**Theoretical Computer Science Notes** **April 9, 2017**

*Natural Proof*

**Chi-Ning Chou**

# 1 Overview

One of the major goal in theoretical CS is to separate complexity classes while one of the most exciting approach is trough *circuit lower bound*. As long as one can prove $\mathbf{NP} \not\subseteq \mathbf{P/poly}$, then we get $\mathbf{P} \neq \mathbf{NP}$. However, life is not so easy, after decades of trials, people still cannot prove something much more weaker such as $\mathbf{NEXP} \not\subseteq \mathbf{P/poly}$! To the best of our knowledge, the state-of-the-art unconditinoal circuit lower bound is by Williams [Wil14] showing that $\mathbf{NEXP} \not\subseteq \mathbf{ACC^0}$. See Figure 1.

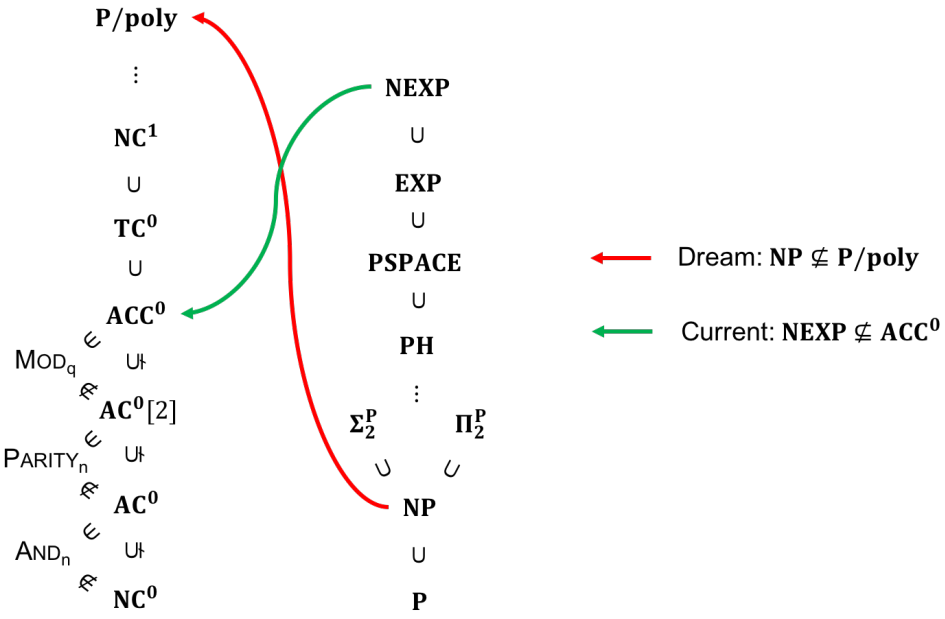


Figure 1: The map of circuit lower bound.

## 1.1 History

In 1980s. there were several exciting progress in proving circuit lower bound. See the corresponding position in Figure 1.

- In 1984, Furst, Saxes, and Sipser showed that PARITY is not in $\mathbf{AC^0}$.
- In 1987, Razborov showed that $MOD_p$ is not in $\mathbf{AC^0}[q]$ for distinct primes $p$ and $q$.
- In 1987, Andreev showed a nearly quadratic lower bound for general formula.

However, people soon found out that it is very difficult to extend the techniques used in these results to higher circuit complexity classes.

As a result, a *natural* question arose:

**Question 1** *Is there any barrier in proving circuit lower bound?*

### 1.2 The limitation of proof techniques

Razborov and Rudich [RR97] answered Question 1 by showing that current (in that day) circuit lower bound techniques all fell into the so called *natural proof barrier*. That is, one cannot hope for using the same techniques to prove lower bound for higher circuit complexity class as long as certain well-believed cryptographic assumption holds.

- If we have a natural proof against circuit class $\mathcal{C}$, then $\mathcal{C}$ cannot contain very strong families of pseudorandom functions.
  - Natural proof against $\mathcal{C}$ implies exists efficient algorithm $A$ that given random boolean function $f$, $A$ can certify $f \notin \mathcal{C}$ with non-negligible probability.
  - Efficient algorithm can never certify $f \notin \mathcal{C}$ if $f$ is a pseudorandom function and computable in $\mathcal{C}$.
- If we use natural proof to show something is harder and harder, then we would simultaneously show that those problems to be easier and easier.
  - Any natural proof showing that these problems took at least $t(n)$ time, would also show that they took at most roughly $2^{t^{-1}(n)}$ time.
  - Thus, no natural proof can show lower bound more than half-exponential.

## 2 Natural proof barrier

### 2.1 Some definitions

Let's start with the definition of natural proof. In the following, we assume some background in basic circuit complexity classes. We say $\mathcal{C}$ is a typical circuit family class if $\mathcal{C} \in \{\mathbf{AC^0}, \mathbf{ACC^0}, \mathbf{TC^0}, \mathbf{NC^1}, \mathbf{P/poly}\}$. Also, we denote $\mathcal{C}/s(n)$ to be the circuit complexity class of size $s(n)$ with the corresponding gates. If not specified, $\mathcal{C} = \mathcal{C}/\mathrm{poly}(n)$.

Let $\mathcal{F}_n := \{f : \{0,1\}^n \to \{0,1\}\}$ to be the set of all boolean function of input length $n$. We denote a family of function by using bold face character such as $\mathbf{f} = \{f_n\}_{n\in\mathbb{N}}$ where $f_n \in \mathcal{F}_n$ for each $n \in \mathbb{N}$.

A property $\mathcal{P} = \{\mathcal{P}_n\}_{n\in\mathbb{N}}$ is a family of collection of functions where $\mathcal{P}_n \subseteq \mathcal{F}_n$ for each $n \in \mathbb{N}$. Semantically, $\mathcal{P}$ contains the functions that satisfy the *property*.

It's not difficult to see that to prove that a typical circuit class $\mathcal{C}$ does not contain a function family $\mathbf{f}$ is equivalent to find a property $\mathcal{P}$ such that

- $\mathbf{f} \in \mathcal{P}$, and
- for any $\mathbf{g} \in \mathcal{C}$, $\mathbf{g} \notin \mathcal{P}$.

Having the above proof paradigm in mind, Razborov and Rudich formalized the following question in order to answer Question 1.

**Question 2** *What kind of property **CANNOT** prove strong circuit lower bound?*

## 2.2 Natural proof

The natural proof barrier by Razborov and Rudich shows that when a property $\mathcal{P}$ satisfies several conditions, then $\mathcal{P}$ **CANNOT** prove strong circuit lower bound under reasonable cryptographic assumption.

**Definition 3** *Let $\Gamma$ and $\Lambda$ be two complexity classes and $\delta(n)$ be some function in $n$, we call a property $\mathcal{P} = \{\mathcal{P}_n\}_{n\in\mathbb{N}}$ is $\Gamma$-natural and $\mathcal{C}$-useful with density $\delta(n)$ if the following conditions hold.*

- *(**constructive**) there exists a test $T$ computable in $\Gamma$ such that for any $\mathbf{f} = \{f_n\}_{n\in\mathbb{N}}$, $T(f_n) = 1$ for all $n \in \mathbb{N}$ iff $\mathbf{f} \in \mathcal{P}$. Note that the parameter used in $\Gamma$ is the size of the truth table of $f_n$, i.e., $2^n$.*
- *(**largeness**) $\mathbb{P}_{f_n \leftarrow \mathcal{F}_n}[f_n \in \mathcal{P}_n] \geq \delta(n)$. Equivalently, $|\mathcal{P}_n| \geq \delta(n) \cdot |\mathcal{F}_n|$.*
- *(**usefulness**) for any $f \in \mathcal{C} \cap \mathcal{F}_n$, $f \notin \mathcal{P}_n$.*

We say a proof is a natural proof if the property $\mathcal{P}$ involved in the proof can be efficiently constructed and is non-negligibly large.

**Definition 4 (natural proof)** *We say $\mathcal{P}$ is a $\Gamma$-natural proof for $\mathbf{f} \notin \mathcal{C}$ for some function family $\mathbf{f}$ if $\mathbf{f} \in \mathcal{P}$ and $\mathcal{P}$ is a $\Gamma$-natural against $\mathcal{C}$ with density $1/\mathrm{poly}(n)$.*

Since presenting the main theorem by Razborov and Rudich requires some background in *pseudorandom generator (PRG)*, here we first informally state the theorem as follows.

**Theorem 5 (natural proof barrier (informal))** *Let $\mathcal{C}/s(n)$ be a circuit complexity class. If there exists a $\mathcal{C}/\mathrm{poly}(n)$-natural proof against $\mathcal{C}/s(n)$, then there's no pseudorandom generator in $\mathcal{C}$ with hardness $2^{s^{-1}(n)}$.*

## 2.3 Why largeness?

Let's consider a toy example for proving circuit lower bound. The idea is based on the *formal complexity measure*, which is a special case of the proof paradigm mentioned in Section 2.1.

**Definition 6 (formal complexity measure)** *We say a non-negative function of Boolean function $\mu : (\{0,1\}^* \to \{0,1\}) \to \mathbb{R}^{\geq 0}$ is a formal complexity measure if*

- *$\forall f, g$, $\mu(f \wedge g), \mu(f \vee g) \leq \mu(f) + \mu(g)$, and*
- *$\forall f, g$, $\mu(f + g), \mu(f - g) \leq \max\{\mu(f), \mu(g)\}$.*

Let $\mu$ be a formal complexity measure and $S > 0$, one can use property $\mathcal{P}_S := \{f : \mu(f) \geq S\}$ and the proof paradigm in Section 2.1 to prove circuit lower bound.

The following theorem show that when using formal complexity measure to prove circuit lower bound, the corresponding property must enjoy largeness.

**Theorem 7** *Let $\mu$ be a formal complexity measure and $f \in \mathcal{F}_n$ be a function such that $\mu(f) \geq S$ for some $S > 0$. Then $\mathbb{P}_{g \leftarrow \mathcal{F}_n}[\mu(g) \geq S/4] \geq 1/4$.*

**Proof:** Let $g$ be randomly sampled from $\mathcal{F}_n$ and $h := f \oplus g$. Note that $g, \neg g, h, \neg h$ are all uniformly random function from $\mathcal{F}_n$. Observe that

$$f = g \oplus h = (g \wedge \neg h) \vee (\neg g \wedge h),$$

Thus, by the definition of formal complexity measure $\mu(f) \leq \mu(g) + \mu(\neg h) + \mu(\neg g) + \mu(h)$. Suppose the claim of the theorem is wrong, then by union bound the probability that $\mu(f) < S$ is non-zero, which is a contradiction. ■

## 2.4 Why constructive?

In [Wil16], Williams showed two theorems indicating the difficulty of proving **NEXP** $\notin$ **P/poly** without constructivity.

**Theorem 8 ([Wil16]: Theorem 1.1)** *For all typical $\mathcal{C}$, **NEXP** $\notin \mathcal{C}$ iff there exists a polynomial time computable property of Boolean function that is useful against $\mathcal{C}$ with $O(\log n)$ bits advice.*

**Theorem 9 ([Wil16]: Theorem 1.2)** *For all typical $\mathcal{C}$, **NEXP** $\notin \mathcal{C}$ iff there exists a polynomial time algorithm that is useful against $\mathcal{C}$.*

The above two theorems say that any **NEXP** lower bound should satisfy two of the three properties of Natural Proof (constructivity and usefulness).

However, note that this doesn't mean that every proof for **NEXP** lower bound should satisfy the two conditions since the theorems only say there *exists* such efficiently computable property. For instance, the $\mathbf{ACC^0}$ lower bound for **NEXP** in [Wil14] bypassed these theorems by designing a slightly faster $\mathbf{ACC^0}$ satisfiability algorithm.

## 2.5 Pseudorandom generator (PRG) and pseudorandom function (PRF)

To show the difficulty of having a circuit lower bound with natural proof, Razborov and Rudich based the hardness on the existence of *pseudorandom generator (PRG)*.

PRG is one of the most important object in the study of pseudorandomness. Intuitively, a PRG extends the length of uniform random bit string in a way that small circuit cannot distinguish the output with pure randomness. Formally, we have the following definition.

**Definition 10 (pseudorandom generator (PRG))** *A function $G_k : \{0,1\}^k \to \{0,1\}^{2k}$ is a PRG with hardness $s(k)$ and error $\epsilon(k)$ if $G$ is computable in polynomial time and for any circuit $C$ of size at most $s(k)$,*

$$|\mathbb{P}[C(G_k(U_k)) = 1] - \mathbb{P}[C(U_{2k}) = 1]| \geq \epsilon(k),$$

*where $U_k$ is the uniform distribution on $\{0,1\}^k$.*

Hastad, Impagliazzo, Levin, and Luby showed that if one way function exists (OWF), then PRG exists. [NR04] constructed a PRG in $\mathbf{TC^0}$ based on the DDH assumption.

Furthermore, one can use a PRG to construct a pseudorandom function family (PRF) by the famous GGM construction [GGM84]. First, let's define PRF as follows.

**Definition 11 (pseudorandom function family (PRF))** *We say $\{f_s : f_s : \{0,1\}^n \to \{0,1\}\}_{s\in\{0,1\}^m}$ is a pseudorandom function family (PRF) with hardness $s(m)$ and error $\epsilon(m)$ if given $x \in \{0,1\}^n$ $f_s(x)$ can be computed in polynomial time for any $s \in \{0,1\}^m$ and for any circuit $C$ of size at most $s(m)$,*

$$|\mathbb{P}_{s\leftarrow\{0,1\}^m}[C^{f_s}(1^n) = 1] - \mathbb{P}_{f\leftarrow\mathcal{F}_n}[C^f(1^n) = 1]| \leq \epsilon(n).$$

**Theorem 12 (GGM construction)** *If we have a PRG $G : \{0,1\}^m \to \{0,1\}^{2m}$ with hardness $s(m)$ and error $\epsilon(m)$, then we have a PRF $\{f_s : \ f_s : \{0,1\}^n \to \{0,1\}\}_{s\in\{0,1\}^m}$ with hardness $s(m)$ error $\epsilon(m)/\text{poly}(m)$.*

## 2.6 Razborov-Rudich theorem

Here, we present the original theorem and proof in the paper of Razborov and Rudich [RR97].

**Theorem 13** *Suppose there exists a $\mathbf{P/poly}$-natural proof against $\mathbf{P/poly}$, then there's no PRG against circuit of size $2^{n^\epsilon}$ for any $\epsilon > 0$.*

**Proof Sketch:** Let's first think about the high-level strategy of the proof. As we want to condition on the existence of PRG/OWF, by the method of contradiction, it suffices to show the *existence* of a distinguisher for the PRG when we assume there exists a natural proof.

To do so, let's see what we have right now. With the help of a natural proof against $\mathbf{P/poly}$, we can efficiently distinguish a function in $\mathbf{P/poly}$ from a random function. As a result, it's natural to have the following strategy:

1. Turn a PRG, which is conjectured to lie in $\mathbf{P/poly}$, into a PRF family $\{f_s\}_{s\in\{0,1\}^m}$ in $\mathbf{P/poly}$.
2. By the property of natural proof, one can efficiently distinguish $\{f_s\}_{s\in\{0,1\}^m}$ from a random function.

When carefully picking the parameters, one can efficiently break the original PRG/OWF assumption and thus we base the difficulty of using natural proof on the assumption of the existence of PRG/OWF. □

**Proof:** Let's prove the theorem by the three steps mentioned above.

1. From PRG to PRF (pseudorandom function)
   Here, we use the well-known HILL (Hastad, Impagliazzo, Levin, and Luby) and GGM (Goldreich, Goldwasser, and Micali) construction to construct PRF from OWF. Concretely, assume we have OWF with hardness $2^{n^\epsilon}$ for some $\epsilon > 0$. Then we have a PRF family with seed length $m$ that has hardness $2^{m^{\epsilon'}}$ for some constant $\epsilon'$ related to only $\epsilon$.

2. Use natural property to efficiently distinguish PRF
   Suppose we have a PRF family of seed length $m$ and hardness $2^{m^\epsilon}$. Take $n = m^{\epsilon/2}$. Now, given an input function $h$, which might be a random function or a function from the PRF family, we have oracle access to $h$ in the following.

   First, pad $h$ into a function $g$ with $n$ input bits and construct the truth table of $g$ in $2^{O(n)}$ time. Next, use the constructivity of the natural proof to distinguish $h$ in $\text{poly}(2^{O(n)}) = 2^{O(n)}$ time.

   - When $h$ is a random function, then by the largeness condition, with probability at least $1/\text{poly}(n)$ the algorithm will correctly identify $h$ as random function.
   - When $h$ is from the PRF family, then by the usefulness condition, the algorithm will always correctly identify $h$ as in the PRF family.

   Namely, the above simple algorithm has inverse polynomial advantage on distinguishing PRF from random function in time $2^{O(n)} = O(2^{m^{\epsilon/2}})$

   Observe that $2^{m^{\epsilon/2}} = o(2^{m^\epsilon})$, which is a contradiction to the hardness of PRF.

■

## References

[GGM84] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct randolli functions. In *Foundations of Computer Science, 1984. 25th Annual Symposium on*, pages 464–479. IEEE, 1984.

[NR04] Moni Naor and Omer Reingold. Number-theoretic constructions of efficient pseudo-random functions. *Journal of the ACM (JACM)*, 51(2):231–262, 2004.

[RR97] Alexander A Razborov and Steven Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, 1997.

[Wil14] Ryan Williams. Nonuniform acc circuit lower bounds. *Journal of the ACM (JACM)*, 61(1):2, 2014.

[Wil16] R Ryan Williams. Natural proofs versus derandomization. *SIAM Journal on Computing*, 45(2):497–529, 2016.