Branching Programs: Complexity Lower Bounds by Communication Games

Martin Sauerhoff, Universität Dortmund

Circuits, Logic & Games 2006

Overview

1. Introduction
2. Some Remarks about
   Lower Bound Techniques
3. New Result:
   Lower Bound for Rand. Read-k BPs
   via Information Complexity.

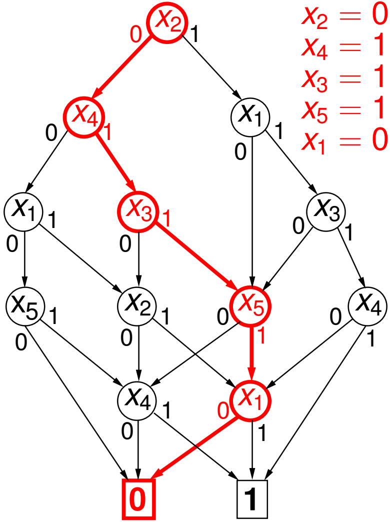
"Master Plan" of Branching Program Theory:

- **Ultimate goal:**
  Proving superlogarithmic space lower bounds, separating L from something interesting (NP, P).
- **On the way:**
  Such lower bounds for less and less restricted variants of branching programs.

Celebrated results of Ajtai / Beame et al.:

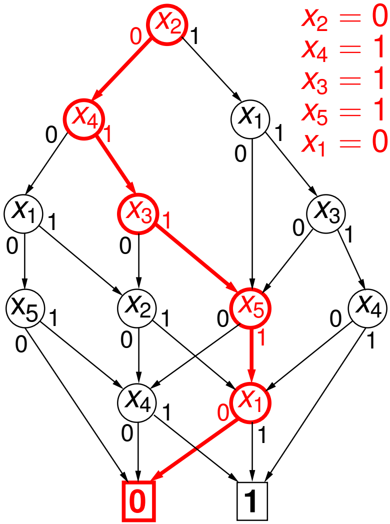
Superlinear time-space tradeoff lower bounds for computation of boolean functions / languages on (nonuniform) RAMs (with arbitrary instruction set).

Branching Programs (BPs)



$x_2 = 0$
$x_4 = 1$
$x_3 = 1$
$x_5 = 1$
$x_1 = 0$

Branching Programs (BPs)



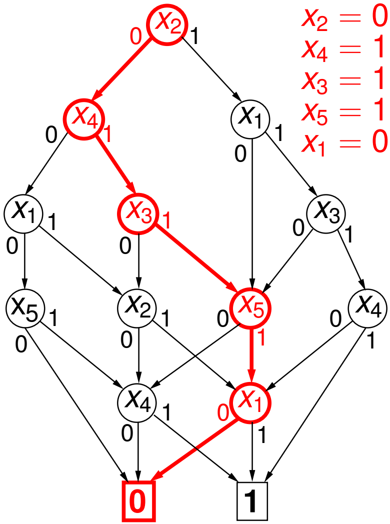
BP space:

log(# nodes in the graph)

BP time:

maximum # edges
on a computation path

Branching Programs (BPs)



$x_2 = 0$
$x_4 = 1$
$x_3 = 1$
$x_5 = 1$
$x_1 = 0$

BP space:

log(# nodes in the graph)

BP time:

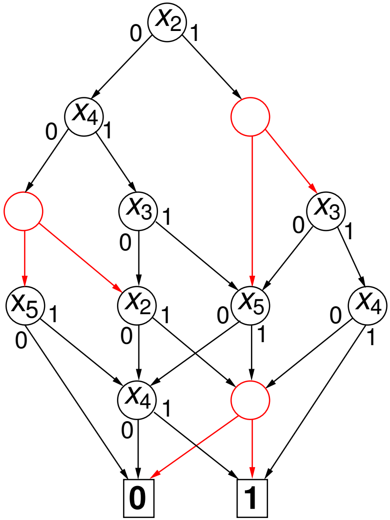
maximum # edges
on a computation path

Key Facts:

BP space $\approx$ RAM space;

BP time $\leq$ RAM time.

Branching Programs (BPs)



Nondeterministic / Randomized variant:

allow unlabeled nodes, choose successor nondeterministically / by (random) coin toss

Here: Always two-sided error bounded by constant smaller than $1/2$.

Zoo of Restricted Variants (A Quick Tour)

- **OBDDs (ordered binary decision diagrams):**
  Read each variable at most once, in a fixed order.
  Data structure, model for data stream algorithms.
- **Read-*k* BPs:**
  Each variable at most k times on each path.
  In particular: Time $\leq kn$, $n = \#$ variables.
- **Time-restricted BPs:**
  Time kn, $k = k(n)$ "small."

  Space lower bound under time restriction $\leftrightarrow$
  Time lower bound under space restriction.
  Time as a function of space $\rightarrow$ "tradeoff lower bound."

Results for Unrestricted BPs / RAMs

Current frontier of what is achievable
with proof techniques:

- **Beame, Jayram, Saks ('98), Ajtai ('99a, '99b), Beame, Saks, Sun, Vee ('00):**

  First superlinear time-space tradeoff lower bounds.
  Largest bound for boolean inputs: $T = \Omega\left(n\sqrt{\frac{\log(n/S)}{\log\log(n/S)}}\right)$.

- **Beame, Vee ('02):**

  So far largest time lower bound for sublinear space:
  $T = \Omega(n \log^2 n)$,
  RAMs with registers of bitlength poly(n).

- **S., Woelfel ('03):**

  Middle-bit of integer multiplication: $T = \Omega(n \log(n/S))$,
  RAMs with registers of bitlength $\Theta(\log n)$.

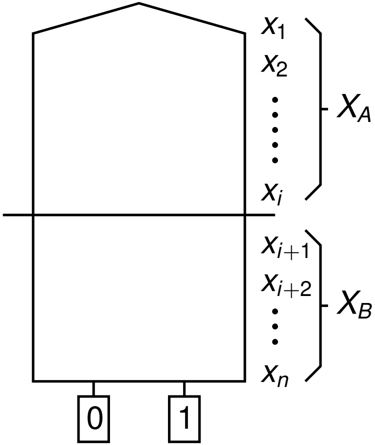
Very rough sketch of the basic idea:

- Cut computations along time axis.
  Computations short $\rightarrow$ only few pieces.
- Show that total amount of information transported over all boundaries has to be large for computing desired function.

  Averaging (few pieces) yields single boundary
  with much information.

  $\Rightarrow$ Need many configurations, i. e. much space.

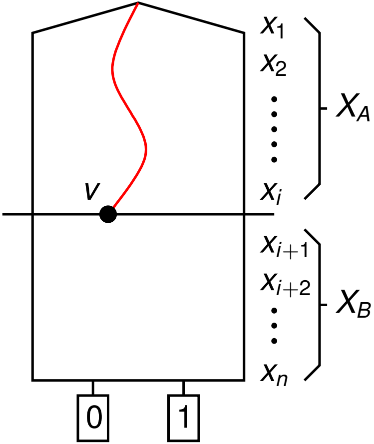
Simple, Prototypic Case: OBDDs



Given OBDD G.

- Cut computation paths $\rightarrow$ partition (X_A, X_B) of vars.
- Communication protocol wrt. (X_A, X_B): Alice / Bob follow partial computation paths in upper / lower part.

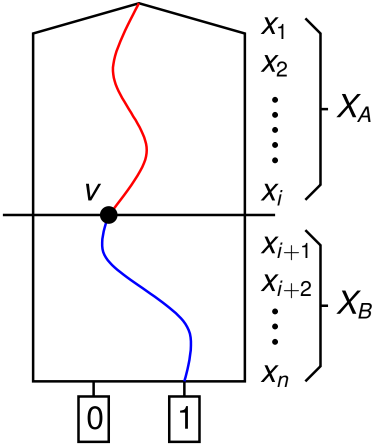
Simple, Prototypic Case: OBDDs



Given OBDD G.

- Cut computation paths $\to$ partition (X_A, X_B) of vars.
- Communication protocol w.r.t. (X_A, X_B): Alice / Bob follow partial computation paths in upper / lower part.

Simple, Prototypic Case: OBDDs



Given OBDD G.

- Cut computation paths $\rightarrow$ partition (X_A, X_B) of vars.
- Communication protocol w.r.t. (X_A, X_B): Alice / Bob follow partial computation paths in upper / lower part.
- $\#$ comm. bits $\leq$ log(# nodes).

On the Way to More General Models (1/2)

First problem:

How do we find sets of variables for the two players?

E. g., variable sequence on computation path:

$x_3 \quad x_6 \quad x_5 \quad x_3 \quad x_1 \quad x_4 \quad x_1 \quad x_6 \quad x_2 \quad x_3 \quad x_1 \quad x_4$

On the Way to More General Models (1/2)

First problem:

How do we find sets of variables for the two players?

E. g., variable sequence on computation path:

$X_A = \{x_1, x_2, x_3\}$, $X_B = \{x_4, x_5, x_6\} \rightarrow$

$x_3 \quad x_6 \quad x_5 \quad x_3 \quad x_1 \quad x_4 \quad x_1 \quad x_6 \quad x_2 \quad x_3 \quad x_1 \quad x_4$:−(

On the Way to More General Models (1/2)

First problem:

How do we find sets of variables for the two players?

E. g., variable sequence on computation path:

$X_A = \{x_1, x_2, x_3\}$, $X_B = \{x_4, x_5, x_6\} \rightarrow$

$x_3 \quad x_6 \quad x_5 \quad x_3 \quad x_1 \quad x_4 \quad x_1 \quad x_6 \quad x_2 \quad x_3 \quad x_1 \quad x_4$:−(

Idea: Alice / Bob obtain disjoint subsets X_A, X_B of set X of all variables.

Want:

- $|X_A|, |X_B|$ "large," ideally: $|X_A|, |X_B| \geq \alpha n$, constant α.
- Not "too many" changes between Alice / Bob variables: Each requires a cut in the computation, bad for averaging.

On the Way to More General Models (2/2)

Second problem:

In general: Different variable sequences
on different comp. paths.

- No longer global variable partition:
  Communication complexity $\rightarrow$ rectangle bounds.
- Group inputs together according to read access patterns to get rectangles (sophisticated averaging arguments).

On the Way to More General Models (2/2)

Second problem:

In general: Different variable sequences
on different comp. paths.

- No longer global variable partition:
  Communication complexity $\rightarrow$ rectangle bounds.
- Group inputs together according to read access patterns
  to get rectangles (sophisticated averaging arguments).

Third problem:

Have to get rid of $X - X_A - X_B$ somehow (useless).

- Models with same variable sequence on each comp. path:
  Set to constants. Still requires suitable function.
- In general: Difficult. Again clever averaging techniques.

Limitations on Achievable Lower Bounds (1/2)

Why we do not get larger time lower bounds:

For read-k / time kn BPs, we get X_A, X_B with:
$|X_A|, |X_B| \geq n \cdot 2^{-(k+1)}$, also essentially optimal.
Thus technique only works for $k \leq \log n$.

Best possible time lower bounds for sublinear space:
$T = kn = \Omega(n \log n)$.

Problem: Two-party communication setting.

Limitations on Achievable Lower Bounds (2/2)

Possible remedy: Multiparty communication setting.

Consider p player, "number-on-the-forehead" model.
Then: $2^{-k} \rightarrow (1 - 1/p)^k \approx 2^{-k/p}$.

But: Multiparty results only for $p \leq \log n$ so far.
Best possible time lower bounds for sublinear space:
$T = \Omega(n \log^2 n)$. (Beame, Vee '02)

Large gap to best possible tradeoff lower bounds
$TS = \Omega(n^2)$!

Goal now:

Lower bounds on randomized BPs for problems that are nondeterministically easy.

Known results for models "above read-once" need tiny error probabilities for randomized BPs.
For time-restricted BPs: Even have to tend to 0 in n.

Proving better results:
Forced to improve lower bound techniques.

Randomness vs. Nondeterminism for Read-k BPs

Best separation so far (S. '03):

- Function: Test whether n-vertex graph contains a triangle (3-clique), $\mathrm{cl}_{3,n}$.
- Easy even for nondeterministic OBDDs: Guess and verify, $\Theta(n^3)$ nodes.
- Randomized read-k BPs for $\mathrm{cl}_{3,n}$ with error bounded by $2^{-\Omega(2^{-2k})}$ require space $\Omega(k^{-2}2^{-4k}\sqrt{n})$.

Here: Bound on the error $\rightarrow 2^{-\Omega(k)}$. :–)

Tool: Information Complexity (1/2)

Recent technique for proving lower bounds on randomized / quantum communication complexity in various settings.

Most important reference here:
Bar-Yossef, Jayram, Kumar, Sivakumar ('02).

- Lower bounds for randomized two-party communication protocols.
- New proof of $\Omega(n)$ bound for set-nondisjointness, i. e., OR of n ANDs on disjoint pairs of variables:

  Via direct sum property and lower bound on information complexity of AND of two bits.

  (Yes, $\mathrm{AND}(x, y) = x \wedge y$.)

Tool: Information Complexity (2/2)

Outline of the approach:

- Consider randomized two-party protocol P on suitable input random variables (X, Y).
- Information complexity of P wrt. (X, Y):
  Information that messages of P reveal about (X, Y).
- **Easy to see:**
  Information complexity is a lower bound on communication complexity.
- **Show:**
  P has small error probability for given function $\Rightarrow$ Information complexity large.

Proof of the Main Result – Ingredients

1. Known lower bound on information complexity for AND.
2. Randomized multipartition protocols.
3. Information complexity lower bound for randomized multipartition protocols computing AND.
4. Application to clique function.

Information Complexity Lower Bound for AND

Result of Bar-Yossef et al.:

- Input distribution: $Z \in \{0,1\}^2$ defined as follows:
  - Choose $D \in \{1,2\}$ randomly.
  - If $D = 1$: $Z = (0, Y)$, Y random bit.
  - If $D = 2$: $Z = (X, 0)$, X random bit.
- Rand. bounded error protocol for AND:
  Information complexity of P wrt. Z is $\Omega(1)$.
- More "low level" version:
  Information complexity lower bounded by
  Hellinger distance of messages distributions
  for inputs $(0,0)$ and $(1,1)$.

  In turn lower bounded by constant for protocols
  computing AND with small error.

Randomized Multipartition Protocols

Definition:

- Distribution $(q_1, \ldots, q_k)$ over collection of randomized two-party communication (sub-)protocols $P_1, \ldots, P_k$.
- Each protocol may have its own input partition.
- Probability of output r on z: $\sum_{i=1}^{k} q_i \cdot \Pr\{P_i(z) = r\}$.
- Complexity: $\log\left(\sum_{i=1}^{k} 2^{c_i}\right)$, where $c_1, \ldots, c_k$ complexities of the subprotocols.

For AND(x, y):

Only possible partitions are $(\{x\}, \{y\})$, $(\{y\}, \{x\})$.

Goal:

Information complexity lower bound for AND in this model.

New Information Complexity Bound

Large steps:

- Info complexity of multipartition protocol lower bounded by avg. info complexity of subprotocols.
- Apply "low level" form of known lower bound for single partition to subprotocols: Information complexity of AND lower bounded by Hellinger distance of message distributions for inputs $(0, 0)$ and $(1, 1)$.
- Hellinger distance $\rightarrow$ L_1-norm.
- Lower bound in terms of avg. L_1-distance of message distributions for $(0, 0)$ and $(1, 1)$ induced by subprotocols.
- Δ-inequality: Same for L_1-distance of complete multipartition protocol. Must be large due to small error.

Application to 3-Clique Function

Sketch:

- Setting variables to constants, simultaneously turn...
  - clique function $\mathrm{cl}_{3,n}$ into disjointness problem;
  - rand. read-k BP into randomized multipartition protocol for this disjointness problem.

  Actually more complicated than described model:
  Each AND may have several partitions where Alice / Bob has both variables.
- Combinatorial result from earlier paper (S. '03) $\rightarrow$
  Can ensure that for each AND there is at least a $2^{-\Theta(k)}$-fraction of all partitions where the variables are split.
  Due to the very small error bound, also the error on these must be small.
- BP space lower bounded in complexity of the obtained protocol, $S = \Omega(k^{-2} 2^{-5k} \sqrt{n})$. $\square$