - Z_0| \geq s/2) \\
&\leq 2e^{-\frac{s^2/4}{2s \cdot 576}} \leq 2 \cdot 2^{-s/3195}.
\end{aligned}$$

Now we can deduce the lemma: Imagine a nondeterministic protocol with B starting, k rounds and $s/3200$ guess bits. This protocol induces a deterministic protocol which accepts at least $2^{-s/3200}$ of all 1-inputs and consequently $2^{-s/3200-1}$ of all inputs. But any corresponding protocol tree with the allowed communication has the property that with probability at most $2^{-s/3195+1}$ a leaf is reached in which there are no paths that have conditioned information at most δ. An accepting leaf of the protocol tree has an associated matrix of inputs which is monochromatic, i.e., the parity of the kth vertices is known. But then $I(f^{(k)}(v^i)|y = a) \geq 1$ for all i and a. So for no i $I(f^{(k)}(v^i)|y) \leq \delta$ and monochromatic leafs are not very nice for any i.
Thus no protocol tree with the allowed communication and depth and B starting can compute a fraction of $2^{-s/3200}$ of all 1-inputs. □
For the conclusion of theorem 2 we take a look at the nondeterministic communication complexity with unlimited nondeterminism.

Proposition 7 $NC(f_k^s) = \Omega(sk \log(n/(sk)))$.

Lemma 2 and proposition 7 imply the lower bound of theorem 2 for the function f_k^{3200s}. The upper bounds are obvious.

5 Hierarchies for 2-way automata

We will present an application of the rounds-communication theorem to automata. First consider automata with unlimited nondeterminism, for which a proper hierarchy on the number of visits holds.

Theorem 3 *Any language that can be computed by a k-visit automaton of size q can also be computed by a $k-j$-visit automaton of size $O(q^{j+2})$ (if $j < k$).*
There is a language $L \subseteq \{0,1\}^n$ such that any k-visit automaton for L needs size at least $\Omega(N^{1/k})$ and at most $O(N^{1/k}\log^2 N)$ for $N = \Theta(2^{n/2})$, the minimal size of a 1-visit automaton for L.

PROOF SKETCH: Consider any language computed by a k-visit automaton A. For every input $x_1, \ldots, x_n$ the crossing sequences c_i contain the sequence of states in which A visits x_i. Since A is k-visit, all crossing sequences are at most k long. A performs a j-excursion between i and l if it starts at i, moves arbitrarily on the positions between i and l such that each cell is visited at most $j+1$ times, and reaches i or l. A j-excursion can be shortcut by nondeterminism: guess a crossing sequence of length $j+1$ on i. Then move to the next cell, while guessing a crossing sequence that is consistent with the former crossing sequence and the input, if possible (else stop and reject). For this the "storage" for $j+1$ states is sufficient. So when l is reached the whole excursion has been performed (if the guessing was good), without the way back to i. If the way back is needed, the whole process can be performed backwards again, until i is reached (guess when to stop). So $O(q^{j+1})$ states are used, but only $j-1$ visits may be saved. Since the beginning of excursions can be guessed, too, it suffices to show that every k-visit tour across the input becomes $k-j$-visit if an optimal set of $j+1$-excursions is shortcut. The proof for this is omitted.
For the second part consider the language $L = \{xy | n/2 = |x| = |y| \wedge x = y\}$. Any k-visit automaton deciding L can be simulated by a k round protocol, in which A receives x and B y, such that the communication in each round is bounded from below by the logarithm of the size. Since the nondeterministic communication complexity of L is $n/2$, at least one round must communicate $n/(2k)$ bits. On the other hand an automaton can read $n/(2k)$ bits of x in each round, move to the right position in y, compare, and iterate. For this counters of together $2\log n$ bits and $n/(2k)$ bits "storage" suffice. □
The hierarchy between k and $k-1$ visits is strict up to $k = \Theta(\sqrt{\log N/\log\log N})$, but the gaps are polynomial. This changes drastically for limited nondeterminism.

Theorem 4 *For any s, k there is a language $F_k^s \subseteq \{0,1\}^n$ which can be decided by a k-visit automaton that works deterministic and has size $kn^{O(s)}$. F_k^s requires size $2^{\Omega(n/(s^2k^3\log n))}$ for any $k-1$-visit automaton with s guess bits.*

PROOF SKETCH: Consider the language of the function f_k^{3200s} from section 4. A k-visit automaton reads and stores $3200s$ pointers, moves to their respective positions and so forth. Obviously k sweeps across the input suffice.
On the other hand with $k-1$ visits the size is at least 2^c for the communication $c = \Omega(n/(s^2k^3\log n))$ of one round in a protocol with s guess bits. □

6 Monotone Circuit Depth

We first turn to unlimited nondeterminism and find that depth-reduction is possible (the proof is omitted).

Theorem 5 *A monotone nondeterministic circuit with size c can be converted to an equivalent monotone nondeterministic formula with depth $\log c + O(1)$ and size $O(c)$. If unbounded fan-in is allowed then depth 2 and size $O(c)$ suffice.*

It is easy to see that a circuit with s nondeterministic bits can be made deterministic within additive depth increase s by ORing over all choices. So strong deterministic lower bounds lead to lower bounds for limited nondeterminism. But what if much more nondeterministic bits than depth are allowed?

Theorem 6 *Let d, n such that $\sqrt{n} \geq d \geq \log n$. There is an explicit Boolean function g_n^d on n variables and a constant $\epsilon > 0$ so that g_n^d can be computed by a monotone deterministic circuit of depth $O(d)$, and every nondeterministic monotone circuit with $\epsilon n/d$ guess bits needs depth $\Omega(d)$. g_n^d can be computed in monotone depth $O(\log n)$ with n/d guess bits.*

PROOF SKETCH: The input consists of n/d^2 bipartite undirected graphs on $2d$ vertices each coded by 0 for edges and 1 for non-edges. The function decides whether all graphs do not contain a perfect matching. This is the direct sum of the dual of the bipartite perfect matching problem, is monotone, and can be computed in monotone depth $O(d)$.
With n/d guess bits choose n/d^2 subsets of the left side vertices of the n/d^2 input graphs. For every subset check whether its neighbor set is smaller than itself. This check suffices due to Hall's theorem. To implemented this by a circuit which is monotone on the input do the following: For every graph G let $a_G^1, \ldots, a_G^d$

denote the guessed subset incidence vector and $e_G^{i,j}$ the edges of G ($e_G^{i,j}=0$ means the edge is in the graph).

$$\bigvee_{k=0}^{d}\left[\sum_{i=1}^{d}\left(\bigwedge_{j=1}^{d}\neg a_G^j\vee e_G^{j,i}\right)\geq d-k+1\wedge\sum_{i=1}^{d}a_G^i\geq k\right]$$

tests whether for subset a_G of the left vertices in graph G the number of neighbors is smaller than its size. This witnesses that G has no perfect matching. The depth of the test is $O(\log d)$ due to the monotone formulae for the majority function given in [V84]. So together with an AND over all G we have depth $\log n+O(\log d)$.

Now to the lower bound. The (dual of the) bipartite perfect matching problem has a lower bound on the depth of deterministic monotone circuits of $\Omega(d)$ for graphs on $2d$ vertices. So we are left to prove that $s=\epsilon n/d$ guess bits do not help (for some $\epsilon>0$).

Given a monotone circuit F of depth t with s guess bits we construct a monotone deterministic circuit of depth $t+\epsilon d+O(\log n)$ for bipartite perfect matching on graphs of $2d$ vertices. This yields a lower bound $t=\Omega(d)$ (via [RW92], for some small $\epsilon>0$).

By fixing the guess bits of F we get 2^s deterministic circuits. At least one of these accepts $1/2^s$ of all 1-inputs. There must be a position in the direct sum where the accepted inputs vary over at least $1/2^{\epsilon d}$ of all bipartite graphs on $2d$ vertices without a perfect matching, otherwise less than $1/2^s$ of all combinations are possible and the circuit accepts too few inputs. In other words: for a fraction of $1/2^{\epsilon d}$ of all bipartite graphs on $2d$ vertices without perfect matching, there are n/d^2-1 other graphs such that the vector of these graphs is accepted. Since the circuit is monotone it is possible to use empty graphs instead of these unknown graphs. Fixing n/d^2-1 positions to empty graphs yields a monotone deterministic circuit accepting $1/2^{\epsilon d}$ of all graphs that must be accepted. Since a nondeterministic circuit works for restricted inputs as well we can iterate this and get $O(d^2 2^{\epsilon d})$ circuits that cover all ones. An OR tree has additional depth $\epsilon d+O(\log d)$. To compute the dual function switch ANDs and ORs. □

Theorem 7 *There are constants $c,\epsilon>0$ such that the following holds: Let d,n such that $n^{1/c}\geq d\geq\log n$. There is an explicit Boolean function h_n^d on n variables that can be computed by a monotone deterministic circuit of polynomial size in depth $O(d)$, and every nondeterministic monotone circuit with $\epsilon\frac{n\log^{3c}d}{d^{3c-1}}$ guess bits needs depth $\Omega(d)$. h_n^d can be computed in monotone depth $O(\log n)$ if $\frac{n\log^{2c}d}{d^{2c}}$ guess bits are allowed.*

PROOF SKETCH: In [RM97] a function on $m=l^3$ inputs is presented that can be computed by a deterministic monotone circuit of depth $\Theta(l^{1/c}\log l)$ (for some constant c) and polynomial size. Take $(n\log^{3c}d)/d^{3c}$ instances on $d^{3c}/\log^{3c}d$ input bits each. Like in theorem 6 it can be shown that the claimed number of guess bits does not help and depth $\Omega(d)$ is needed. On the other hand it can also be shown that l guess bits per instance suffice for depth $O(\log n)$. □

Theorem 8 *The depth of a monotone circuit for s,t-connectivity using $(n/k)\log n$ nondeterministic bits ($1\leq k\leq n$) is $\Omega(\log^2 k+\log n)$ and $O(\log n\log k)$.*

PROOF SKETCH: For the lower bound consider a nondeterministic monotone circuit for s,t-connectivity having access to $(n/k)\log n$ guess bits. The lower bound $\Omega(\log n)$ is trivial. If $\log^2 k=O(\log n)$ then the whole lower bound holds, else assume $\epsilon\log^2 k\geq\log n$ and we have $r=(n/k)\epsilon\log^2 k$ nondeterministic bits. Partition the set of vertices into n/k disjoint sets of size k each. By fixing vertices s_i,t_i for each set and including the edges t_i,s_{i+1} we get the direct sum of n/k instances of s,t-connectivity on graphs of k vertices, i.e., $s=s_1$ is connected to $t=t_{n/k}$ iff for all i there is a path from s_i to t_i.

Like in theorem 6 it can be shown that a circuit with r guess bits for the direct sum on n/k graphs with k vertices each has size $\Omega(\log^2 k)$ (using [KW90]).

For the upper bound observe the following. In depth $O(\log n\log k)$ one can compute for every vertex pair the predicate "has distance at most k" by (monotone) matrix multiplication. So guessing n/k vertices and testing if every consecutive pair of them has distance at most k and if s and t are among them suffices. □

Theorem 9 *For every $k\geq 2$ and $s\geq n$ there is a function q_k^s on $N=O(sn^k)$ inputs that can be computed by a deterministic monotone circuit with fan-in $O(s)$, depth $k+1$, and size $O(N)$. Any monotone circuit with depth k, unbounded fan-in, and s nondeterministic bits for q_k^s has size $2^{\Omega((N/s)^{1/k}/k)}$.*
q_k^s can be computed by a monotone circuit with unbounded fan-in, $O(s\log n)$ nondeterministic bits, depth k, and size $O(N)$ (with the exception of $k=2$, where the size is $O(N\log N)$).

PROOF SKETCH: The (unscaled) function q_k^s is defined by an alternating AND-OR-tree of depth $k+1$, where the top AND gate has fan-in s and all other gates have fan-in n. So there are sn^k inputs. This is the direct sum of s/n functions q_k defined by an alternating AND-OR-tree of depth $k+1$, where the top AND gate and all other gates have fan-in n.

The depth $k+1$ upper bound is immediate by definition. For the lower bound we can argue that a monotone depth k circuit with ϵs guess bits leads to a monotone deterministic depth k circuit with the same size t which accepts a fraction of $2^{\epsilon n}$ of all ones of q_k, and to an equivalent monotone unbounded fan-in formula with the same depth and size t^k. An OR over $O(2^{\epsilon n} n^k)$ such formulae gives us a monotone depth $k+1$ formula for q_k with a top OR gate and size $O(2^{\epsilon n} n^k \cdot t^k)$. This leads to the claimed size bound via [KPY84] or [NW93].
In the case of less limited nondeterminism guess strings $a_1, \ldots, a_s$ of $\log n$ bits each. $a_i = j$ indicates that the jth child $F_{i,j}$ of the ith gate among the children of the top gate in the defining circuit of q_k^s outputs 1. So it has to be checked whether for all i, j $a_i = j$ implies that this subcircuit $F_{i,j}$ outputs 1, i.e., $\bigwedge_{i,j} \neg(a_i = j) \vee F_{i,j}$. This leads to a circuit of depth $k+1$, but the term $\neg(a_i = j)$ can be ORed with all children of the $F_{i,j}$, which leads to depth $k-1$ and size $O(N)$, if the children are gates, otherwise ($k = 2$) the depth is k and the size is $O(N \log N)$). □

References

[BS90] R.B.Boppana, M.Sipser. The Complexity of Finite Functions. *Handbook of Theoretical Computer Science A*. Elsevier, 1990.

[CC97] L.Cai, J.Chen. On the amount of nondeterminism & the power of verifying. *SIAM Journ. Comput.*, vol.26, pp.733–750, 1997.

[CY91] J.Chen, C.K.Yap. Reversal Complexity. *SIAM Journ. Comput.*, vol.20, pp.622–638, 1991.

[DGS87] P.Duris, Z.Galil, G.Schnitger. Lower Bounds on Communication Complexity. *Information and Computation*, vol.73, pp.1–22, 1987.

[GLM96] J.Goldsmith, M.A.Levy, M.Mundhenk. Limited Nondeterminism. *SIGACT News*, vol.27(2), pp.20–29, 1996.

[GKW90] J.Goldstine, C.M.R.Kintala, D.Wotschke. On Measuring Nondeterminism in Regular Languages. *Information and Computation*, vol.86, pp.179–194, 1990.

[HR93] B.Halstenberg, R.Reischuk. Different Modes of Communication. *SIAM Journal Comput.*, vol.22, pp.913–934, 1993.

[Hr91] J.Hromkovič. Reversals-Space-Parallelism Tradeoffs for Language Recognition. *Math. Slovaca*, vol.2, pp. 121–136, 1991.

[Hr97] J.Hromkovič. Communication Complexity and Parallel Computing. Springer, 1997.

[HrS96] J.Hromkovič, G.Schnitger. Nondeterministic Communication with a Limited Number of Advice Bits. *Proc. 28th ACM Symp. on Theory of Comp.*, pp. 451–560, 1996.

[K88] M.Karchmer. Communication Complexity: A New Approach to Circuit Depth. Dissertation, 1988.

[KW90] M.Karchmer, A.Wigderson. Monotone Circuits for Connectivity Require Superlogarithmic Depth. *SIAM Journ. Discrete Math.*, vol.3, pp.255–265, 1990.

[Kl97] H.Klauck. On the Size of Probabilistic Formulae. *8th Int. Symp. on Algorithms and Computation*, pp.243–252, 1997.

[KPY84] M.Klawe, W.Paul, N.Pippenger, M.Yannakakis. On Monotone Formulae With Restricted Depth. *Proc. 16th ACM Symp. on Theory of Comp.*, pp. 480–487, 1984.

[KN96] E.Kushilevitz, N.Nisan. Communication Complexity. Cambridge University Press, 1996.

[N66] E.I.Nečiporuk. A Boolean function. *Sov. Math. Dokl.*, vol.7, pp.999-1000, 1966.

[Ne91] I.Newman. Private vs. Common Random Bits in Communication Complexity. *Information Processing Letters*, vol.39, pp.67–71, 1991.

[NW93] N.Nisan, A.Wigderson. Rounds in communication complexity revisited. *SIAM Journ. Comput.*, vol.22, pp.211-219, 1993.

[MR95] R.Motwani, P.Raghavan. Randomized Algorithms. Cambridge University Press, 1995.

[PST83] W.Paul, N.Pippenger, E.Szemerédi, W.Trotter. On Determinism Versus Non-Determinism and Related Problems. *24th Symp. Found. Comput. Science*, pp.429–438, 1983.

[RM97] R.Raz, P.McKenzie. Separation of the Monotone NC Hierarchy. *38th Symp. Found. Comput. Science*, pp.234–243, 1997.

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

[V84] L.G. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, vol.5, pp. 363–366, 1984.

[W87] I.Wegener. The Complexity of Boolean Functions. Wiley, 1987.

A Appendix: Proofs of Propositions

Proposition 1 *The s-nondeterministic Nečiporuk function is a lower bound on the size of nondeterministic Boolean formulae with s guess bits.*

PROOF: For a given partition of the inputs we show how a nondeterministic formula F can be simulated by the k communication games, such that the limited nondeterministic one-way communication of game i is bounded by the number of leaves in the subtree of F that contains exactly the variables belonging to B.
Given a nondeterministic formula the two players choose a deterministic formula with their common nondeterministic bits. Player A knows all inputs except those in y_i. Consider the formula subtree which has as leaves the variables in y_i of B and as root the top gate. Let V_i be the vertices of that tree having fan-in 2 and P_i the set of paths which start in V_i or at a leave and end in V_i or at the root without touching any other vertex in V_i. Obviously it suffices to let A send 2 bits for each path in P_i determining whether the uppermost gate of the path computes $0, 1, g$, or $\neg g$ for the function g computed on the lowest gate. Since there are at most $2|V_i| + 1$ such paths an overall communication of $4|V_i| + 2$ is sufficient. The set L_i of leaves labeled with variables from y_i has size $|V_i| + 1$, so we have $NC_s^{A,1}(f^i) \leq 4|V_i| + 2 < 4|L_i|$ and thus $1/4 \sum_i NC_s^{A,1}(f^i)$ is a lower bound on the size of nondeterministic formulae with s guess bits. □

Proposition 2 *Let z be a row of M' that appears t times in M'. Let the rows of M appearing at the same positions have δn difference positions. Then z has at most $2\binom{n}{s}s^s/2^{\delta s/9}$ ones.*

PROOF: Consider t rows of M with δn difference positions that agree with all ones of z. There is a set C of $\binom{n}{s}s^s$ columns/sets that are ones in the first of these rows, all other columns are invalid. Let $k = \delta s$. First we count the columns in C which intersect with at most $k/2$ of the sets U_i of the difference positions. A column in C is chosen by choosing s positions in the vector of sets and choosing an intersecting element for each position.
If s positions are chosen randomly then expected δs positions are difference positions. Thus by a Chernoff inequality with probability at most $2^{-\delta s/9}$ a random column/set in C does not intersect the U_i of at least $k/2$ difference positions. So these columns contribute at most $2^{-\delta s/9}\binom{n}{s}s^s$ ones.
Now consider the columns/sets in C which intersect the U_i of at least $k/2$ difference positions. How large is the probability for these to intersect with all set vectors on all s nonempty positions? On each difference position there are two set vectors which have different sets. The probability that one element out of s lies in the intersection of two sets that have distance $s/2$ is at most $1/2$. So the probability that a column set in C intersects with all sets in the vectors at its s nonempty positions is at most $(1/2)^{k/2}$ and so at most $\binom{n}{s}s^s/2^{k/2}$ of these ones are legal in the row.
Overall at most a fraction of $2^{-\delta s/9+1}$ of all ones can be valid. □

Proposition 4 *Let $W \subseteq Y$ be an event that does not contain an elementary event fixing path i. An input chosen randomly under the condition W reaches a leaf which is not very nice for i with probability at most $22k\sqrt{\epsilon} + k/n$.*

PROOF: First we prove that the first two items of being very nice hold in depth j in the protocol tree with probability at least $1 - j/n$. Consider the protocol tree. Restrict the tree to those inputs that satisfy W. So the new protocol tree has a root which is reached by all inputs that satisfy W. Call the matrix of inputs at a node z satisfying W $H_A^z \times H_B^z$, the uniformly distributed random variables h_A^z and h_B^z are defined on columns and rows. The restricted protocol tree induces a protocol, let the communication up to node z (measured in bits) in this protocol be called c'_z.
So at the root $I(h_A^z) = I(h_B^z) = 0 = 2c'_z$ and thus $I(h_A^z|y), I(h_B^z|y) \leq \log|Y| \leq 2c'_z + \log|Y|$. [Observe that

$$I(x) \leq I(x|y) \leq I(x) + H(y) \leq I(x) + \log|Y|$$

for all X and Y.]
Now we proceed downwards in the tree. Let w be a random child of a node z with $I(h_A^z), I(h_B^z) \leq 2c'_z$ (the child is chosen with the probability determined by the number of inputs arriving there). Assume A speaks at z and let a'_w be the length of the "message" leading to w, thus $c'_w = c'_z + a'_w$.
$Pr(I(h_B^w|y) > 2c'_w + \log|Y|) = 0$, because B sent nothing.
$Pr(I(h_A^w|y) > 2c'_w + \log|Y|) \leq 1/n$ as follows. Let μ denote the uniform distribution on inputs satisfying W. A child is chosen with probability $\mu(H_A^w)/\mu(H_A^z)$, and

$$\begin{aligned} I(h_A^w|y) &\leq I(h_A^w) + \log|Y| \\ &= -\log\mu(H_A^w) + \log|Y|. \end{aligned}$$

Thus

$$Pr(I(h_A^w|y) > 2c'_w + \log|Y|)$$

$$
\begin{aligned}
&\leq Pr(I(h_A^w) > 2c'_w) \\
&= Pr(\mu(H_A^w) < 2^{-2c'_w}) \\
&\leq Pr(\mu(H_A^w)/\mu(H_A^z) < 2^{-2a'_w}) \\
&\leq \frac{1}{n}\sum_w 2^{-a'_w} \leq \frac{1}{n}.
\end{aligned}
$$

This holds because $a'_w \geq \log n$, since the restricted protocol still has to communicate one edge of chain i. The last inequality follows from the Kraft inequality.

So the probability of reaching a vertex in depth j with $I(h_A^w|y) > 2c'_w + \log|Y|$ or $I(h_B^w|y) > 2c'_w + \log|Y|$ is at most j/n. But $I(f_A^w|y) \leq I(h_A^w|y) + \log|Y|$ for all w and $c_w \geq c'_w$ and thus the first two items of being very nice hold in depth j with probability at least $1 - j/n$ (obviously $|Y| \leq 2sk\log n$).

Now we prove that the root of the original protocol tree satisfies the third item of being very nice.

$$
I(f_A^z(v_0^i)|y) \leq \frac{(s-1)k}{n}\log n \leq \delta,
$$

where the first inequality holds because B starts the communication and thus $I(f_A^z(v_0^i)|y = a) \leq \log n$ if a contains the edge at v_0^i and 0 otherwise (B knows v_0^i). The second inequality holds for all s, k, n for which a nonconstant lower bound is claimed.

We continue by proving that a random child w, chosen with probability proportional to the number of inputs satisfying W which reach that child, is not very nice for i with probability $22\sqrt{\epsilon}$, given that the father z is very nice and that the son satisfies the first two items of being very nice. This implies the proposition. First a useful fact from [NW93], which allows to bound the probability of events under a distribution with small information by their uniform probability.

Fact 8 *For $W \subseteq X$ let q be the uniform probability of W. If*

$$
\Delta = \sqrt{\frac{4I(x)}{q}} \leq \frac{1}{10},
$$

then $|Pr(W) - q| \leq q\Delta$.

Assume A speaks at z, and let a_w be the length of the message leading from z to w, thus $c_w = c_z + a_w$. We have $I(f_A^w|y) \leq 2c_w + 4sk\log n \leq \epsilon n$. But then $\sum_{v\in V_A} I(f_A^w(v)|y) \leq I(f_A^w|y) \leq \epsilon n$. So if a v were chosen uniformly, then $Pr_U(I(f_A^w(v)|y) > \delta) \leq \epsilon/\delta$ by the Markov inequality. But $v_t = f_B^z(v_{t-1})$ and thus $I(v_t) = I(f_B^z(v_{t-1}) \leq I(f_B^z(v_{t-1})|y) \leq \delta$ and with fact 8:

$$
\begin{aligned}
&Pr(I(f_A^w(v_t)|y) > \delta) \\
&\leq \epsilon/\delta\left(1 + \sqrt{\frac{4\delta}{\epsilon/\delta}}\right) \\
&= 22\sqrt{\epsilon}.
\end{aligned}
$$

□

Proposition 7 $NC(f_k^s) = \Omega(sk\log(n/(sk)))$.

PROOF SKETCH: We will show a reduction from the following problem:

$$
p(x_1, \ldots, x_s, y_1, \ldots, y_s,) = \bigoplus_{i=1}^{s}(x_i = y_i)
$$

on strings $x_i, y_i \in \{0,1\}^{\Theta(k\log(n/(sk)))}$. This problem has nondeterministic communication complexity $\Omega(sk\log(n/(sk)))$, which can be proved using the method of the largest 1-chromatic submatrix.

First partition the vertices of each side into s sets of n/s vertices. A *monotone path* in segment i is a path which alternates between left vertices and right vertices from the ith set, such that the path begins at the first vertex on the left and ends at a vertex with PARITY 1. The path must have a sequence of vertex numbers which is nondecreasing.

Let the set of monotone paths of length k in segment i be called P_i. Clearly $|P_i| = \Omega(\binom{n/s}{k})$. Now identify the left and the right side of a monotone path with each other and with a string in $\{0,1\}^{\log|P_i|}$. This is sensible, because the left and the right side of different monotone paths do not fit together: the different paths may be parallel up some point. Then one path goes to a higher numbered vertex than the other. Here the path of one left/right combination breaks off, but the other combination breaks off later because there are too few edges on one side.

Consider the set of inputs that correspond to all s-tuples of monotone paths of length k, where all other edges point to vertex 0 (in segment 0). We can identify each vector of paths with a vector of strings and get the claimed reduction: each incorrectly matched path contributes a 0, each correctly matched path contributes a 1. □