Conspiracies between Learning Algorithms, Lower Bounds, and Pseudorandomness

Igor Carboni Oliveira
University of Oxford

Joint work with **Rahul Santhanam** (Oxford)

Context

Minor algorithmic improvements imply lower bounds (Williams, 2010).

NEXP not contained in $\mathbf{ACC^0}$ (Williams, 2011), and extensions.

This Work

Analogue of Williams' celebrated lower bound program in **Learning Theory**.

Combining and extending existing connections.

Further applications of the "**Pseudorandom Method**":

Hardness of **MCSP**,
Karp-Lipton Theorems for **BPEXP**.
etc.

Lower bounds from learning

Learning Model (Randomized, MQs, Uniform Dist.)

A Boolean circuit class **C** is fixed.

$f : \{0,1\}^n \to \{0,1\}$ from **C**[**s**(**n**)] is selected.

$$\boxed{\textbf{Randomized Algorithm}} \; \begin{array}{c} \xrightarrow{\quad a \quad} \\ \xleftarrow[\quad f(a) \quad]{} \end{array} \; f$$

Learner must output w.h.p a hypothesis h such that:

$$\Pr_{x \in \{0,1\}^n}[h(x) = f(x)] \geq 1 - 1/n.$$

Some learning algorithms

$$\overbrace{\mathsf{DNF} \subsetneq \mathsf{AC}^0 \subsetneq \mathsf{AC}^0[p] \subsetneq}^{\text{Combinatorial lower bounds}} \overbrace{\mathsf{ACC}^0 \subseteq \mathsf{TC}^0 \subseteq \mathsf{Formula}[\mathsf{poly}] \subseteq \mathsf{Circuit}[\mathsf{poly}]}^{\text{Lower bounds are unknown, or obtained via diagonalization}}.$$

[Jac97] **DNFs** can be learned in **polynomial** time.

Harmonic-Sieve/Boosting

[LMN93] **AC⁰** circuits learnable in **quasi-polynomial** time.

Fourier Concentration

[CIKK16] **AC⁰[p]** learnable in **quasi-polynomial** time.

Pseudorandomness/Natural Property

Can we learn **AC^0** circuits with **Mod 6** gates in **sub-exponential time**?

As far as I know, open even for:

AND o OR o MAJ circuits, **MOD_2 o AND o THR** circuits.

Definition. **Non-trivial learning algorithm:**

- Runs in randomized time $\leq \dfrac{2^n}{n^{\omega(1)}}$.
- For every function **f** in **C**:

$$\Pr_{x \in \{0,1\}^n}[h(x) = f(x)] \geq \frac{1}{2} + \frac{1}{n}.$$

Non-trivial learning implies lower bounds

Let **BPE** = **BPTIME**[$2^{O(n)}$].

Theorem. Let **C** be any subclass of Boolean circuits closed under restrictions.

Example: C = (depth-6)-ACC^0, AND o OR o THR, etc.

If for each k>1, **C**[n^k] admits a non-trivial learning algorithm, then for each k > 1, **BPE** is not contained in **C**[n^k].

LBs from Proofs, Derandomization, Learning

	Non-trivial **SAT/Proof System**	Non-trivial **Derandomization**	Non-trivial **Deterministic Exact Learning**	Non-trivial **Randomized Learning**
Assumption	Proofs checked in **deterministic** time $2^n/n^{\omega(1)}$	Algorithm runs in **deterministic** time $2^n/n^{\omega(1)}$	Learner runs in **deterministic** time $< 2^n$	Learner runs in **randomized** time $2^n/n^{\omega(1)}$
Consequence	LBs for **NEXP**	LBs for **NEXP**	LBs for **EXP**	LBs for **BPEXP**
Reference	[**Wil10**]	[**Wil10**], [**SW13**]	[**KKO13**]	[**This Work**]

Remarks on lower bounds from Learning

- Learning approach won't directly work for classes containing PRFs.

- Conceivable that one can design non-trivial learning algorithms for a class **C** under the assumption that **BPEXP** is contained in **P/poly**.

- Learning connection applies to virtually any circuit class of interest, and there is **no depth blow-up**.

It can lead to new lower bounds for restricted classes such as **THR o THR** and **$\mathbf{ACC^0}$**.

Previous work on learning vs. lower bounds

- Systematic investigation initiated about 10 years ago:

[FK06] Lower bounds for **BPEXP** from **polynomial time learnability**.

[HH11] Lower bounds for **EXP** from **deterministic exact learning**.

[KKO13] Optimal lower bounds for **EXP** from **deterministic exact learning**.

[Vol14] Lower bounds for **BPP/1** from polynomial time learnability.

[Vol'15] Further results for **learning arithmetic circuits**.

A Challenge in Getting Lower Bounds from Randomized Learning

Williams' lower bounds from non-trivial SAT algorithms: a **non-trivial algorithm** can be used to violate a **tight hierarchy theorem** for **NTIME**.

Challenge in **Randomized Learning**:
lack of strong hierarchy theorems for **BPTIME.**

The approach has to be indirect, and we must do something different …

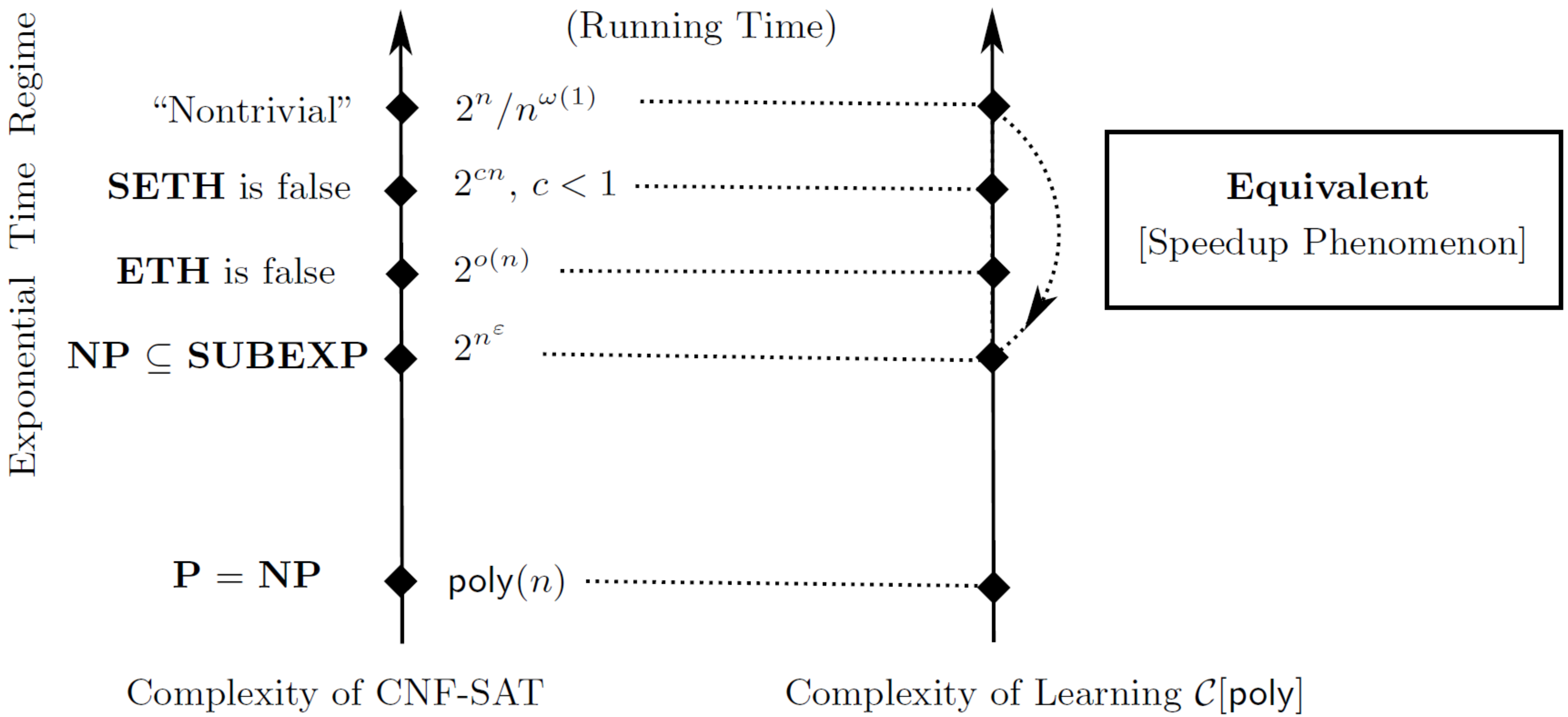
Speedup Phenomenon in Learning Theory

Speedup Lemma. Let $\mathcal{C}$ be any class of Boolean circuits containing $AC^0[2]$.

Suppose that for each $k \geq 1$ the class $\mathcal{C}(n^k)$ admits **a non-trivial learning** algorithm.

Then for each $k \geq 1$ and $\varepsilon > 0$, the class $\mathcal{C}(n^k)$ is **strongly learnable** in time $O(2^{n^{\varepsilon}})$.

SAT Algorithms vs. Learning Algorithms



Main Techniques: "Speedup Lemma"

Non-trivial Learner + NW-Generator + Hardness Amplification $\Rightarrow$ Faster Learner!

1. Given oracle access to $f : \{0,1\}^n \to \{0,1\}$ in **C**[**poly**], implicitly construct a "**pseudorandom**" **ensemble** of functions in **C**[**poly**] on **n**$^{\boldsymbol{\delta}}$ bits.

(using NW-generator + Hardness Amplification [**CIKK16**])

Intuition: **Non-trivial learner** can **distinguish** this ensemble from random functions. This can be done in time $2^{O(n^{\delta})}$.

2. This distinguisher (**i.e. the non-trivial learner**) and the reconstruction procedures of **NW-generator** and **Hardness Amplification** can be used to strongly learn f in time $2^{O(n^{\delta})}$.

Main Techniques: "LBs from Learning"

1. Starting from **non-trivial learner**, apply the **Speedup Lemma** to obtain a **sub-exponential time learner**.

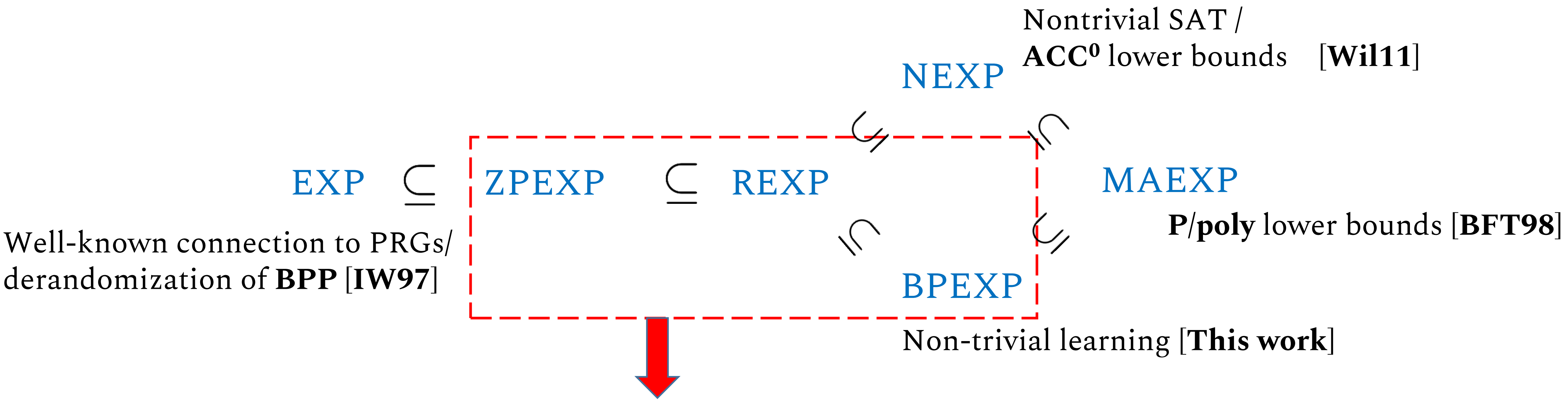
$$2^n/n^{\omega(1)} \xrightarrow{\textbf{Speedup}} 2^{O(n^\delta)}$$

$$2^{O(n^\delta)} \Downarrow \mathsf{BPE} \not\subseteq \mathsf{C}[n^k]$$

2. Adapting the techniques from [**KKO13**], **randomized sub-exponential** time learnability of **C[poly]** implies **BPE lower bounds** against **C[n^k]**.

3. Using an additional **win-win argument**, this holds under **minimal assumptions** on C, and **with no blow-up in the reduction.**

Combining and extending existing connections



[**OS17**] Connections to **pseudo-deterministic** algorithms.

- Further motivation for the following question:

Which algorithmic **upper bounds** imply **lower bounds** for **ZPEXP** and **REXP**, respectively?

One-sided error: Lower bounds for REXP

We combine the satisfiability and learning connections to lower bounds to show:

[**Informal**]
If a circuit class **C** admits both **non-trivial SAT** and **non-trivial Learning** then **REXP** is not contained in **C**.

Corollary. [ACC⁰ lower bounds from non-trivial learning]
If for every depth **d>1** and modulo **m>1** there is $\varepsilon > 0$ such that $\mathsf{ACC}^0_{d,m}(2^{n^\varepsilon})$ has non-trivial learning algorithms, then $\mathsf{REXP} \not\subseteq \mathsf{ACC}^0(\mathsf{poly}(n))$.

Indicates that combining the two frameworks might have further benefits.

Zero-error: Lower bounds for ZPEXP

[**IKW02**], [**Wil13**] Connections between **natural properties without density condition**, **Satisfiability Algorithms**, and **NEXP** lower bounds.

[**CIKK16**] Connections between **BPP-natural properties** and **Learning Algorithms.**

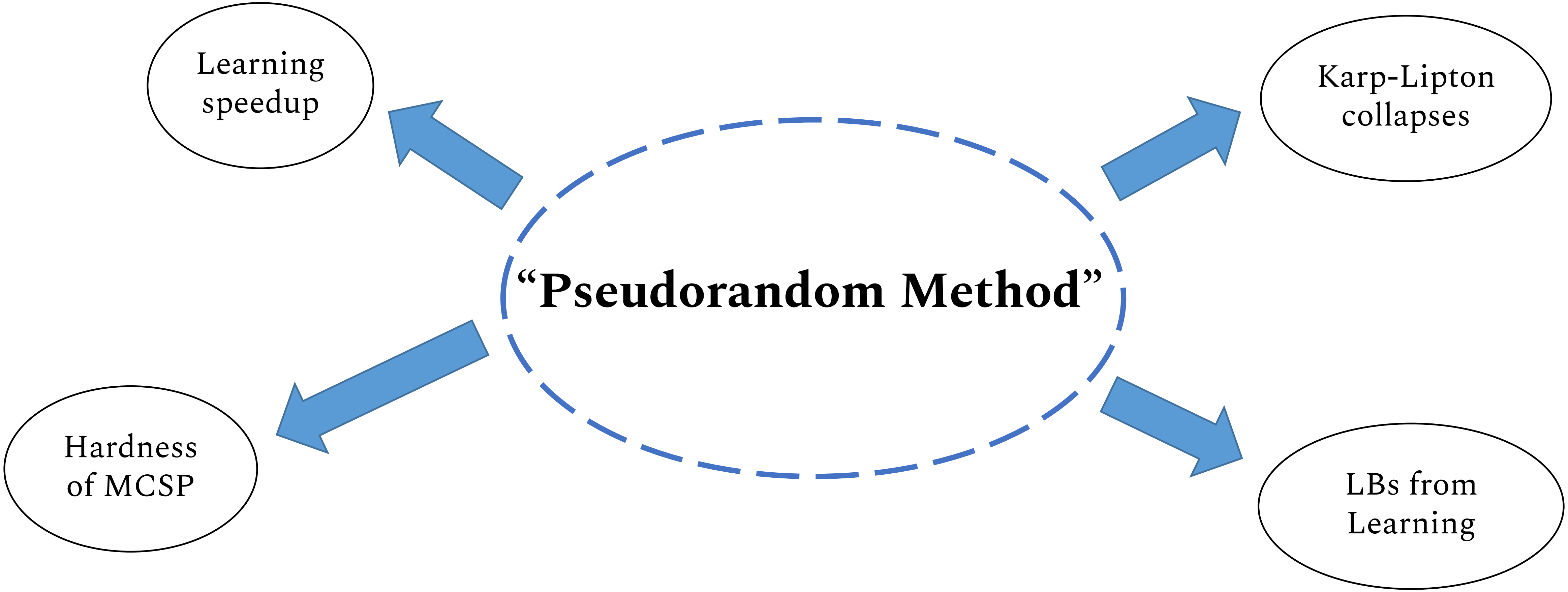
We give a new connection between **P-natural properties** and **ZPEXP** lower bounds.

Let $\mathcal{C}(\mathsf{poly}) \subseteq \mathsf{P}/\mathsf{poly}$ be a circuit class closed under restrictions.

Theorem. [ZPEXP lower bounds from natural properties]
If for some $\delta > 0$ there are P-natural properties against $\mathcal{C}(2^{n^\delta})$ then $\mathsf{ZPEXP} \not\subseteq \mathcal{C}(\mathsf{poly}(n))$.

Further Applications of our Techniques

A rich web of techniques and connections



Use of (conditional) **PRGs** and related tools, often in contexts where (**pseudo**)**randomness** is not intrinsic.

Karp-Lipton Collapses

Connection between **uniform** class and **non-uniform** circuit class:

[**KL80**] If $\mathsf{NP} \subseteq \mathsf{P}/\mathsf{poly}$ then $\mathsf{PH} = \Sigma_2^p \cap \Pi_2^p$.

Assumption	Consequence	Major Application
EXP in **P/poly**	**EXP** = **MA** [**BFT98**]	**MA$_{\mathbf{EXP}}$** not in **P/poly** [**BFT98**]
NEXP in **P/poly**	**NEXP** = **EXP** [**IKW02**]	**SAT / LB** Connection [**Wil10**]
Randomized Exponential Classes such as **BPEXP** ?		

Karp-Lipton for randomized classes

Theorem 1. If $\mathsf{BPE} \subseteq \mathsf{i.o.SIZE}[n^k]$ then $\mathsf{BPEXP} \subseteq \mathsf{i.o.EXP}/O(\log n)$.

The advice is needed for technical reasons. But it can be eliminated in some cases:

Theorem 2. If $\mathsf{BPE} \subseteq \mathsf{i.o.SIZE}[n^k]$ then $\mathsf{REXP} \subseteq \mathsf{i.o.EXP}$.

- Check paper for Karp-Lipton collapses for **ZPEXP**, and related results.

Hardness of MCSP

Minimum Circuit Size Problem:

Given $\mathbf{1^s}$ and a Boolean function $f : \{0,1\}^n \to \{0,1\}$ represented as an N-bit string,

Is it computed by a circuit of size at most s?

Recent work on MCSP and its variants: [**KC00**], [**ABK+06**], [**AHM+08**], [**KS08**], [**AD14**], [**HP15**], [**AHK15**], [**MW15**], [**HP15**], [**AGM15**], [**HW16**].

[**ABK+06**] **MCSP** is not in $\mathbf{AC^0}$.

Open. Prove that **MCSP** is not in $\mathbf{AC^0[2]}$!

Our result

We prove the first hardness result for **MCSP** for a standard complexity class beyond $\mathbf{AC}^0$:

Theorem. If **MCSP** is in $\mathbf{TC}^0$ then $\mathbf{NC}^1$ collapses to $\mathbf{TC}^0$.

The argument describes a non-uniform $\mathbf{TC}^0$ reduction from $\mathbf{NC}^1$ to **MCSP** via **pseudorandomness**.

Additional applications of our techniques

- **Equivalences** between **truth-table compression** [**CKK+14**] and **randomized learning models** in the sub-exponential time regime.

    (For instance, **equivalence queries** can be eliminated in sub-exp time randomized learning of expressive concept classes.)

- A **dichotomy** between **Learnability** and **Pseudorandomness** in the non-uniform exponential-security setting:

    “A circuit class is either **learnable** or contains **pseudorandom functions**, but not both.”

    In other words, **learnability is the only obstruction to pseudorandomness.**

(Morally, $\mathbf{ACC^0}$ is either learnable in sub-exp time or contains exp-secure PRFs.)

Problems and Directions

Is there a **speedup phenomenon** for complex classes (say $AC^0[p]$ and above) for

learning under the uniform distribution with **random examples**?

Can we establish **new** lower bounds for modest circuit classes by designing non-trivial learning algorithms?

Towards lower bounds against NC?

Non-trivial learning implies lower bounds:

First example of lower bound connection from **non-trivial randomized algorithms**.

Problem. Establish a connection between **non-trivial randomized SAT** algorithms and **lower bounds**.

(First step in a program to obtain unconditional lower bounds against **NC**.)

Thank you