Towards P $\neq$ NP from Extended Frege lower bounds*

Ján Pich
University of Oxford

Rahul Santhanam
University of Oxford

September 2023

Abstract

We prove that if conditions I-II (below) hold and there is a sequence of Boolean functions f_n hard to approximate by p-size circuits such that p-size circuit lower bounds for f_n do not have p-size proofs in Extended Frege system EF, then $\mathsf{P} \neq \mathsf{NP}$.

I. S^1_2 proves that a concrete function in E is hard to approximate by subexponential-size circuits.

II. **Learning from $\neg\exists$ OWF.** S^1_2 proves that a p-time reduction transforms circuits breaking one-way functions to p-size circuits learning p-size circuits over the uniform distribution, with membership queries.

Here, S^1_2 is Buss's theory of bounded arithmetic formalizing p-time reasoning.

Further, we show that any of the following assumptions implies that $\mathsf{P} \neq \mathsf{NP}$, if EF is not p-bounded:

1. **Feasible anticheckers.** S^1_2 proves that a p-time function generates anticheckers for SAT.
2. **Witnessing $\mathsf{NP} \not\subseteq \mathsf{P/poly}$.** S^1_2 proves that a p-time function witnesses an error of each p-size circuit which fails to solve SAT.
3. **OWF from $\mathsf{NP} \not\subseteq \mathsf{P/poly}$ & hardness of E.** Condition I holds and S^1_2 proves that a p-time reduction transforms circuits breaking one-way functions to p-size circuits computing SAT.

The results generalize to stronger theories and proof systems.

*A version of the paper with a less technical introduction appeared on ECCC.

1 Introduction

The *proof complexity* approach to the P versus NP problem, sometimes refereed to as the Cook-Reckhow program [3, 22], proceeds by proving lower bounds on lengths of proofs of tautologies in increasingly powerful proof systems - $\mathsf{NP} \neq \mathsf{coNP}$ (and hence $\mathsf{P} \neq \mathsf{NP}$) iff all propositional proof systems have hard sequences of tautologies that require superpolynomial proof size. A problem with the approach is that we do not know if we ever reach the point of proving a superpolynomial lower bound for all proof systems, if we focus only on concrete ones.

There are several results which mitigate this situation. For example, 'lifting' theorems provide a method for deriving monotone circuit lower bounds from lower bounds for weak proof systems [11]. In the algebraic setting, superpolynomial lower bounds for CNF tautologies in the Ideal Proof System imply that the Permanent does not have p-size arithmetic circuits [12]. Proving that a proof system P is not p-bounded implies also that $\mathsf{SAT} \notin \mathsf{P/poly}$ is consistent with a theory T_P corresponding to the proof system P via the standard correspondence in bounded arithmetic [22].

However, no generic way of deriving explicit circuit lower bounds for unrestricted Boolean circuits from proof complexity lower bounds for concrete propositional proof systems of interest has been discovered.[1] In the present paper we connect this problem to several classical questions in complexity theory.

1.1 Our contribution

1.1.1 Core idea: Self-provability of circuit upper bounds

The initial point for our considerations is an observation about a potential 'self-provability' of circuit upper bounds: As it turns out, under certain assumptions, the mere truth of $\mathsf{NP} \subseteq \mathsf{P/poly}$ implies its own provability. To explain the self-provability we first consider a simple-to-state question about a witnessing of $\mathsf{NP} \not\subseteq \mathsf{P/poly}$.

Witnessing $\mathsf{NP} \not\subseteq \mathsf{P/poly}$. Fix a constant $k \geq 1$. Suppose that for each sufficiently big n, no circuit with n inputs and size n^k finds a satisfying assignment for each satisfiable propositional formula of size n, i.e. no n^k-size circuit solves the search version of SAT. Can we witness this assumption 'feasibly' by a p-time function f such that for each sufficiently big n, for each n^k-size circuit C with n inputs and $\leq n$ outputs, $f(C)$ outputs a formula ϕ of size n together with a satisfying assignment a of ϕ such that $\neg\phi(C(\phi))$?

Similar kinds of witnessing have been considered before in the literature, using diagonalization techniques [13, 2, 5]. Indeed, Bogdanov, Talwar and Wan [5] call a similar

[1] It is essentially a trivial observation that there is some proof system P such that if P is not p-bounded then $\mathsf{NP} \neq \mathsf{coNP}$. To see that, consider two cases: I. If $\mathsf{NP} \neq \mathsf{coNP}$, we can take arbitrary P; II. If $\mathsf{NP} = \mathsf{coNP}$, we can take a p-bounded proof system for P. It would be dramatically different to obtain such a connection for a concrete proof system.

feasible witnessing in the uniform setting a "dreambreaker" (following Adam Smith) and show that such a feasible witnessing can be constructed. However, in our applications, it is crucial that the witnessing function finds an error on the input length of the given circuit assuming just that the circuit errs on the given input length, and we do not yet know an unconditional answer to the question in the previous paragraph.

A witnessing function f from the penultimate paragraph exists under the assumption of the existence of a one-way function and a function in E hard for subexponential-size circuits [28, 26], see also Lemma 2. Is it, however, possible to construct it without assuming more than the assumption we want to witness? Formally, we are asking if there is a p-time function f such that for each big enough n propositional formula $w^k_n(f)$, defined by

$$w^k_n(f) := [\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, C(x))] \vee [\mathsf{SAT}_n(f_1(C), f_2(C)) \wedge \neg \mathsf{SAT}_n(f_1(C), C(f_1(C)))],$$

is a tautology. Here, $\mathsf{SAT}_n(x,y)$ is a p-time predicate saying that x is an n-bit string encoding a propositional formula satisfied by assignment y, $C(z)$ says that free variables C represent a circuit with n inputs, $\leq n$ outputs and size n^k which outputs $C(z)$ on z, and $f(C)$ outputs a pair of strings $\langle f_1(C), f_2(C)\rangle$. We do not specify the precise encoding of $w^k_n(f)$. The proof systems we work with simulate Extended Frege system EF and are therefore strong enough to reason efficiently with any natural encoding of $w^k_n(f)$.

If we had a function f such that for some n_0 and all $n > n_0$, $w^k_n(f)$ would be a tautology, we could define an extension of EF, denoted $\mathsf{EF} + w^k(f)$, such that $\mathsf{EF} + w^k(f)$ proofs are EF-proofs which are, in addition, allowed to derive substitutional instances of $w^k_n(f)$, for $n > n_0$.

Self-provability of $\mathsf{NP} \subseteq \mathsf{P/poly}$. Assuming we have a p-time function f such that $w^k_n(f), n > n_0$, is a tautology, we can now turn to the self-provability of $\mathsf{SAT} \in \mathsf{P/poly}$.

Suppose that $\forall n > 1, \mathsf{SAT}_n \in \mathsf{Circuit}[n^{k'}]$, where $\mathsf{Circuit}[s(n)]$ stands for the set of all single-output circuits with n inputs and size $\leq s(n)$. Then, there is a sequence of circuits C with n inputs, $\leq n$ outputs and size n^k falsifying the right disjunct in $w^k_n(f)$, for some $k > k'$ and all $n > 1$. Therefore, $\mathsf{EF} + w^k(f)$ admits p-size proofs of $\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, C(x))$. That is, the mere validity of $\mathsf{SAT}_n \in \mathsf{Circuit}[n^{k'}]$ implies an efficient propositional provability of $\mathsf{SAT}_n \in \mathsf{P/poly}$. The efficient provability of $\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, C(x))$ further implies that $\mathsf{EF} + w^k(f)$ is p-bounded: To prove a tautology ϕ of size n in $\mathsf{EF} + w^k(f)$ it suffices to check out that $\neg\mathsf{SAT}_n(\neg\phi, C(\neg\phi))$ (which implies that $\mathsf{SAT}_n(\phi, y)$ and ϕ hold).

First-order setting. We can formulate also a first-order version of the observation from the previous paragraph. Consider Buss's theory of bounded arithmetic S^1_2 formalizing p-time reasoning, cf. §2.2. Let $W^k_{n_0}(f)$ denote a natural $\forall\Pi^b_1$-formalization of the statement "$\forall n > n_0, w^k_n(f)$", see §2.2 for the definition of Π^b_1. By the correspondence between S^1_2

and EF, cf. [22], if $W^k_{n_0}(f)$ was provable in S^1_2, for some p-time f, then tautologies $w^k_n(f)$, for $n \geq n_0$, would have p-size proofs in EF.

To summarize, we proved the following.

Theorem 1 (Circuit complexity from proof complexity & witnessing of $\mathsf{NP} \not\subseteq \mathsf{P/poly}$)**.** *Let $k \geq 1$ be a constant.*

1. *Suppose that there is a p-time function f such that for each big enough n, $w^k_n(f)$ is a tautology. If $\mathsf{EF} + w^k(f)$ is not p-bounded, then $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^{\epsilon k}]$ for infinitely many n.*

2. *Suppose that there is a p-time function f such that for some n_0, $\mathsf{S}^1_2 \vdash W^k_{n_0}(f)$. If EF is not p-bounded, then $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^{\epsilon k}]$ for infinitely many n.*

In Items 1 and 2, $\epsilon > 0$ is a universal constant (independent of k).

Notably, if for all $k \geq 1$ there is a p-time function f^k such that for each big enough n, $w^k_n(f^k)$ is a tautology, then $\mathsf{NP} \not\subseteq \mathsf{SIZE}[n^\ell]$, for all $\ell \in \mathbb{N}$. This is because by Theorem 1, the existence of such functions f^k and the assumption that for each k, $\mathsf{EF} + w^k(f^k)$ is not p-bounded imply $\mathsf{NP} \not\subseteq \mathsf{P/poly}$. On the other hand, if for some k, $\mathsf{EF} + w^k(f^k)$ is p-bounded, then $\mathsf{NP} = \mathsf{coNP}$ and $\mathsf{NP} \not\subseteq \mathsf{SIZE}[n^\ell]$, for each $\ell \in \mathbb{N}$, by Kannan's lower bound [19].

Corollary 1 (Circuit lower bounds from witnessing)**.**
If for all $k \geq 1$ there is a p-time function f^k such that for each big enough n, $w^k_n(f^k)$ is a tautology, then $\mathsf{NP} \not\subseteq \mathsf{SIZE}[n^\ell]$, for all $\ell \in \mathbb{N}$.

The last observation tells us also that obtaining an unconditional reduction of non-uniform circuit lower bounds to proof complexity of concrete systems will be challenging.

Proposition 1. *If there is a proof system R such that showing that R is not p-bounded implies $\mathsf{NP} \not\subseteq \mathsf{P/poly}$, then $\mathsf{NP} \not\subseteq \mathsf{SIZE}[n^\ell]$, for all $\ell \in \mathbb{N}$.*

Restricting nonuniformity. In Theorem 1, we can restrict the number of nonuniform bits in the concluded lower bounds by adapting formulas $w^k_n(f)$: Assume that the circuit C includes a hardwired description of a fixed universal Turing machine U. Moreover, interpret C as encoding an algorithm A described by $\leq \log n$ bits with $u(n) \leq n^k$ nonuniform bits of advice $a(n)$. The algorithm A and its nonuniform advice are described by free variables. We assume that u is p-time. On each input $z \in \{0,1\}^n$, $C(z)$ uses U to simulate the computation of A on z with access to $a(n)$ up to n^k steps. That is, now the size of C is $poly(n^k)$. Denote the resulting formulas by $w^{k,u}_n(f)$. If we have a p-time function f which witnesses errors of n^k-time algorithms described by $\log n$ bits with $u(n)$ bits of advice attempting to solve the search version of SAT, i.e. such that formulas $w^{k,u}_n(f)$ are tautologies for big enough n, we can define proof system $\mathsf{EF} + w^{k,u}(f)$. Further, we can define $\forall\Pi_1^b$ formulas $W_n^{k,u}(f)$ expressing "$\forall n > n_0,\ w_n^{k,u}(f)$." Denote by $\mathsf{Time}[n^k]/u(n)$ the class of problems solvable by uniform algorithms with $\leq u(n)$ bits of nonuniform advice running in time $O(n^k)$. The proof of Theorem 1 works in this case as well.

Corollary 2 (Circuit complexity from proof complexity & witnessing of $\mathsf{P} \neq \mathsf{NP}$)**.** *Let $k \geq 1$ be a constant and u a p-time function such that $u(n) \leq n^k$.*

1. *Suppose that there is a p-time function f such that for each big enough n, $w_n^{k,u}(f)$ is a tautology. If $\mathsf{EF} + w^{k,u}(f)$ is not p-bounded, then $\mathsf{SAT} \notin \mathsf{Time}[n^{\Omega(k)}]/u(n)$.*
2. *Suppose that there is a p-time function f such that for some n_0, $\mathsf{S}_2^1 \vdash W_{n_0}^{k,u}(f)$. If EF is not p-bounded, then $\mathsf{SAT} \notin \mathsf{Time}[n^{\Omega(k)}]/u(n)$.*

The significance of Corollary 2 is that in the uniform setting, a similar kind of feasible witnessing is known to exist using diagonalization techniques [13, 5]. It is unclear whether diagonalization techniques will suffice to establish that $w_n^{k,u}(f)$ is a tautology for large enough n in some concrete proof system, but there is at least a strong motivation for considering the question, given its implications for deriving strong computational complexity lower bounds from proof complexity lower bounds.

Nonuniform witnessing. If we allow f to be nonuniform, we obtain a version of formulas $w_n^k(f)$ which are unconditionally tautological. This follows from a theorem of Lipton and Young [24], who showed that for each sufficiently big n and each Boolean function f with n inputs which is hard for circuits of size s^3, $s \geq n^3$, there is a set $A_n^{f,s} \subseteq \{0,1\}^n$ of size $poly(s)$ such that no s-size circuit computes f on $A_n^{f,s}$. The set $A_n^{f,s}$ is the set of *anticheckers* of f w.r.t. s. Let n be sufficiently big and α_n^s be tautologies defined by

$$\alpha_n^s := \big(\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, B(x))\big) \vee \Big(\bigvee_{z \in A} C(z) \neq \mathsf{SAT}_n(z)\Big),$$

where $\mathsf{SAT}_n(z) \in \{0,1\}$ is such that $\mathsf{SAT}_n(z) = 1 \Leftrightarrow \exists y, \mathsf{SAT}_n(z,y)$. A is $A_n^{\mathsf{SAT}_n,s}$ if $\mathsf{SAT}_n \notin \mathsf{Circuit}[s^3]$ and an arbitrary $poly(s)$-size subset of $\{0,1\}^n$ otherwise. $C(z)$ in the right disjunct of α_n^s stands for the output of a single-output s-size circuit C with input z. The circuit C is represented by free variables. The circuit B in the left disjunct is a fixed $poly(s)$-size circuit, with n inputs and $\leq n$ outputs, obtained from a fixed s^3-size single-output circuit B' with n inputs such that

$$\mathsf{SAT}_n \in \mathsf{Circuit}[s^3] \Leftrightarrow \forall x \in \{0,1\}^n, B'(x) = \mathsf{SAT}_n(x).$$

The circuit B' exists by considering two cases: I. If $\mathsf{SAT}_n \notin \mathsf{Circuit}[s^3]$, we can take arbitrary s^3-size circuit B'; II. If $\mathsf{SAT}_n \in \mathsf{Circuit}[s^3]$, we can let B' be an s^3-size circuit computing SAT_n. B is obtained from B' in a standard way so that B solves the search version of SAT_n if $\mathsf{SAT}_n \in \mathsf{Circuit}[s^3]$. Similarly as before, we get the following.

Theorem 2 (Circuit complexity from nonuniform proof complexity)**.**
Let $k \geq 3$ be a constant. If there are tautologies without p-size EF*-derivations from substitutional instances of tautologies $\alpha_n^{n^k}$, then* $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^k]$ *for infinitely many n.*

Theorem 2 shows that $\mathsf{EF} + \alpha^{n^k}$, defined analogously as $\mathsf{EF} + w^k(f)$, is in certain sense optimal: If $\forall n, \mathsf{SAT}_n \in \mathsf{Circuit}[n^k]$, then $\mathsf{EF} + \alpha^{n^k}$ has *poly*(n)-size proofs of all tautologies. In fact, there is a p-size circuit which given a tautology ϕ of size n outputs its proof in $\mathsf{EF} + \alpha^{n^k}$. Cook and Krajíček [10] constructed an optimal proof system with 1 bit of nonuniform advice. Their system differs from $\mathsf{EF} + \alpha^{n^k}$ in that it is based on a diagonalization simulating all possible proof systems.

1.1.2 Meta-mathematical implications

We use observations from the previous section to shed light on some classical questions in complexity theory. Specifically, we show that the assumption of the existence of a p-time witnessing function in Theorem 1 can be replaced by an efficient algorithm generating antecheckers for SAT or by efficient reductions collapsing some of Impagliazzo's worlds (together with a hardness of E for subexponential-size circuits).

I. Feasible anticheckers.

Theorem 3 ('CC ← PC' from feasible anticheckers - Informal, cf. Theorem 7)**.**
Let $k \geq 3$ be a constant and assume that there is a p-time function f such that S^1_2 proves:

> "*$\forall 1^n$, $f(1^n)$ outputs a poly(n^k)-size circuit B such that*
>
> $$\forall x, y \in \{0,1\}^n, [\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, B(x))]$$
>
> *or $\big(f(1^n)$ outputs sets $A_n^{\mathsf{SAT}_n, n^k}, A' \subseteq \{0,1\}^n$ of size poly(n^k) such that A' is a set of assignment y_x of formulas $x \in A^{\mathsf{SAT}_n, n^k}$ such that y_x satisfies x if x is satisfiable, and $\forall$ n^k-size circuit C,*
>
> $$\exists x \in A_n^{\mathsf{SAT}_n, n^k}, \mathsf{SAT}_n(x, y_x) \neq C(x)\big)."$$

Then, proving that EF *is not p-bounded implies* $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^k]$ *for infinitely many n.*

Theorem 3 follows easily from the proof of Theorem 1, see §3.

Similarly as in Corollary 1, if for each $k \geq 3$ there is a p-time function f such that S^1_2 proves the statement about the existence of anticheckers from the assumption of Theorem 3, then $\mathsf{NP} \not\subseteq \mathsf{SIZE}[n^\ell]$, for all $\ell \in \mathbb{N}$.

II. One-way functions from $\mathsf{NP} \not\subseteq \mathsf{P/poly}$.

Denote by $\mathsf{tt}(f_n, s)$ a propositional formula expressing that Boolean function $f_n : \{0,1\}^n \mapsto \{0,1\}$ represented by its truth-table is not computable by a Boolean circuit of size s represented by free variables, see §2.4. So $\mathsf{tt}(f_n, s)$ is a tautology if and only if f_n is hard for circuits of size s. The size of the formula $\mathsf{tt}(f_n, s)$ is $poly(2^n, s)$. Similarly, let $\mathsf{tt}(f_n, s, t)$ be a formula expressing that circuits of size s fail to compute f on $\geq t$-fraction of inputs.

Given a function $h \in \mathsf{E}$ such that for some n_0, for each $n \geq n_0$, each $s(2^n)$-size circuit with n inputs fails to compute h_n on $\geq t(2^n)$-fraction of inputs, where s, t are p-time functions in 2^n, we define a proof system $\mathsf{EF} + \mathsf{tt}(h, s, t)$ as an extension of EF which is allowed to derive in its proofs substitutional instances of $\mathsf{tt}(h_n, s(2^n), t(2^n))$, for $n \geq n_0$.

Theorem 4 ('CC ← PC' from 'OWF ← $\mathsf{NP} \not\subseteq \mathsf{P/poly}$' & hard E, cf. Theorem 8)**.** *Assume that for each sufficiently big n, each $2^{n/4}$-size circuit fails to compute $h' \in \mathsf{E}$ on $\geq 1/2 - 1/2^{n/4}$ of all inputs. Further, assume that there is a p-time function $h : \{0,1\}^n \mapsto \{0,1\}^{u(n)}$ such that for each constants c, d, there is a p-time function f and a constant $0 < \epsilon < 1$ such that S^1_2 proves:*

"$\forall n, \forall cn^c$-size circuit C with $u(m)$ inputs and m outputs such that $n \leq dm^d$, $f(C)$ outputs a $poly(n)$-size circuit B such that

$$\forall x, y \in \{0,1\}^{n^\epsilon}, [\mathsf{SAT}_{n^\epsilon}(x, y) \rightarrow \mathsf{SAT}_{n^\epsilon}(x, B(x))]$$

or

$$\Pr_{x \in \{0,1\}^m}[h(C(h(x))) = h(x)] < 1/2."$$

Then, proving that $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ is not p-bounded implies $\mathsf{SAT} \notin \mathsf{P/poly}$.

We remark that if there is a p-time function f' and constant n_0 such that S^1_2 proves that for each $n \geq n_0$, $f'(1^{2^n})$ outputs the truth-table of a function h' with n inputs which is hard on average for $2^{n/4}$-size circuits, then EF and $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ are p-equivalent. (Note that $h' \in \mathsf{E}$.)

The proof of Theorem 4 is based on a formalization of the already-mentioned fact that given a one-way function h and a function in E hard for subexponential-size circuits, we can construct a p-time function witnessing errors of p-size circuits attempting to solve SAT. (The witnessing function outputs formulas encoding the statement $h(x) = b$, with free variables x and suitable constants b, cf. Lemma 2.) We formalize the conditional witnessing in a theory $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$, where $dWPHP(\mathsf{PV})$ stands for a dual weak pigeonhole principle, see §2.2. Combining this with the assumption that S^1_2 proves that a one-way function can be obtained from the hardness of SAT, we obtain the 'ideal' $(\mathsf{S}^1_2 + dWPHP(\mathsf{PV}))$-provable witnessing similar to the tautology $w^k_n(f)$. Having the ideal witnessing statement we proceed as in the proof of Theorem 1 with the difference that the axiom $dWPHP(\mathsf{PV})$ and the assumed hardness of E lead to the system $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ instead of EF, see §4.

III. Learning from the non-existence of OWFs.

Theorem 5 (‘CC $\leftarrow$ PC’ from ‘Learning $\leftarrow \nexists$ OWF’ & hardness of E, cf. Theorem 9)**.** *Let $k, t \geq 1$ be constants. Assume that for each sufficiently big n, each $2^{n/4}$-size circuit fails to compute $h' \in \mathsf{E}$ on $\geq 1/2 - 1/2^{n/4}$ of all inputs. Further, assume that there is a p-time functions $h : \{0,1\}^n \mapsto \{0,1\}^{u(n)}$ such that for each constants c, d, there is a p-time function f and a constant $0 < \epsilon < 1$ such that S^1_2 proves:*

> *“$\forall n$, $\forall cn^c$-size circuits C with $u(m)$ inputs and m outputs such that $n \leq dm^d$, $f(C)$ outputs a $poly(n)$-size circuit B such that B learns $n^{\epsilon t}$-size circuits with n^ϵ inputs, over the uniform distribution, up to error $1/2 - 1/n^\epsilon$, with membership queries and confidence $1/n^\epsilon$,*
> *or*
> $$\Pr_{x \in \{0,1\}^m}[h(C(h(x))) = h(x)] < 1/2.\text{”}$$

Then, there are constants b and a (depending on $k, t, h, h', c, d, f, \epsilon$) such that for each n, the existence of a function $g_n : \{0,1\}^n \mapsto \{0,1\}$ such that no bn^b-size circuit computes g_n on $\geq (1/2 + 1/n)$ fraction of inputs and such that $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ does not have a 2^{an}-size proof of $\mathsf{tt}(g_n, n^t)$ implies that $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^k]$.

Theorem 5 is proved similarly as Theorem 4 with the difference that the provability of efficient learning allows us to prove efficiently only circuit lower bounds instead of all tautologies, see §5.

[29, Lemma 4] shows that, assuming E is S^1_2-provably hard as in Theorem 5, learning algorithms for small circuits can be S^1_2-provably constructed from circuits automating EF on tt-formulas.[2] Theorem 5 thus implies that (assuming E is S^1_2-provably hard) S^1_2-deriving automatability of EF from the non-existence of one-way functions would reduce circuit complexity to EF lower bounds.

Generalization to stronger proof systems. S^1_2 in Theorems 1 & 3-5 can be replaced by essentially an arbitrary first-order theory T containing S^1_2 and satisfying some basic properties, if we simultaneously replace EF in conclusions of Theorems 1 & 3-5 by a suitable propositional proof system P_T such that propositional translations of Π^b_1 theorems of T have p-size proofs in P_T.

Weakening the assumptions. The core component of Theorems 1 & 3-5 is the existence of a suitable reduction. For example, in case of Theorem 5 we need a p-time reduction constructing learning algorithms from circuits breaking one-way functions. If such a reduction exists, even without assuming its provability in S^1_2, we can build a propositional proof system P by adding tautologies encoding the correctness of the reduction to EF.

[2] [29, Lemma 4] assumes also the existence of a prime. The assumption can be removed after moving to the propositional setting.

Then, showing that the resulting proof system $P + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ is not p-bounded on tt-tautologies would separate P and NP. This shows that the S^1_2-provability in the assumptions of our theorems can be weakened just to the validity of the respective statements, if we use stronger systems than EF in their conclusions. It also shows that the most of technicalities in the present paper stem from making the presented approach work for EF.

Moreover, we remark that if our final goal is to prove that $\mathsf{P} \neq \mathsf{NP}$, then the first assumption of Theorems 4-5 postulating the existence of a hard Boolean function in E is given to us 'for free', as otherwise, if all functions in E can be approximated by subexponential-size circuits, it is not hard to show that $\mathsf{P} \neq \mathsf{NP}$.

1.1.3 Plausibility of the assumptions

Impagliazzo's worlds. In a famous survey of Impagliazzo [15], he described 5 possible worlds of average-case complexity: Algorithmica, Heuristica, Pessiland, Minicrypt and Cryptomania. Recently, there have been various approaches proposed to rule out Heuristica and Pessiland (see, for example, [14, 32, 25]) by studying the complexity of problems about compression, such as the Minimum Circuit Size Problem (MCSP) and the problem of computing time-bounded Kolmogorov complexity. Our results have implications for the feasibility of such efforts - provable collapses of Impagliazzo's worlds would imply a new and surprising link between proof complexity and circuit complexity. For example, the reduction assumed in Theorem 5 asks for a construction of learning algorithms from circuits breaking one-way functions. Morally, the existence of such a reduction would rule out Pessiland out of Impagliazzo's worlds. Similarly, the reduction assumed in Theorem 4 would rule out Pessiland and Heuristica.

Feasible MinMax theorem. The proofs of the existence of anticheckers we are aware of use the efficient MinMax theorem [1, 24, 27] or similar methods. If we had a proof of MinMax which would use counting with only polynomial precision (formally, APC_1-counting) and if we could replace p-time f in Theorem 3 by the existential quantifiers (see §3 for a discussion of the issue), we could prove the existence of anticheckers in APC_1 and obtain the desired reduction of circuit complexity to proof complexity. Here, APC_1 is Jeřábek's theory of approximate counting, see §2.2.

S^1_2-provability of a circuit lower bound. If we want to replace $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ in Theorems 4 & 5 by EF, it suffices to assume the S^1_2-provability of a subexponential circuit lower bound for E. This assumption has an interesting status. Razborov's conjecture about hardness of Nisan-Wigderson generators implies a conditional hardness of formulas $\mathsf{tt}(h, n^{O(1)})$ for Frege (for every h), cf. [31], and it is possible to consider extensions of the conjecture to all standard proof systems, even set theory ZFC. A conditional hardness of tt-formulas (for EF) follows also from a conjecture of Krajíček [21, Conjecture

7.9]. If the tt-formulas expressing subexponential lower bounds for E are hard for EF, then S^1_2 cannot prove the lower bounds either. On the other hand, it is not known how to prove hardness of $\mathsf{tt}(h, 2^{n/4})$, for all h, for Frege, under any standard complexity-theoretic hardness assumption. Moreover, all major circuit lower bounds for weak circuit classes and explicit Boolean functions are known to be provable in S^1_2, cf. [30, 26].[3] It is thus perfectly possible that subexponential average-case circuit lower bounds for E are provable in a theory such as S^1_2.[4]

1.1.4 Revising the status of the Cook-Reckhow program

Showing that statements like $\mathsf{P} \neq \mathsf{NP}$ follow from proof complexity lower bounds for concrete proof systems is considered so challenging that there have not been practically any conscious attempts to approach it. Our results show that the significant efforts that have been made in order to address some of the central problems in cryptography and learning theory are, in fact, aiming to establish precisely that. This can be interpreted as an evidence for the hardness of resolving the relevant problems in cryptography and learning theory, but also as showing that proving that the Cook-Reckhow program could be in principle realized successfully might not be completely out of reach. In any case, the presented results demonstrate a new fundamental connection between proof complexity, cryptography and learning theory.

1.1.5 Self-provability from random self-reducibility

In §6 we show that a self-provability of circuit upper bounds can be obtained for some random-selfreducible problems. We exhibit this in the case of the discrete logarithm: Assuming the existence of a hard Boolean function in E and the existence of efficient circuits solving the discrete logarithm problem, we show that an explicit proof system admits p-size proofs of the fact that the discrete logarithm problem is solvable efficiently.

1.2 Open problems

1. Prove that any superpolynomial EF lower bound separates P and NP by constructing a witnessing function such that, for all big enough n, tautologies $w^k_n(f)$ have p-size EF-proofs. It would be interesting to construct such witnessing functions f even assuming $\mathsf{P} \not\subseteq \mathsf{SIZE}[n^\ell]$, for all $\ell \in \mathbb{N}$. By Corollary 1, an unconditional construction of such witnessing functions would imply $\mathsf{NP} \not\subseteq \mathsf{SIZE}[n^\ell]$, for all $\ell \in \mathbb{N}$.

[3] This has not been verified for lower bounds obtained via the algorithmic method of Williams [33].

[4] We emphasize that Theorems 4 & 5 do not require that S^1_2 proves a circuit lower bound. Further, we can replace the system $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ in Theorems 4 & 5 by, say, a propositional proof system corresponding to ZFC, if we assume that a subexponential circuit lower bound for E is provable in (the set theory) ZFC.

2. Prove that the p-time functions defining reductions in Theorems 3-5 can be replaced by the existential quantifiers, see §3 for a discussion of the problem.

2 Preliminaries

2.1 Learning algorithms

$[n]$ denotes $\{1, \dots, n\}$. 1^n stands for a string of n 1s. $\mathsf{Circuit}[s]$ denotes fan-in two Boolean circuits of size at most s. The size of a circuit is the number of its gates. A function $f : \{0,1\}^n \mapsto \{0,1\}$ is γ-approximated by a circuit C, if $\Pr_x[C(x) = f(x)] \geq \gamma$.

Definition 1 (PAC learning)**.** *A circuit class $\mathcal{C}$ is learnable over the uniform distribution by a circuit class $\mathcal{D}$ up to error ϵ with confidence δ, if there are randomized oracle circuits L^f from $\mathcal{D}$ such that for every Boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ computable by a circuit from $\mathcal{C}$, when given oracle access to f, input 1^n and the internal randomness $w \in \{0,1\}^*$, L^f outputs the description of a circuit satisfying*

$$\Pr_w[L^f(1^n, w)\ (1-\epsilon)\text{-}\mathit{approximates}\ f] \geq \delta.$$

L^f uses non-adaptive membership queries if the set of queries which L^f makes to the oracle does not depend on the answers to previous queries. L^f uses random examples if the set of queries which L^f makes to the oracle is chosen uniformly at random.

In this paper, PAC learning always refers to learning over the uniform distribution. While, a priori, learning over the uniform distribution might not reflect real-world scenarios very well (and on the opposite end, learning over all distributions is perhaps overly restrictive), as far as we can tell it is possible that PAC learning of p-size circuits over the uniform distribution implies PAC learning of p-size circuits over all p-samplable distributions. Binnendyk, Carmosino, Kolokolova, Ramyaa and Sabin [4] proved the implication, if the learning algorithm in the conclusion is allowed to depend on the p-samplable distribution.

2.2 Bounded arithmetic and propositional logic

Theories of bounded arithmetic capture various levels of feasible reasoning and present a uniform counterpart to propositional proof systems.

The first theory formalizing p-time reasoning was introduced by Cook [8] as an equational theory PV. We work with its first-order conservative extension PV_1 from [23]. The language of PV_1, denoted PV as well, consists of symbols for all p-time algorithms given by Cobham's characterization of p-time functions, cf. [7]. A PV-formula is a first-order formula in the language PV. Σ^b_0 (=Π^b_0) denotes PV-formulas with only sharply bounded quantifiers $\exists x, x \leq |t|$, $\forall x, x \leq |t|$, where $|t|$ is "the length of the binary representation of t". Inductively, Σ^b_{i+1} resp. Π^b_{i+1} is the closure of Π^b_i resp. Σ^b_i under positive Boolean combinations, sharply bounded quantifiers, and bounded quantifiers $\exists x, x \leq t$ resp. $\forall x, x \leq t$. Predicates definable by Σ^b_i resp. Π^b_i formulas are in the Σ^p_i resp. Π^p_i level of the polynomial hierarchy, and vice versa. PV_1 is known to prove $\Sigma^b_0(\mathsf{PV})$-induction:

$$A(0) \wedge \forall x\ (A(x) \rightarrow A(x+1)) \rightarrow \forall x A(x),$$

for Σ^b_0-formulas A, cf. Krajíček [20].

Buss [6] introduced the theory S^1_2 extending PV_1 with the Σ^b_1-length induction:

$$A(0) \wedge \forall x < |a|, (A(x) \rightarrow A(x+1)) \rightarrow A(|a|),$$

for $A \in \Sigma^b_1$. S^1_2 proves the sharply bounded collection scheme $BB(\Sigma^b_1)$:

$$\forall i < |a|\ \exists x < a, A(i,x) \rightarrow \exists w\ \forall i < |a|, A(i, [w]_i),$$

for $A \in \Sigma^b_1$ ($[w]_i$ is the ith element of the sequence coded by w), which is unprovable in PV_1 under a cryptographic assumption, cf. [9]. On the other hand, S^1_2 is $\forall\Sigma^b_1$-conservative over PV_1. This is a consequence of Buss's witnessing theorem stating that $\mathsf{S}^1_2 \vdash \exists y, A(x,y)$ for $A \in \Sigma^b_1$ implies $\mathsf{PV}_1 \vdash A(x, f(x))$ for some PV-function f.

Following a work by Krajíček [21], Jeřábek [16, 17, 18] systematically developed a theory APC_1 capturing probabilistic p-time reasoning by means of approximate counting.[5] The theory APC_1 is defined as $\mathsf{PV}_1 + dWPHP(\mathsf{PV})$ where $dWPHP(\mathsf{PV})$ stands for the dual (surjective) pigeonhole principle for PV-functions, i.e. for the set of all formulas

$$x > 0 \rightarrow \exists v < x(|y|+1) \forall u < x|y|,\ f(u) \neq v,$$

where f is a PV-function which might involve other parameters not explicitly shown. We devote §2.3 to a more detailed description of the machinery of approximate counting in APC_1.

Any Π^b_1-formula ϕ provable in S^1_2 can be expressed as a sequence of tautologies $||\phi||_n$ with proofs in the Extended Frege system EF which are constructible in p-time (given a string of the length n), cf. [8]. We refer to Krajíček [22] for basic notions in proof complexity such as EF. As it is often easier to present a proof in a theory of bounded arithmetic than in the corresponding propositional system, bounded arithmetic functions, so to speak, as a uniform language for propositional logic.

[5] Krajíček [21] introduced a theory BT defined as $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$ and proposed it as a theory for probabilistic p-time reasoning.

2.3 Approximate counting

In order to prove our results we will need to use Jeřábek's theory of approximate counting. This section recalls the properties of $\mathsf{APC_1}$ we will need.

By a definable set we mean a collection of numbers satisfying some formula, possibly with parameters. When a number a is used in a context which asks for a set it is assumed to represent the integer interval $[0, a)$, e.g. $X \subseteq a$ means that all elements of set X are less than a. If $X \subseteq a$, $Y \subseteq b$, then $X \times Y := \{bx + y \mid x \in X, y \in Y\} \subseteq ab$ and $X \dot{\cup} Y := X \cup \{y + a \mid y \in Y\} \subseteq a + b$. Rational numbers are assumed to be represented by pairs of integers in the natural way. We use the notation $x \in Log \leftrightarrow \exists y,\ x = |y|$ and $x \in LogLog \leftrightarrow \exists y,\ x = ||y||$.

Let $C : 2^n \to 2^m$ be a circuit and $X \subseteq 2^n, Y \subseteq 2^m$ definable sets. We write $C : X \twoheadrightarrow Y$ if $\forall y \in Y \exists x \in X,\ C(x) = y$. Jeřábek [18] gives the following definitions in $\mathsf{APC_1}$, but they can be formulated in $\mathsf{PV_1}$ as well.

Definition 2. *Let $X, Y \subseteq 2^n$ be definable sets, and $\epsilon \leq 1$. The size of X is approximately less than the size of Y with error ϵ, written as $X \preceq_\epsilon Y$, if there exists a circuit C, and $v \neq 0$ such that*

$$C : v \times (Y \dot{\cup} \epsilon 2^n) \twoheadrightarrow v \times X.$$

$X \approx_\epsilon Y$ stands for $X \preceq_\epsilon Y$ and $Y \preceq_\epsilon X$.

Since a number s is identified with the interval $[0, s)$, $X \preceq_\epsilon s$ means that the size of X is at most s with error ϵ.

The definition of $X \preceq_\epsilon Y$ is an unbounded $\exists \Pi^b_2$-formula even if X, Y are defined by circuits so it cannot be used freely in bounded induction. Jeřábek [18] solved this problem by working in HARD^A, a conservative extension of $\mathsf{APC_1}$, defined as a relativized theory $\mathsf{PV_1}(\alpha) + dWPHP(\mathsf{PV}(\alpha))$ extended with axioms postulating that $\alpha(x)$ is a truth-table of a function on $||x||$ variables hard on average for circuits of size $2^{||x||/4}$, see §2.4.2. In HARD^A there is a $\mathsf{PV_1}(\alpha)$ function $Size$ approximating the size of any set $X \subseteq 2^n$ defined by a circuit C so that $X \approx_\epsilon Size(C, 2^n, 2^{\epsilon^{-1}})$ for $\epsilon^{-1} \in Log$, cf. [18, Lemma 2.14]. If $X \cap t \subseteq 2^{|t|}$ is defined by a circuit C and $\epsilon^{-1} \in Log$, we can define

$$\Pr_{x<t}[x \in X]_\epsilon := \frac{1}{t} Size(C, 2^{|t|}, 2^{\epsilon^{-1}}).$$

The presented definitions of approximate counting are well-behaved:

Proposition 2 (Jeřábek [18])**.** *(in $\mathsf{PV_1}$) Let $X, X', Y, Y', Z \subseteq 2^n$ and $W, W' \subseteq 2^m$ be definable sets, and $\epsilon, \delta < 1$. Then*

i) $X \subseteq Y \Rightarrow X \preceq_0 Y$,
ii) $X \preceq_\epsilon Y \wedge Y \preceq_\delta Z \Rightarrow X \preceq_{\epsilon+\delta} Z$,
iii) $X \preceq_\epsilon X' \wedge W \preceq_\delta W' \Rightarrow X \times W \preceq_{\epsilon+\delta+\epsilon\delta} X' \times W'$.
iv) $X \preceq_\epsilon X' \wedge Y \preceq_\delta Y'$ *and X', Y' are separable by a circuit, then* $X \cup Y \preceq_{\epsilon+\delta} X' \cup Y'$.

Proposition 3 (Jeřábek [18]). *(in APC_1)*
1. Let $X, Y \subseteq 2^n$ be definable by circuits, $s, t, u \leq 2^n$, $\epsilon, \delta, \theta, \gamma < 1, \gamma^{-1} \in Log$. Then
i) $X \preceq_\gamma Y$ *or* $Y \preceq_\gamma X$,
ii) $s \preceq_\epsilon X \preceq_\delta t \Rightarrow s < t + (\epsilon + \delta + \gamma)2^n$,
iii) $X \preceq_\epsilon Y \Rightarrow 2^n \backslash Y \preceq_{\epsilon+\gamma} 2^n \backslash X$,
iv) $X \approx_\epsilon s \wedge Y \approx_\delta t \wedge X \cap Y \approx_\theta u \Rightarrow X \cup Y \approx_{\epsilon+\delta+\theta+\gamma} s + t - u$.

2. (Disjoint union) Let $X_i \subseteq 2^n$, $i < m$ be defined by a sequence of circuits and $\epsilon, \delta \leq 1$, $\delta^{-1} \in Log$. If $X_i \preceq_\epsilon s_i$ for every $i < m$, then $\bigcup_{i<m}(X_i \times \{i\}) \preceq_{\epsilon+\delta} \sum_{i<m} s_i$.

When proving Σ^b_1 statements in APC_1 we can afford to work in $\mathsf{S}^1_2 + dWPHP(\mathsf{PV}) + BB(\Sigma^b_2)$ and, in fact, assuming the existence of a single hard function in PV_1 gives us the full power of APC_1. Here, $BB(\Sigma^b_2)$ is defined as $BB(\Sigma^b_1)$ but with $A \in \Sigma^b_2$.

Lemma 1 ([26]). *Suppose $\mathsf{S}^1_2 + dWPHP(\mathsf{PV}) + BB(\Sigma^b_2) \vdash \exists y A(x, y)$ for $A \in \Sigma^b_1$. Then, for every $\epsilon < 1$, there is k and PV-functions g, h such that PV_1 proves*

$$|f| \geq |x|^k \wedge \exists y, |y| = ||f||, C_h(y) \neq f(y) \rightarrow A(x, g(x, f))$$

where $f(y)$ is the yth bit of f, $f(y) = 0$ for $y > |f|$, and C_h is a circuit of size $\leq 2^{\epsilon||f||}$ generated by h on f, x. Moreover, $\mathsf{APC}_1 \vdash \exists y A(x, y)$.

2.4 Formalizing complexity-theoretic statements

2.4.1 Circuit lower bounds

An 'almost everywhere' circuit lower bound for circuits of size s and a function f says that for every sufficiently big n, for each circuit C with n inputs and size s, there exists an input y on which the circuit C fails to compute $f(y)$.

If $f : \{0,1\}^n \rightarrow \{0,1\}$ is an NP function and $s = n^k$ for a constant k, this can be written down as a $\forall\Sigma^b_2$ formula $\mathsf{LB}(f, n^k)$,

$$\forall n,\ n > n_0\ \forall \text{ circuit } C \text{ of size } \leq n^k\ \exists y,\ |y| = n,\ C(y) \neq f(y),$$

where n_0 is a constant and $C(y) \neq f(y)$ is a Σ^b_2 formula stating that a circuit C on input y outputs the opposite value of $f(y)$. The intended meaning of '$\exists y, |y| = n$' is to say that y is a string from $\{0,1\}^n$. This is a slight abuse of notation as, formally, $|y| = n$ fixes the first bit of y to 1.

If we want to express $s(n)$-size lower bounds for $s(n)$ as big as $2^{O(n)}$, we add an extra assumption on n stating that $\exists x,\ n = ||x||$. That is, the resulting formula $\mathsf{LB}_{\mathsf{tt}}(f, s(n))$ has form '$\forall x, n; n = ||x|| \rightarrow \dots$'. Treating x, n as free variables, $\mathsf{LB}_{\mathsf{tt}}(f, s(n))$ is Π^b_1 if f is, for instance, SAT because $n = ||x||$ implies that the quantifiers bounded by $2^{O(n)}$ are sharply bounded. Moreover, allowing $f \in \mathsf{NE}$ lifts the complexity of $\mathsf{LB}_{\mathsf{tt}}(f, s(n))$ just to $\forall\Sigma^b_1$. The function $s(n)$ in $\mathsf{LB}_{\mathsf{tt}}(f, s(n))$ is assumed to be a PV-function with input x (satisfying $||x|| = n$).

In terms of the *Log*-notation, $\mathsf{LB}(f, n^k)$ implicitly assumes $n \in Log$ while $\mathsf{LB}_{\mathsf{tt}}(f, n^k)$ assumes $n \in LogLog$. By chosing the scale of n we are determining how big objects are going to be 'feasible' for theories reasoning about the statement. In the case $n \in LogLog$, the truth-table of f (and everything polynomial in it) is feasible. Assuming just $n \in Log$ means that only the objects of polynomial-size in the size of the circuit are feasible. Likewise, the theory reasoning about the circuit lower bound is less powerful when working with $\mathsf{LB}(f, n^k)$ than with $\mathsf{LB}_{\mathsf{tt}}(f, n^k)$. (The scaling in $\mathsf{LB}_{\mathsf{tt}}(f, s)$ corresponds to the choice of parameters in natural proofs and in the formalizations by Razborov [30].)

We can analogously define formulas $\mathsf{LB}_{\mathsf{tt}}(f, s(n), t(n))$ expressing an average-case lower bound for f, where f is a free variable (with $f(y)$ being the yth bit of f and $f(y) = 0$ for $y > |f|$). More precisely, $\mathsf{LB}_{\mathsf{tt}}(f, s(n), t(n))$ generalizes $\mathsf{LB}_{\mathsf{tt}}(f, s(n))$ by saying that each circuit of size $s(n)$ fails to compute f on at least $t(n)$ inputs, for PV-functions $s(n), t(n)$. Since $n \in LogLog$, $\mathsf{LB}_{\mathsf{tt}}(f, s(n), t(n))$ is Π^b_1.

Propositional version. An $s(n)$-size circuit lower bound for a function $f : \{0,1\}^n \to \{0,1\}$ can be expressed by a $poly(2^n, s)$-size propositional formula $\mathsf{tt}(f, s)$,

$$\bigvee_{y \in \{0,1\}^n} f(y) \neq C(y)$$

where the formula $f(y) \neq C(y)$ says that an $s(n)$-size circuit C represented by $poly(s)$ variables does not output $f(y)$ on input y. The values $f(y)$ are fixed bits. That is, the whole truth-table of f is hard-wired in $\mathsf{tt}(f, s)$.

The details of the encoding of the formula $\mathsf{tt}(f, s)$ are not important for us as long as the encoding is natural because systems like EF considered in this paper can reason efficiently about them. We will assume that $\mathsf{tt}(f, s)$ is the formula resulting from the translation of Π^b_1 formula $\mathsf{LB}_{\mathsf{tt}}(h, s)$, where $n_0 = 0$, n, x are substituted after the translation by fixed constants so that $x = 2^{2^n}$, and h is a free variable (with $h(y)$ being the yth bit of h and $h(y) = 0$ for $y > |h|$) which is substituted after the translation by constants defining f.

Analogously, we can express average-case lower bounds by propositional formulas $\mathsf{tt}(f, s(n), t(n))$ obtained by translating $\mathsf{LB}_{\mathsf{tt}}(h, s(n), t(n)2^n)$, with $n_0 = 0$, fixed $x = 2^{2^n}$ and h substituted after the translation by f.

2.4.2 Learning algorithms

A circuit class $\mathcal{C}$ is defined by a PV-formula if there is a PV-formula defining the predicate $C \in \mathcal{C}$. Definition 1 can be formulated in the language of HARD^A: A circuit class $\mathcal{C}$ (defined by a PV-formula) is learnable over the uniform disribution by a circuit class $\mathcal{D}$ (defined by a PV-formula) up to error ϵ with confidence δ, if there are randomized oracle circuits L^f from $\mathcal{D}$ such that for every Boolean function $f : \{0,1\}^n \mapsto \{0,1\}$ (represented by its truth-table) computable by a circuit from $\mathcal{C}$, for each $\gamma^{-1} \in Log$, when given oracle access to f, input 1^n and the internal randomness $w \in \{0,1\}^*$, L^f outputs the description of a circuit satisfying

$$\Pr_w[L^f(1^n, w)\ (1-\epsilon)\text{-approximates } f]_\gamma \geq \delta.$$

The inner probability of approximability of f by $L^f(1^n, w)$ is counted exactly. This is possible because f is represented by its truth-table, which implies that $2^n \in Log$.

Propositional version. In order to translate the definition of learning algorithms to propositional formulas and to the language of PV_1 we need to look more closely at the definition of HARD^A.

PV_1 can be relativized to $\mathsf{PV}_1(\alpha)$. The new function symbol α is then allowed in the inductive clauses for introduction of new function symbols. This means that the language of $\mathsf{PV}_1(\alpha)$, denoted also $\mathsf{PV}(\alpha)$, contains symbols for all p-time oracle algorithms.

Proposition 4 (Jeřábek [16]). *For every constant $\epsilon < 1/3$ there exists a constant n_0 such that APC_1 proves: for every $n \in LogLog$ such that $n > n_0$, there exist a function $f : 2^n \to 2$ such that no circuit of size $2^{\epsilon n}$ computes f on $\geq (1/2 + 1/2^{\epsilon n})2^n$ inputs.*

Definition 3 (Jeřábek [16]). *The theory HARD^A is an extension of the theory $\mathsf{PV}_1(\alpha) + dWPHP(\mathsf{PV}(\alpha))$ by the axioms*

1. *$\alpha(x)$ is a truth-table of a Boolean function in $||x||$ variables,*
2. *$\mathsf{LB}_{\mathsf{tt}}(\alpha(x), 2^{||x||/4}, 2^{||x||}(1/2 - 1/2^{||x||/4}))$, for constant n_0 from Proposition 4,*
3. *$||x|| = ||y|| \to \alpha(x) = \alpha(y)$.*

By inspecting the proof of Lemma 2.14 in [18], we can observe that on each input $C, 2^n, 2^{\epsilon^{-1}}$ the $\mathsf{PV}_1(\alpha)$-function $Size$ calls α just once (to get the truth-table of a hard function which is then used as the base function of the Nisan-Wgiderson generator). In fact, $Size$ calls α on input x which depends only on $|C|$, the number of inputs of C and w.l.o.g. also just on $|\epsilon^{-1}|$ (since decreasing ϵ leads only to a better approximation). In combination with the fact that the approximation $Size(C, 2^n, 2^{\epsilon^{-1}}) \approx_\epsilon X$, for $X \subseteq 2^n$ defined by C, is not affected by a particular choice of the hard boolean function generated by α, we get that APC_1 proves

$$\mathsf{LB}_{\mathsf{tt}}(y, 2^{||y||/4}, 2^{||y||}(1/2 - 1/2^{||y||/4})) \wedge ||y|| = S(C, 2^n, 2^{\epsilon^{-1}}) \to\ Sz(C, 2^n, 2^{\epsilon^{-1}}, y) \approx_\epsilon X,$$

where Sz is defined as $Size$ with the only difference that the call to $\alpha(x)$ on $C, 2^n, 2^{\epsilon^{-1}}$ is replaced by y and $S(C, 2^n, 2^{\epsilon^{-1}}) = ||x||$ for a PV-function S. (S is given by a subcomputation of $Size$ specifying $||x||$, for x on which $Size$ queries $\alpha(x)$.)

This allows us to express $\Pr_{x<t}[x \in X]_\epsilon = a$, where $\epsilon^{-1} \in Log$ and $X \cap t \subseteq 2^{|t|}$ is defined by a circuit C, without a $\mathsf{PV}_1(\alpha)$ function, by formula

$$\forall y\ (\mathsf{LB}_{\mathsf{tt}}(y, 2^{||y||/4}, 2^{||y||}(1/2 - 1/2^{||y||/4})) \wedge ||y|| = S(C, 2^{|t|}, 2^{\epsilon^{-1}}) \rightarrow Sz(C, 2^{|t|}, 2^{\epsilon^{-1}}, y)/t = a).$$

We denote the resulting formula by $\Pr^y_{x<t}[x \in X]_\epsilon = a$. We will use the notation $\Pr^y_{x<t}[x \in X]_\epsilon$ in equations with the intended meaning that the equation holds for the value $Sz(\cdot,\cdot,\cdot,\cdot)/t$ under corresponding assumptions. For example, $t \cdot \Pr^y_{x<t}[x \in X]_\epsilon \preceq_\delta a$ stands for '$\forall y, \exists v, \exists$ circuit $\hat{C}$ (defining a surjection) which witnesses that $\mathsf{LB}_{\mathsf{tt}}(y, 2^{||y||/4}, 2^{||y||}(1/2 - 1/2^{||y||/4})) \wedge ||y|| = S(C, 2^{|t|}, 2^{\epsilon^{-1}})$ implies $Sz(C, 2^{|t|}, 2^{\epsilon^{-1}}, y) \preceq_\delta a$'.

The definition of learning can be now expressed without a $\mathsf{PV}_1(\alpha)$ function: If circuit class $\mathcal{C}$ is defined by a PV-function, the statement that a given oracle algorithm L (given by a PV-function with oracle queries) learns a circuit class $\mathcal{C}$ over the uniform distribution up to error ϵ with confidence δ can be expressed as before with the only difference that we replace $\Pr_w[L^f(1^n, w)\ (1-\epsilon)\text{-approximates } f]_\gamma \geq \delta$ by

$$\Pr_w^y[L^f(1^n, w)\ (1-\epsilon)\text{-approximates } f]_\gamma \geq \delta.$$

Since the resulting formula A defining learning is not Π^b_1 (because of the assumption $\mathsf{LB}_{\mathsf{tt}}$) we cannot translate it to propositional logic. We will sidestep the issue by translating only the formula B obtained from A by deleting subformula $\mathsf{LB}_{\mathsf{tt}}$ (but leaving $||y|| = S(\cdot,\cdot,\cdot)$ intact) and replacing the variables y by fixed bits representing a hard boolean function. In more detail, Π^b_1 formula B can be translated into a sequence of propositional formulas $\mathsf{lear}^y_\gamma(L, \mathcal{C}, \epsilon, \delta)$ expressing that "if $C \in \mathcal{C}$ is a circuit computing f, then L querying f generates a circuit D such that $\Pr[D(x) = f(x)] \geq 1 - \epsilon$ with probability $\geq \delta$, which is counted approximately with precision γ". Note that C, f are represented by free variables and that there are also free variables for error γ from approximate counting and for Boolean functions y. As in the case of tt-formulas, we fix $|f| = 2^n$, so n is not a free variable. Importantly, $\mathsf{lear}^y_\gamma(L, \mathcal{C}, \epsilon, \delta)$ does not postulate that y is a truth-table of a hard boolean function. Nevertheless, for any fixed (possibly non-uniform) bits representing a sequence of Boolean functions $h = \{h_m\}_{m>n_0}$ such that h_m is not $(1/2 + 1/2^{m/4})$-approximable by any circuit of size $2^{m/4}$, we can obtain formulas $\mathsf{lear}^h_\gamma(L, \mathcal{C}, \epsilon, \delta)$ by substituting bits h for y.

Using a single function h in $\mathsf{lear}^h_\gamma(L, \mathcal{C}, \epsilon, \delta)$ does not ruin the fact that (the translation of function) Sz approximates the respective probability with accuracy γ because Sz queries a boolean function y which depends just on the number of atoms representing γ^{-1} and on the size of the circuit D defining the predicate we count together with the number of inputs of D. The size of D and the number of its inputs are w.l.o.g. determined by the number of inputs of f.

If we are working with formulas $\mathsf{lear}^h_\gamma(L, \mathcal{C}, \epsilon, \delta)$, where h is a sequence of bits representing a hard boolean function, in a proof system which cannot prove efficiently that h

is hard, our proof system might not be able to show that the definition is well-behaved - it might not be able to derive some standard properties of the function Sz used inside the formula. Nevertheless, in our theorems this will never be the case: our proof systems will always know that h is hard.

In formulas $\mathsf{lear}^y_\gamma(L,\mathcal{C},\epsilon,\delta)$ we can allow L to be a sequence of nonuniform circuits, with a different advice string for each input length. One way to see that is to use additional input to L in Π^b_1 formula B, then translate the formula to propositional logic and substitute the right bits of advice for the additional input. Again, the precise encoding of the formula $\mathsf{lear}^y_\gamma(L,\mathcal{C},\epsilon,\delta)$ does not matter very much to us but in order to simplify proofs we will assume that $\mathsf{lear}^y_\gamma(L,\mathsf{Circuit}[n^k],\epsilon,\delta)$ has the from $\neg\mathsf{tt}(f,n^k)\rightarrow R$, where n,k are fixed, f is represented by free variables and R is the remaining part of the formula expressing that L generates a suitable circuit with high probability.

2.4.3 Nisan-Wigderson generators

The core theorem underlying approximate counting in APC_1 is the following formalization of Nisan-Wigderson generators (NW), cf. [16, Proposition 4.7].

Theorem 6 (Jeřábek [16])**.** *Let $0<\gamma<1$. There are constants $c>1$ and $\delta'>0$ such that for each $\delta<\delta'$ there is a poly(2^m)-time function*

$$NW:\{0,1\}^{cm}\times\{0,1\}^{2^m}\mapsto\{0,1\}^{\lfloor 2^{\delta m}\rfloor}$$

such that S^1_2 proves: "If $2^m\in Log$ and $f:\{0,1\}^m\mapsto\{0,1\}$ is a Boolean function such that no circuit of size $2^{\epsilon m}$ computes f on $>(1/2+1/2^{\epsilon m})2^m$ inputs, then for each $(2^{\epsilon m}-\lceil 2^{(\delta+\gamma)m}\rceil-1)$-size circuit D with $\lfloor 2^{\delta m}\rfloor$ inputs,

$$2^{\lfloor 2^{\delta m}\rfloor}\times\{z<2^{cm}\mid D(NW_f(z))=1\}\succeq_e 2^{cm}\times\{x<2^{\lfloor 2^{\delta m}\rfloor}\mid D(x)=1\},$$

where $e:=\lceil 2^{\delta m}\rceil/2^{\epsilon m}$ and $NW_f(z):=NW(z,f)$."

Theorem 6 shows that $\Pr^y_x[D(x)=1]_\theta$ is S^1_2-provably similar to $\Pr^y_z[D(NW_f(z))=1]_\theta$, for $\theta^{-1}\in Log$. To see this, note that

$$2^{cm}\Pr_z^y[D(NW_f(z))=1]_\theta\approx_\theta\{z<2^{cm}\mid D(NW_f(z))=1\}.$$

Hence, by Proposition 2 *iii)*,

$$2^{\lfloor 2^{\delta m}\rfloor}2^{cm}\Pr_z^y[D(NW_f(z))=1]_\theta\succeq_\theta 2^{\lfloor 2^{\delta m}\rfloor}\times\{z<2^{cm}\mid D(NW_f(z))=1\}.$$

Similarly,

$$2^{cm}\times\{x<2^{\lfloor 2^{\delta m}\rfloor}\mid D(x)=1\}\succeq_\theta 2^{\lfloor 2^{\delta m}\rfloor}2^{cm}\Pr_x^y[D(x)=1]_\theta.$$

Therefore, by Proposition 2 *ii*), the conclusion of Theorem 6 implies

$$2^{\lfloor 2^{\delta m}\rfloor} 2^{cm} \Pr_z^y[D(NW_f(z)) = 1]_\theta \succeq_{2\theta+e} 2^{\lfloor 2^{\delta m}\rfloor} 2^{cm} \Pr_x^y[D(x) = 1]_\theta.$$

If the size of D is $\leq 2^{\epsilon m} - \lceil 2^{(\delta+\gamma)m} \rceil - 2$, the same inequality holds for $\neg D$ instead of D.

3 Feasible anticheckers.

If there is an n^k-size circuit computing SAT_n, there is a $poly(n^k)$-size circuit B with n inputs and $\leq n$ outputs such that $\forall x, y \in \{0,1\}^n, (\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, B(x)))$. We use this to formulate the existence of anticheckers for SAT as a $\forall \Pi_1^b$ statement.

Theorem 7 (‘CC ← PC’ from feasible anticheckers)**.**
Let $k \geq 3$ be a constant and assume that there is a p-time function f such that S_2^1 proves:

“$\forall 1^n$, $f(1^n)$ *is a* $poly(n^k)$*-size circuit* B *such that*

$$\forall x, y \in \{0,1\}^n, [\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, B(x))]$$

or $\big(f(1^n)$ *outputs sets* $A_n^{\mathsf{SAT}_n, n^k}, A' \subseteq \{0,1\}^n$, $D \subseteq A_n^{\mathsf{SAT}_n, n^k} \times A'$ *of size* $poly(n^k)$ *such that*

$$\forall x \in A_n^{\mathsf{SAT}_n, n^k} [\exists y_x \in A', \langle x, y_x \rangle \in D \wedge \forall z, y \in A', (\langle x, y \rangle \in D \wedge \langle x, z \rangle \in D \rightarrow y = z)]$$

and $\forall$ n^k*-size circuit* C,

$$\forall x \in A_n^{\mathsf{SAT}_n, n^k} \forall y \in \{0,1\}^n \ [\mathsf{SAT}_n(x,y) \rightarrow \mathsf{SAT}_n(x, y_x)] \wedge$$

$$\exists x \in A_n^{\mathsf{SAT}_n, n^k}, \mathsf{SAT}_n(x, y_x) \neq C(x)\big).”$$

Then, proving that EF *is not p-bounded implies* $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^k]$ *for infinitely many* n.

Proof. The statement assumed to have an S_2^1-proof is $\forall \Pi_1^b$, so there are p-size EF-proofs of its propositional translation. If we now assume that $\exists n_0 \forall n > n_0, \mathsf{SAT}_n \in \mathsf{Circuit}[n^k]$, there are circuits C and $y \in \{0,1\}^n$ falsifying the second disjunct of the translated assumption for $n > n_0$. Consequently, EF proves efficiently that the circuits generated by $f(1^n)$ solve SAT_n, which implies that EF is p-bounded. □

Existential quantifiers instead of witnessing. If we used the existential quantifiers instead of function f in Theorem 7, the resulting statement S formalizing the existence of anticheckers would be $\forall \Sigma_2^b$. By the KPT theorem [23], PV_1-provability of S would then imply the existence of p-time functions $f_1, \ldots, f_c$, for a constant c, with a PV_1-proof of:

"$\forall 1^n$, $\forall x^1,\dots,x^c,y^1,\dots,y^c,\tilde{y}^1,\dots,\tilde{y}^c\in\{0,1\}^n$, $\forall n^k$-size circuits $C^1,\dots,C^c$, $f_1(1^n)$ outputs a $poly(n^k)$-size circuit B and $A_n^{\mathsf{SAT}_n,n^k}$, $A'\subseteq\{0,1\}^n$, $D\subseteq A_n^{\mathsf{SAT}_n,n^k}\times A'$ of size $poly(n^k)$ such that the following predicate, denoted $P_{f_1}(x^1,y^1,C^1,\tilde{y}^1)$, holds:

$$\big(\mathsf{SAT}_n(x^1,y^1)\rightarrow\mathsf{SAT}_n(x^1,B(x^1))\big)\vee\big(D'(\tilde{y}^1)\wedge\exists\tilde{x}\in A_n^{\mathsf{SAT}_n,n^k},\mathsf{SAT}_n(\tilde{x},y_{\tilde{x}})\neq C(\tilde{x})\big),$$

where $D'(\tilde{y}^1)$ stands for the remaining part of the Σ_0^b subformula of S,
or $f_2(1^n,x^1,y^1,C^1,\tilde{y}^1)$ outputs a $poly(n^k)$-size circuit B and $A_n^{\mathsf{SAT}_n,n^k}$, A',D of size $poly(n^k)$ such that $P_{f_2}(x^2,y^2,C^2,\tilde{y}^2)$ holds, or
...
or $f_c(1^n,x^1,\dots,x^{c-1},y^1,\dots,y^{c-1},C^1,\dots,C^{c-1},\tilde{y}^1,\dots,\tilde{y}^{c-1})$ outputs a $poly(n^k)$-size circuit B and $A_n^{\mathsf{SAT}_n,n^k}$, A',D of size $poly(n^k)$ such that $P_{f_c}(x^c,y^c,C^c,\tilde{y}^c)$."

The resulting $\forall\Pi_1^b$-statement could be translated to propositional tautologies with p-size EF-proofs. However, given $\forall n,\mathsf{SAT}_n\in\mathsf{Circuit}[n^k]$, we could not directly obtain p-size EF-proofs of tautologies stating that one of the functions $f_1,\dots,f_c$ generates a circuit solving SAT_n. This is because B and $A_n^{\mathsf{SAT}_n,n^k}$ generated by f_2 depend on y^1. For the same reason, it seems possible for EF to prove efficiently $\mathsf{SAT}\in\mathsf{P/poly}$ (using the formalization based on the KPT witnessing) without proving efficiently all tautologies.

4 One-way functions from $\mathsf{NP}\not\subseteq\mathsf{P/poly}$

Theorem 8 ('CC $\leftarrow$ PC' from 'OWF $\leftarrow$ $\mathsf{NP}\not\subseteq\mathsf{P/poly}$' & hardness of E)**.**
Assume that for each sufficiently big n, each $2^{n/4}$-size circuit fails to compute $h'\in\mathsf{E}$ on $\geq(1/2-1/2^{n/4})$ of all inputs. Further, assume that there is a p-time function $h:\{0,1\}^n\mapsto\{0,1\}^{u(n)}$ such that for each constants c,d, there is a p-time function f_2 and a constant $0<\epsilon<1$ such that S^1_2 proves:

"$\forall n,\forall$ cn^c-size circuit C with $u(m)$ inputs and m outputs such that $n\leq dm^d$, $\big(f_2(C)$ is a $poly(n)$-size circuit B such that

$$\forall x,y\in\{0,1\}^{\lfloor n^\epsilon\rfloor},[\mathsf{SAT}_{\lfloor n^\epsilon\rfloor}(x,y)\rightarrow\mathsf{SAT}_{\lfloor n^\epsilon\rfloor}(x,B(x))]$$

or

$$\Pr^{y}_{x\in\{0,1\}^m}[h(C(h(x)))=h(x)]_{\frac{1}{m}}<1/2\big)."$$

Then, proving that $\mathsf{EF}+\mathsf{tt}(h',2^{n/4},1/2-1/2^{n/4})$ is not p-bounded implies $\mathsf{SAT}\notin\mathsf{P/poly}$.

The system $\mathsf{EF}+\mathsf{tt}(h',2^{n/4},1/2-1/2^{n/4})$ is defined in the same way as in the introduction. That is, $\mathsf{EF}+\mathsf{tt}(h',2^{n/4},1/2-1/2^{n/4})$ is an extension of EF which can use substitutional instances of $\mathsf{tt}(h'_n,2^{n/4},1/2-1/2^{n/4})$, for sufficiently big n, in its proofs.

The proof of Theorem 8 is based on the following lemma formalizing a conditional witnessing of $\mathsf{NP}\not\subseteq\mathsf{P/poly}$, cf. [28, 26].

Lemma 2 (Formalized witnessing of $\mathsf{NP} \not\subseteq \mathsf{P/poly}$ from OWF & hardness of E)**.**
Let $k \geq 1$ be a constant. For each p-time functions $h : \{0,1\}^n \mapsto \{0,1\}^{u(n)}$ and f_1, there are p-time functions f_0, f_{-1}, f_{-2} and constants b, n_1 such that $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$ proves: "*$\forall 1^n > n_1$, $\forall m$ such that $n/2^b \leq 2^{bm} \leq n$, if*

$$\mathsf{LB_{tt}}'(f_1(1^{2^m}), 2^{m/4}, 2^m(1/2 - 1/2^{m/4})),$$

then $f_0(1^n, m)$ outputs sets $A, A' \subseteq \{0,1\}^n$ of size poly(n) such that

$$\forall x \in A\ \exists y_x \in A'\ \mathsf{SAT}_n(x, y_x)$$

and $\big(\forall\ n^k$-size circuit C with n inputs and $\leq n$ outputs,

$$\exists x \in A, \neg \mathsf{SAT}_n(x, C(x))$$

or $f_{-1}(C, m)$ outputs a poly(n)-size circuit C' with $u(n')$ inputs and n' outputs, where $n \leq f_{-2}(1^{n'})$, such that

$$\Pr^{y}_{x \in \{0,1\}^{n'}}[h(C'(h(x))) = h(x)]_{\frac{1}{n'}} \geq 1/2\big)."$$

Here, $\mathsf{LB_{tt}}'$ is obtained from $\mathsf{LB_{tt}}$ by setting $m_0 = 0$ and skipping the universal quantifier on m (so m in $\mathsf{LB_{tt}}'$ is the same as the universally quantified m in the S^1_2-provable statement).

Proof of Theorem 8 from Lemma 2. Intuitively, Theorem 8 assumes that the hardness of SAT yields a function h which is hard to invert. Lemma 2 shows that such h can be used to find an error of each small circuit attempting to compute SAT. Combining the assumption of Theorem 8 with Lemma 2 we obtain a p-time function f such that for each small circuit C, either C solves SAT or $f(C)$ finds an error of C. Moreover, this holds provably in $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$, so the propositional translation of the correctness of the witnessing statement has short proofs in $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$. This will allow us to derive the desired implication similarly as in the proof of Theorem 1.

We proceed with a formal proof.

The assumption of Theorem 8 in combination with Lemma 2 implies that for each $k \geq 1$, for p-time f_1 generating the truth-table of h', there is $0 < \epsilon < 1$ and b, n_1 such that $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$ proves the following statement S:

"$\forall 1^n > n_1$, $\forall m, n/2^b \leq 2^{bm} \leq n$, if

$$\mathsf{LB_{tt}}'(f_1(1^{2^m}), 2^{m/4}, 2^m(1/2 - 1/2^{m/4})),$$

then $f_0(1^n, m)$ outputs $A, A' \subseteq \{0,1\}^n$ such that

$$\forall x \in A\ \exists y_x \in A'\ \mathsf{SAT}_n(x, y_x)$$

and $\big(\forall n^k$-size circuit C with n inputs and $\leq n$ outputs,

$$\exists x \in A, \neg \mathsf{SAT}_n(x, C(x))$$

or $f_2(f_{-1}(C,m))$ outputs a circuit B such that

$$\forall x, y \in \{0,1\}^{\lfloor n^\epsilon \rfloor}, [\mathsf{SAT}_{\lfloor n^\epsilon \rfloor}(x,y) \to \mathsf{SAT}_{\lfloor n^\epsilon \rfloor}(x, B(x))]$$

or y' does not satisfy the assumption of $\Pr_x^{y'}[\cdot]_{1/n'} \geq 1/2\big)$."

Since S is $\forall\Sigma_1^b$, by Lemma 1, there is a p-time function f_3 and a constant ℓ such that PV_1 proves: "$\forall 1^n > n_1, \forall m, n/2^b \leq 2^{bm} \leq n$, if $|h'| \geq n^\ell$ and a $2^{||h'||/4}$-size circuit generated by a p-time function fails to compute h', then $f_3(1^n, m, h', C, x, y, y')$ outputs a circuit falsifying

$$\mathsf{LB_{tt}}'(f_1(1^{2^m}), 2^{m/4}, 2^m(1/2 - 1/2^{m/4})),$$

or $f_3(1^n, m, h', C, x, y, y')$ outputs a circuit falsifying the assumption of $\Pr_x^{y'}[\cdot]_{1/n'} \geq 1/2$ or F' holds," where F' is the rest of the statement S.

Consequently, EF proves efficiently the propositional translation of the PV_1-theorem. Substituting h' for y', $\mathsf{EF}^+ := \mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ proves the formula F encoding F'. We now proceed as in the proof of Theorem 1. Assuming that $\mathsf{SAT} \in \mathsf{P/poly}$, there is some k such that for all $1^n > n_1 \geq 1$ we can efficiently falsify the first disjunct of F. Therefore, there is a p-size circuit B such that EF^+ proves efficiently $\mathsf{SAT}_{\lfloor n^\epsilon \rfloor}(x,y) \to \mathsf{SAT}_{\lfloor n^\epsilon \rfloor}(x, B(x))$, which means that EF^+ is p-bounded. □

Proof of Lemma 2. Let $f_0(1^n, m)$ output the set of propositional formulas $A := \{\phi_z(x) \mid z \in \{0,1\}^{cm}\}$, where $\phi_z(x)$ uses free variables x together with some auxiliary variables and encodes the statement

$$h(x) = h(NW_f(z)).$$

Here, $NW_f : \{0,1\}^{cm} \mapsto \{0,1\}^{\lfloor 2^{\delta m} \rfloor}$ and c are given by Theorem 6, for some $0 < \gamma, \delta < 1$ specified later, and $f = f_1(1^{2^m})$. The size of $\phi_z(x)$ is $\lfloor 2^{K\delta m} \rfloor$, for a constant K depending only on h. We set m so that $n/2^{K\delta} \leq 2^{K\delta m} \leq n$ and treat formulas $\phi_z(x)$ as formulas of size n. Let $A' := \{NW_f(z) \mid z \in \{0,1\}^{cm}\}$.

We reason in $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$. Suppose that an n^k-size circuit C with n inputs and $\leq n$ outputs finds a satisfying assignment for all formulas in A. Then, there is an n^k-size circuit C' with $u(\lfloor 2^{\delta m} \rfloor)$ inputs such that

$$2^{\lfloor 2^{\delta m} \rfloor} \times \{z \in \{0,1\}^{cm} \mid h(C'(h(NW_f(z)))) \neq h(NW_f(z))\} \preceq_0 0. \tag{4.1}$$

The circuit C' is obtained from C by a p-time algorithm f_{-1} depending on δ, h and m. The predicate $h(C'(h(x))) \neq h(x)$ is computable by a $2^{K'\delta m}$-size circuit D with $\lfloor 2^{\delta m} \rfloor$ inputs, for a constant K' depending only on k and h. Now, we set a sufficiently small γ

and δ so that $2^{K'\delta m} \leq 2^{m/4} - \lceil 2^{(\delta+\gamma)m} \rceil - 1$. Therefore, by Theorem 6, the assumption that f is hard on average for $2^{m/4}$-size circuits implies that

$$2^{\lfloor 2^{\delta m} \rfloor} \times \{z < 2^{cm} \mid D(NW_f(z)) = 1\} \succeq_e 2^{cm} \times \{x < 2^{\lfloor 2^{\delta m} \rfloor} \mid D(x) = 1\},$$

where $e = \lceil 2^{\delta m} \rceil / 2^{m/4}$. Consequently, by (4.1) and Proposition 2 *ii*),

$$0 \succeq_e 2^{cm} \times \{x < 2^{\lfloor 2^{\delta m} \rfloor} \mid h(C'(h(x))) \neq h(x)\}. \tag{4.2}$$

We want to show that $\Pr^y_{x \in \{0,1\}^{\lfloor 2^{\delta m} \rfloor}}[h(C'(h(x))) = h(x)]_{\frac{1}{\lfloor 2^{\delta m} \rfloor}} \geq 1/2$. For the sake of contradiction, assume that this is not the case. Then, $\{x < 2^{\lfloor 2^{\delta m} \rfloor} \mid h(C'(h(x)) = h(x)\} \preceq_{1/\lfloor 2^{\delta m} \rfloor} 2^{\lfloor 2^{\delta m} \rfloor - 1}$. By Item 1 *iii*) of Proposition 3,

$$\{x < 2^{\lfloor 2^{\delta m} \rfloor} \mid h(C'(h(x)) \neq h(x)\} \succeq_{2/\lfloor 2^{\delta m} \rfloor} 2^{\lfloor 2^{\delta m} \rfloor - 1}.$$

By Proposition 2 *iii*),

$$2^{cm} \times \{x < 2^{\lfloor 2^{\delta m} \rfloor} \mid h(C'(h(x)) \neq h(x)\} \succeq_{2/\lfloor 2^{\delta m} \rfloor} 2^{\lfloor 2^{\delta m} \rfloor - 1 + cm}. \tag{4.3}$$

By Proposition 2 *ii*), (4.2) and (4.3) yield $0 \succeq_{2/\lfloor 2^{\delta m} \rfloor + e} 2^{\lfloor 2^{\delta m} \rfloor - 1 + cm}$. Hence, by Item 1 *ii*) of Proposition 3, $2^{\lfloor 2^{\delta m} \rfloor - 1 + cm} < (3/\lfloor 2^{\delta m} \rfloor + e) 2^{\lfloor 2^{\delta m} \rfloor - 1 + cm}$, which is a contradiction for a sufficiently big m. □

5 Learning from the non-existence of OWFs

Theorem 9 ('CC ← PC' from 'Learning ← $\nexists$ OWF' & hardness of E).
Let $k, t \geq 1$ be constants. Assume that for each sufficiently big n, each $2^{n/4}$-size circuit fails to compute $h' \in \mathsf{E}$ on $\geq 1/2 - 1/2^{n/4}$ of all inputs. Further, assume that there is a p-time function $h : \{0,1\}^n \mapsto \{0,1\}^{u(n)}$ such that for each constants c, d, there is a p-time function f_2 and constants n_0 and $0 < \epsilon < 1$ such that S^1_2 proves:

> *"$\forall n$, $\forall$ cn^c-size circuit C with $u(m)$ inputs and m outputs such that $n \leq dm^d$, $\big($$f_2(C)$ outputs a poly(n)-size circuit B learning $\lfloor n^\epsilon \rfloor^t$-size circuits with $\lfloor n^\epsilon \rfloor$ inputs over the uniform distribution, up to error $1/2 - 1/\lfloor n^\epsilon \rfloor$, with confidence $1/\lfloor n^\epsilon \rfloor$; formally, $\forall f : \{0,1\}^{\lfloor n^\epsilon \rfloor} \mapsto \{0,1\}$, $\forall$ $\lfloor n^\epsilon \rfloor^t$-size circuit D computing f,*
>
> $$\Pr^y_w[B(1^{\lfloor n^\epsilon \rfloor}, w) \ (1/2 + 1/\lfloor n^\epsilon \rfloor)\text{-approximates } f]_{1/2\lfloor n^\epsilon \rfloor} \geq 1/\lfloor n^\epsilon \rfloor;$$
>
> *or*
>
> $$\Pr^y_{x \in \{0,1\}^m}[h(C(h(x))) = h(x)]_{\frac{1}{m}} < 1/2\big)."$$

Then there are constants b and a (depending on $k, t, h, h', c, d, f_2, n_0, \epsilon$) such that for each n the existence of a function $g_n : \{0,1\}^n \mapsto \{0,1\}$ such that no circuit of size bn^b computes g_n on $(1/2 + 1/n)$ fraction of inputs and such that $\mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ does not have 2^{an}-size proofs of $\mathsf{tt}(g_n, n^t)$ implies that $\mathsf{SAT}_n \notin \mathsf{Circuit}[n^k]$.

Note that the S^1_2-theorem in the assumption of Theorem 9 assumes $2^{\lfloor n^\epsilon \rfloor} \in Log$.

Proof. The assumption of Theorem 9 in combination with Lemma 2 implies that for any given $k, t \geq 1$, for p-time f_1 generating the truth-table of h', there are constants $0 < \epsilon < 1$ and b, n_1 such that $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$ proves the following statement S:

"$\forall 1^n > n_1, \forall m, n/2^b \leq 2^{bm} \leq n$, if

$$\mathsf{LB_{tt}}'(f_1(1^{2^m}), 2^{m/4}, 2^m(1/2 - 1/2^{m/4})),$$

then $f_0(1^n, m)$ outputs $A, A' \subseteq \{0,1\}^n$ such that

$$\forall x \in A\ \exists y_x \in A'\ \mathsf{SAT}_n(x, y_x)$$

and $\big(\forall n^k$-size circuit C with n inputs and $\leq n$ outputs,

$$\exists x \in A, \neg \mathsf{SAT}_n(x, C(x))$$

or $f_2(f_{-1}(C, m))$ outputs a circuit B such that $\forall f : \{0,1\}^{\lfloor n^\epsilon \rfloor} \mapsto \{0,1\}$, $\forall\ \lfloor n^\epsilon \rfloor^t$-size circuit D computing f,

$$\Pr^y_w[B(1^{\lfloor n^\epsilon \rfloor}, w)\ (1/2 + 1/\lfloor n^\epsilon \rfloor)\text{-}approximates\ f]_{1/2\lfloor n^\epsilon \rfloor} \geq 1/\lfloor n^\epsilon \rfloor,$$

or y' does not satisfy the assumption of $\Pr^{y'}_x[\cdot]_{1/n'} \geq 1/2\big)$."

Since S is $\forall \Sigma^b_1$, by Lemma 1, there is a p-time function f_3 and a constant ℓ such that PV_1 proves: "$\forall 1^n > n_1, \forall m, n/2^b \leq 2^{bm} \leq n$, if $|h'| \geq 2^{\ell \lfloor n^\epsilon \rfloor}$ and a $2^{||h'||/4}$-size circuit generated by a p-time function fails to compute h', then $f_3(1^n, m, h', C, f, D, y, y')$ outputs a circuit falsifying

$$\mathsf{LB_{tt}}'(f_1(1^{2^m}), 2^{m/4}, 2^m(1/2 - 1/2^{m/4})),$$

or $f_3(1^n, m, h', C, f, D, y, y')$ outputs a circuit falsifying the assumption of $\Pr^{y'}_x[\cdot]_{1/n'} \geq 1/2$ or it outputs a circuit falsifying the assumption of $\Pr^y_w[\cdot]_{1/2\lfloor n^\epsilon \rfloor} \geq 1/\lfloor n^\epsilon \rfloor$ or F' holds," where F' is the rest of the statement S.

Consequently, $\mathsf{EF}^+ := \mathsf{EF} + \mathsf{tt}(h', 2^{n/4}, 1/2 - 1/2^{n/4})$ proves efficiently the propositional translation of F'. If we now fix $n > 1$ and assume that $\mathsf{SAT}_n \in \mathsf{Circuit}[n^{k''}]$, then there is some $k = O(k'')$ such that we can efficiently falsify the first disjunct of the propositional translation of F' in EF^+. Therefore, there is a $poly(n)$-size circuit B and a 2^{Kn}-size EF^+

proof of $\mathsf{lear}^{h'}_{1/2n}(B, \mathsf{Circuit}[n^t], 1/2 - 1/n, 1/n)$, for a constant K independent of n. Recall that this means that EF^+ proves efficiently $\neg\mathsf{tt}(f, n^t) \to R$, for a formula R.

Let $b \geq t$ be such that B has size $\leq bn^b$. We claim that for each Boolean function $g_n : \{0,1\}^n \mapsto \{0,1\}^n$ which is not $(1/2 + 1/n)$-approximable by any circuit of size bn^b, there is a 2^{an}-size EF^+-proof of $\mathsf{tt}(g_n, n^t)$, for a constant a independent of n. This is because in order to prove $\mathsf{tt}(g_n, n^t)$ in EF^+, it suffices to check in EF^+ that $\neg R$ holds for $f = g_n$. $\neg R$ holds for $f = g_n$ as otherwise there would be a bn^b-size circuit $(1/2 + 1/n)$-approximating g_n. Moreover, the fact that $\neg R$ holds for $f = g_n$ is efficiently provable in EF^+ as w.l.o.g. $\neg R$, for $f = g_n$, does not contain any free variables (we can assume that the auxiliary variables are substituted by suitable constants). □

6 Self-provability from random self-reducibility

We show that the random self-reducibility of the discrete logarithm problem can be used to derive a conditional self-provability of the statement that the discrete logarithm problem can be solved by p-size circuits.

For simplicity, we consider the discrete logarithm problem for $\mathbb{Z}_q^\times$, multiplicative groups of integers modulo a prime q. Let G be such a cyclic group. Then there are p-time algorithms A_1, A_2 such that $A_1(g, h, q) = g \cdot h \in G$, for $g, h \in G$, and $A_2(g, q) = g^{-1}$, for $g \in G$. That is, A_1 (given q) defines the multiplication of two elements in G and A_2 outputs the inverse of each $g \in G$.

The discrete logarithm problem for a cyclic group G generated by g is defined as follows. Given $b \in G$, we want to find a such that $g^a = b$. The discrete logarithm problem is 'random self-reducible': If we have a circuit C which solves the problem for a p-fraction of all $b \in G$, we can turn it efficiently into a randomized circuit C' which solves the problem on each $b \in G$ with probability $\geq p$. The circuit C' interprets its random bits r as $r \in [|G|]$. Then C' applies C on bg^r. Since bg^r is a uniformly random element of G, C succeeds in finding ℓ such that $g^\ell = bg^r$ with probability $\geq p$. Finally, C' outputs $\ell - r$, which is the correct answer with probability $\geq p$. In other words, the following implication holds

$$\Pr_{b \in G}[g^{C(b)} = b] \geq p \to \forall b \in G, \Pr_{r \in [|G|]}[g^{C'(b,r)} = b] \geq p, \tag{6.1}$$

where C' is generated from C and g by a p-time function.

We want to express (6.1) by a propositional formula. To do so, we approximate probabilities by a Nisan-Wigderson generator based on a hard function $f \in \mathsf{E}$. Fix a constant k and assume that $q \in (2^{n-1}, 2^n]$, $n \in \mathbb{N}$. Then, for each n^k-size circuit C, the predicate $g^{C(b)} = b$ with input b and the predicate $g^{C'(b,r)} = b$ with input r are computable respectively by circuits D_1 and D_2 with n inputs and size $poly(n)$. Circuits D_1, D_2 reject all inputs not in G, so in particular $\Pr_{r \in [|G|]}[g^{C'(b,r)} = b] \leq 2 \Pr_{r \in \{0,1\}^n}[D_2(r) = 1]$. Further, for each $\epsilon < 1$, there are constants c', c'' and $poly(n)$-time computable generator $NW_f : \{0,1\}^{c'\lceil\log n\rceil} \mapsto \{0,1\}^n$ such that if $f : \{0,1\}^{c''\lceil\log n\rceil} \mapsto \{0,1\}$ is hard to $(1/2+1/2^{\epsilon c''\lceil\log n\rceil})$-approximate by circuits of size $2^{\epsilon c''\lceil\log n\rceil}$, then

$$\left|\Pr_{z\in\{0,1\}^{c'\lceil\log n\rceil}}[D_1(NW_f(z))=1] - \Pr_{b\in\{0,1\}^n}[D_1(b)=1]\right| \leq 1/n,$$

$$\left|\Pr_{z\in\{0,1\}^{c'\lceil\log n\rceil}}[D_2(NW_f(z))=1] - \Pr_{r\in\{0,1\}^n}[D_2(r)=1]\right| \leq 1/n.$$

Therefore, if f is hard, we have

$$\Pr_{z\in\{0,1\}^{c'\lceil\log n\rceil}}[D_1(NW_f(z))=1] \geq p \rightarrow \forall b \in G, \Pr_{z\in\{0,1\}^{c'\lceil\log n\rceil}}[D_2(NW_f(z))=1] \geq \frac{p}{2} - \frac{3}{2n}. \tag{6.2}$$

The advantage of (6.2) is that it can be expressed by $poly(n)$-size tautologies $\mathsf{self}_n(p,b,C)$ with free variables for n^k-size circuits C, n-bit strings b and n-bit parameters p (among other extension variables). The tautologies have p-size proofs in some proof system which includes a p-size proof of the primality of q and a p-size proof of the fact that g is a generator of G. Here, we use the property that g generates G if and only if $g^{(q-1)/d} \not\equiv 1 \pmod q$ for every prime d dividing $q-1$. We can thus define a Cook-Reckhow propositional proof system P_ϵ as EF with the additional axioms which allow the system P_ϵ to derive any substitutional instance of $\mathsf{self}_n(p,b,C)$ and $\mathsf{tt}(f, 2^{\epsilon n}, 1/2 - 1/2^{\epsilon n})$ in a single step of the proof, for each sufficiently big n.

Theorem 10 (Self-provability for the discrete logarithm)**.**
Let k be a constant. Assume that for some $\epsilon < 1$ we have a Boolean function $f \in \mathsf{E}$ such that for each sufficiently big n, f is not $(1/2+1/2^{\epsilon n})$-approximable by any $2^{\epsilon n}$-size circuit. Let P_ϵ be the propositional proof system defined above. If there are n^k-size circuits solving the discrete logarithm problem for $\mathbb{Z}_q^\times$, where $q \in (2^{n-1}, 2^n]$, then there are p-size circuits D such that P_ϵ has p-size proofs of tautologies encoding the statement "$\forall b \in \mathbb{Z}_q^\times, g^{D(b)} = b$."

Proof. Given an n^k-size circuit B solving the discrete logarithm problem for $\mathbb{Z}_q^\times$, where $q \in (2^{n-1}, 2^n]$, P_ϵ can derive $\mathsf{self}_n(1/2-1/n, b, B)$. Since the assumption of $\mathsf{self}_n(1/2-1/n, b, B)$ is true and since it contains essentially no free variables, it can be proven efficiently in EF. (Essentially, in order to do so, it suffices to evaluate a $\mathsf{P/poly}$-predicate inside EF.) Consequently, P_ϵ proves efficiently $\forall b \in \mathbb{Z}_q^\times, g^{D(b)} = b$, for a suitable p-size circuit D obtained by simulating C' on all $z \in \{0,1\}^{c'\lceil\log n\rceil}$. □

Notably, Theorem 10 establishes a conditional equivalence between a circuit lower bound and a proof complexity lower bound (for propositional formulas which might not be tautological).

Acknowledgement

Ján Pich received support from the Royal Society University Research Fellowship URF\R1\211106 "Proof complexity and circuit complexity: a unified approach." Rahul Santhanam was partially funded by the EPSRC New Horizons grant EP\V048201\1: "Structure versus Randomness in Algorithms and Computation." Part of this work was done while the authors were visiting the Simons Institute for the Theory of Computing. For the purpose of Open Access, the authors have applied a CC BY public copyright lincence to any Author Accepted Manuscript version arising from this submission.

References

[1] Althöfer I.; *On sparse approximations to randomized strategies and convex combinations*; Linear Algebra and its Applications, 199(1):339-355, 1994.

[2] Atserias A.; *Distinguishing SAT from polynomial-size circuits, through black-box queries*; Computational Complexity Conference (CCC), 2006.

[3] Beame P., Pitassi T.; *Propositional Proof Complexity: Past, Present and Future*; Current Trends in Theoretical Computer Science, Entering the 21st Century, 42-70, 2001.

[4] Binnendyk E., Carmosino M., Kolokolova A., Ramyaa R., Sabin M.; *Learning with distributional inverters*; Algorithmic Learning Theory (ALT), 2022.

[5] Bogdanov A., Talwar K., Wan A.; *Hard instances for satisfiability and quasi-one-way functions*; ICS, 2010.

[6] Buss S.; *Bounded arithmetic*; Bibliopolis, 1986.

[7] Cobham A.; *The intrinsic computational difficulty of functions*; Proceedings of the 2nd International Congress of Logic, Methodology and Philosophy of Science, North Holland, pp. 24-30, 1965.

[8] Cook S.A.; *Feasibly constructive proofs and the propositional calculus*; Symposium on Theory of Computing (STOC), 1975.

[9] Cook S.A., Thapen N.; *The strength of replacement in weak arithmetic*; ACM Transactions on Computational Logic, 7(4):749-764, 2006.

[10] Cook S.A., Krajíček J.; *Consequences of the Provability of NP$\subseteq$P/poly*; Journal of Symbolic Logic, 72:1353-1357, 2007.

[11] de Rezende S., Göös M., Robere R.; *Proofs, Circuits and Communication*; SIGACT News Complexity Theory Column, 2022.

[12] Grochow J., Pitassi T.; *Circuit Complexity, Proof Complexity, and Polynomial Identity Testing: The Ideal Proof System*; Journal of the ACM, 65(6), 37:1-59, 2018.

[13] Gutfreund D., Shaltiel R., Ta-Shma A.; *If NP languages are hard in the worst-case then it is easy to find their hard instances*; Computational Complexity, 16(4), 412-441, 2007.

[14] Hirahara S.; *Non-black-box worst-case to average-case reductions within* NP; Foundations of Computer Science (FOCS), 2018.

[15] Impagliazzo R.; *A personal view of average-case complexity*; Structure in Complexity Theory (SCT), 1995.

[16] Jeřábek E.; *Dual weak pigeonhole principle, Boolean complexity and derandomization*; Annals of Pure and Applied Logic, 129:1-37, 2004.

[17] Jeřábek E.; *Weak pigeonhole principle and randomized computation*; Ph.D. thesis, Charles University in Prague, 2005.

[18] Jeřábek E.; *Approximate counting in bounded arithmetic*; Journal of Symbolic Logic, 72:959-993, 2007.

[19] Kannan R.; *Circuit-size lower bounds and non-reducibility to sparse sets*, Information and Control, 55:40-56, 1982.

[20] Krajíček J.; *Bounded arithmetic, propositional logic, and complexity theory*; Cambridge University Press, 1995.

[21] Krajíček J.; *On the weak pigeonhole principle*; Fundamenta Mathematicae, 170(1-3):123-140, 2001.

[22] Krajíček J.; *Proof complexity*; Cambridge University Press, 2019.

[23] Krajíček J., Pudlák P., Takeuti G.; *Bounded arithmetic and the polynomial hierarchy*; Annals of Pure and Applied Logic, 52:143-153, 1991.

[24] Lipton R.J., Young N.E.; *Simple strategies for large zero-sum games with applications to complexity theory*; Symposium on Theory of Computing (STOC), 1994.

[25] Liu Y., Pass R.; *On one-way functions from* NP*-complete problems*; Computational Complexity Conference (CCC), 2022.

[26] Müller M., Pich J.; *Feasibly constructive proofs of succinct weak circuit lower bounds*; Annals of Pure and Applied Logic, 2019.

[27] Newman I.; *Private vs common random bits in communication complexity*; Information Processing Letters, 39:67-71, 1991.

[28] Pich J.; *Circuit lower bounds in bounded arithmetics*; Annals of Pure and Applied Logic, 166(1):29-45, 2015.

[29] Pich J., Santhanam R.; *Learning algorithms versus automatability of Frege systems*; arXiv, 2021.

[30] Razborov A.A.; *Bounded arithmetic and lower bounds in boolean complexity*; Feasible Mathematics II, 344-386, 1995.

[31] Razborov A.A.; *Pseudorandom generators hard for k-DNF Resolution and Polynomial Calculus*; Annals of Mathematics, 181(2):415-472, 2015.

[32] Santhanam R.; *Pseudorandomness and the Minimum Circuit Size Problem*; Innovations in Theoretical Computer Science (ITCS), 2020.

[33] Williams R.; *Non-uniform ACC circuit lower bounds*; Computational Complexity Conference (CCC), 2011.