Strong Co-Nondeterministic Lower Bounds for NP cannot be Proved Feasibly

Ján Pich*
Czech Academy of Sciences
& University of Oxford

Rahul Santhanam†
University of Oxford

March 2021

Abstract

We show unconditionally that Cook's theory PV_1 formalizing poly-time reasoning cannot prove, for any non-deterministic poly-time machine M defining a language $L(M)$, that $L(M)$ is inapproximable by co-nondeterministic circuits of sub-exponential size. In fact, our unprovability result holds also for a theory which supports a fragment of Jeřábek's theory of approximate counting APC_1. We also show similar unconditional unprovability results for the conjecture of Rudich about the existence of super-bits.

1 Introduction

It is widely accepted that strong complexity lower bounds are difficult to prove. Despite significant efforts over the past half a century and more, we have not even made much progress on showing super-linear circuit size lower bounds for NP, let alone flagship problems in the area such as the NP vs P problem or the $\mathsf{NP} \not\subseteq \mathsf{P/poly}$ problem.

But is this difficulty fundamental, or have we simply not been lucky enough or ingenious enough? Recently the Sensitivity Conjecture, a major problem in Boolean function analysis open for nearly 30 years, was settled with a simple one-page proof [18]. Might flagship complexity lower bound problems yield similarly to clever tricks or combinations of known ideas?

Since very early in the history of complexity theory, there have been attempts to show that the difficulty in proving lower bounds is indeed fundamental. It has been speculated that the NP vs P problem might be independent of Peano Arithmetic. More relevant to the practice of complexity theory, various barriers have been formalized, such as the relativization barrier [3], the natural proofs barrier [45] and the algebraization barrier [2], to show that certain classes of techniques are unlikely to be able to show strong complexity lower bounds.

*jan.pich@cs.ox.ac.uk
†rahul.santhanam@cs.ox.ac.uk

A natural goal in this direction is to identify a logical system within which much of current-day complexity theory can be formalized, but so that strong complexity lower bounds are hard to show in the system. For example, one could consider subsystems of Peano arithmetic such as Cook's system PV_1 of bounded arithmetic formalizing polynomial-time reasoning [11], or else a standard propositional proof system such as the Extended Frege proof system, and attempt to establish that super-polynomial circuit size lower bounds for SAT are unprovable in these systems.

Such attempts have also not been successful so far. Indeed, some of the same difficulties that apply to showing strong complexity lower bounds seem also to apply to showing these same unprovability results. Just as we don't understand the power of poly-time computation well enough to separate NP from P, we don't understand the power of poly-time reasoning well enough to show that $\mathsf{NP} \not\subseteq \mathsf{P/poly}$ is unprovable using poly-time reasoning. It seems possible that we could live in the worst of all possible worlds for complexity theorists - a world where complexity lower bounds are hard but it is also difficult to understand *why* they are hard!

In this paper, we rule out this pessimistic situation for certain kinds of strong complexity lower bounds, namely for lower bounds for NP against co-nondeterministic circuits. As our main result, we show *unconditionally* that average-case lower bounds for NP against sub-exponential size co-nondeterministic circuits are unprovable in PV_1. In fact, the unprovability holds for a stronger theory $\mathsf{T}^0_{\mathsf{APC}_1}$ defined as T_{PV}, the true universal theory of natural numbers in the language consisting of function symbols for all p-time algorithms, extended by the so called dual weak pigeonhole principle for all p-time functions (without parameters[1]), cf. Section 2.1. In particular, $\mathsf{T}^0_{\mathsf{APC}_1}$ contains a significant fragment of Jeřábek's theory APC_1 which formalizes probabilistic poly-time reasoning [20, 21, 19] and proves the existence of a hard Boolean function, cf. Section 2.1. Our techniques also allow us to rule out provability in $\mathsf{T}^0_{\mathsf{APC}_1}$ of the existence of Rudich's super-bits, a version of pseudorandom generator safe against nondeterministic circuits [46]. Since much of current-day complexity theory is formalizable in PV_1 and APC_1, cf. Section 2.1, this gives evidence that simple tricks or clever combinations of known techniques are *not* sufficient to prove the complexity separations we consider. New techniques or strategies of higher logical complexity are required.

Before stating our results in more detail, we give more motivation for the specific setting we consider. Our starting point for this discussion is the influential natural proofs framework of Razborov and Rudich [45]. A natural proof against a circuit class $\mathfrak{C}$ is a property of Boolean functions (given by their truth tables) that is dense (i.e., contain a significant fraction of Boolean functions), constructive (i.e., checkable in polynomial time as a function of the size of the truth table) and elusive (i.e., implies hardness against $\mathfrak{C}$). Razborov and Rudich showed that known circuit lower bounds yielded natural proofs, in the sense that a dense, constructive and elusive property can be extracted from the lower bound proof. On the other hand, under the standard cryptographic assumption that exponentially hard one-way functions exist, natural proofs against super-polynomial size circuits do not exist. This suggests that new "non-naturalizing" techniques are required to prove the strong circuit lower bounds we desire.

The Razborov-Rudich framework is very interesting conceptually because it demonstrates

[1]The superscript 0 refers to the fact that p-time functions in the dual weak pigeonhole principle are not allowed to have hidden parameters.

the *self-defeating* nature of circuit lower bounds. The assumption about one-way functions is itself a circuit lower bound assumption, and this implies a *barrier* to actually establishing circuit lower bounds. This is reminiscent of diagonal arguments such as the ones in Gödel's celebrated Incompleteness Theorems, but in a resource-bounded setting.

However, as a meta-mathematical result, the Razborov-Rudich framework suffers from some defects. Natural proofs are computational objects, not proofs in some logical system. It is unclear precisely what it means for a proof technique to be "naturalizable". Moreover, "naturalness" is not closed under logical implication - it leaves open the possibility that some naturalizing lower bound against a weak circuit class could imply a strong non-naturalizing circuit lower bound using a reduction or some other form of logical implication. Indeed, this possibility has been explored in recent work on hardness magnification, cf. [9] and references therein.

In recent work [40], we showed that an analogue of the self-defeating phenonemon of natural proofs holds in the setting of propositional proof complexity. We considered two natural candidate distributions on formulas that are believed to be hard for any propositional proof system, namely *random circuit lower bound tautologies* and *random k-DNFs* of large enough linear size. We showed that if random circuit lower bound tautologies are indeed hard for every propositional proof system (with advice), then there *exists* a propositional proof system Q such that it is *hard* for *any* propositional proof system to prove lower bounds on the Q-proof size of these tautologies or for random k-DNFs. Thus, again, the very truth of proof complexity lower bounds for these tautologies implies that they are hard to prove.

Our motivation in [40] was to understand better why proof complexity lower bounds have been hard to show, but the results also have implications for the ability of propositional proof systems to argue about specific cases of the NP vs $\mathsf{coNP/poly}$ question. The results about random truth table tautologies in [40] have implications for the propositional provability of an implicit formulation of the average-case hardness of MCSP against co-nondeterministic circuits, and similarly the results about random k-DNFs have implications for the provability of average-case hardness of SAT.

In the present work, we are interested in pursuing further this direction concering meta-mathematics of circuit complexity - what inherent limitations do logical systems have in arguing about questions such as NP not in $\mathsf{coNP/poly}$ on average? In addressing such questions, there are significant advantages to working with the uniform setting of *theories of bounded arithmetic* rather than with the non-uniform setting of propositional proof complexity. First, as complexity theorists, our reasoning in practice is far closer to the setting of bounded arithmetic using first-order statements and induction, than to propositional proof complexity. Second, bounded arithmetic offers much more *flexibility* in formalizing and studying complexity questions, since we are allowed to deal with first-order statements with multiple quantifiers rather than with propositional tautologies. The NP vs P question, for instance, is naturally formalized as a Π_2 statement. Third, we can consider the more natural explicit formulation of a complexity-theoretic statement than the implicit formulation of [40] where the complexity-theoretic implications follow from unprovability results of an ensemble of statements rather than of a single statement. Note that dealing with the explicit formulation makes unprovability *harder* to show.

1.1 Our Results

Our main result is that average-case lower bounds for NP against sub-exponential size co-nondeterministic circuits are unprovable in $\mathsf{T}^0_{\mathsf{APC}_1}$.

Theorem 1 (Informal Statement, cf. Theorem 6)**.** *Let n_0 be any positive integer, $\delta > 0$ be any positive rational, and M be a non-deterministic Turing machine running in polynomial time. Then the following statement is unprovable in $\mathsf{T}^0_{\mathsf{APC}_1}$: "For every $n > n_0$ and every co-nondeterministic circuit D on n variables of size $\leq 2^{n^\delta}$, there are $2^n/n$ inputs x_i of length n such that for each i, $D(x_i) \neq M(x_i)$".*

Namely, for any non-deterministic machine M running in poly-time defining a language $L(M)$, $\mathsf{T}^0_{\mathsf{APC}_1}$ cannot prove that $L(M)$ cannot be approximated on almost all of its inputs by co-nondeterministic circuits of sub-exponential size. Note that this unprovability result is *unconditional* and that it holds for *any* NP language $L(M)$. This is an advantage over the results in [40] which apply to the specific NP languages MCSP and SAT, and where the formalization of the lower bound is non-standard.

Theorem 1 provides yet another illustration of the self-defeating nature of strong complexity lower bounds. The proof of Theorem 1 proceeds by contradiction - it is shown that the *provability* of the complexity lower bound statement implies its *falsehood*, which contradicts the soundness of $\mathsf{T}^0_{\mathsf{APC}_1}$.

While at a high level this argument might seem similar to the argument in [40], the details are completely different. We use the KPT witnessing theorem and Nisan-Wigderson pseudorandom generators, in contrast to [40] which builds on an idea of Razborov [43] and connections between hitting set generators and the zero-error average case hardness of MCSP.

In more detail, the proof proceeds by contradiction, and exploits a technique of Krajíček [26] (further elaborated on in [38]). If we assume that the lower bound for $L(M)$ is provable in $\mathsf{T}^0_{\mathsf{APC}_1}$, we can employ the KPT theorem [32] to conclude the existence of an interactive game witnessing errors of co-nondeterministic circuits attempting to compute $L(M)$. The next crucial step is to interpret a Nisan-Wigderson (NW) generator based on $L(M)$ as a particular co-nondeterministic circuit. Therefore, the interactive game allows us to witness errors of the NW-generator based on $L(M)$. This can be used to construct a *deterministic* circuit approximating $L(M)$, which contradicts the assumption that $L(M)$ is hard.

The similarity in the high level structure of the proof to that in [40] does suggest the presence of a more general self-defeating phenomenon, which would be interesting to investigate further. Both the argument in [40] and in this paper can be thought of as diagonalization arguments which go via pseudorandomness, one in the setting of propositional proof complexity and the other in the setting of bounded arithmetic.

Theorem 1 also has implications for proof complexity lower bounds. Consider non-uniform propositional proof systems defined as standard propositional proof systems but with proofs verifiable by p-size circuits rather than p-time algorithms. Such proof systems have been investigated by Cook-Krajíček [12]. Theorem 1 shows that $\mathsf{T}^0_{\mathsf{APC}_1}$ cannot show that any samplable distribution on formulas is hard for every non-uniform propositional proof system. This is significant since all known proof complexity lower bounds (expressed as Π^b_1 statements, cf. Section 2.1) are provable in T_{PV}.

Theorem 1 talks about average-case lower bounds for NP against co-nondeterministic circuits. Using known connections from pseudorandomness, we are also able to show that $\mathsf{T}^0_{\mathsf{APC}_1}$ cannot prove the existence of injective super-bits, a version of pseudorandom generator secure against nondeterministic adversaries introduced by Rudich [46] (cf. Section 3.4).

Theorem 2 (Informal Statement, cf. Theorem 7)**.** *Let $\{g_n\}$ be any poly-time computable family of injective functions from n bits to $n+1$ bits. $\mathsf{T}^0_{\mathsf{APC}_1}$ cannot prove that $\{g_n\}$ is a super-bit.*

Finally, in Section 4 we provide a propositional version of our results. We define a notion of KPT interpolation property, generalizing the standard notion of feasible interpolation for propositional proof systems, and show conditional lower bound for all propositional proof systems admitting KPT interpolation.

1.2 Related Work

Investigations of complexity theory in bounded arithmetic have a long history. In particular, formalizations of complexity-theoretic concepts, constructions and results started with Paris, Wilkie and Buss. We refer to [28, Section 22.5] for a robust list of references. These results involve circuit complexity but at that time there were no circuit lower bounds to consider. After the development of the first circuit lower bounds for weak circuit classes, Razborov [43] was the first to consider unprovability of strong circuit lower bounds in bounded arithmetic. He used the natural proofs framework [45] to show unprovability of circuit lower bounds for SAT in a rather weak first-order theory. These ideas formed a part of the motivation for the later development of the notion of proof complexity generators [1, 24]. Razborov has also shown, in a tour de force [44], that circuit lower bounds for SAT are not efficiently provable propositionally in k-DNF Resolution. It is, however, unclear if k-DNF Resolution is strong enough to capture any interesting reasoning with respect to the formalization of circuit lower bounds used by Razborov.

Buss [4] in his seminal work on bounded arithmetic showed that if S^1_2 proves that a predicate is in $\mathsf{NP}\cap\mathsf{coNP}$, then the predicate is already in P. This is a direct consequence of his witnessing theorem for the theory S^1_2 introduced in [4]. Buss's result, however, does not say that S^1_2 cannot prove a p-time lower bound for an $\mathsf{NP}\cap\mathsf{coNP}$ predicate. The conclusion that an S^1_2-provably $\mathsf{NP}\cap\mathsf{coNP}$ predicate is in P is derived outside S^1_2. In [26, 27] Krajíček showed that theories such as PV_1 cannot prove specific $\mathsf{NP/poly}\cap\mathsf{coNP/poly}$ lower bounds unless $\mathsf{NP/poly}\cap\mathsf{coNP/poly}$ can be approximated by circuits of subexponential size.[2] Note that in a world where $\mathsf{NP/poly}\cap\mathsf{coNP/poly}\subseteq\mathsf{P/poly}$ and $\mathsf{NP}\cap\mathsf{coNP}=\mathsf{P}$, both results would trivialize - this distinguishes them from our results, which are unconditional and concerned with separations between NP and coNP. Pich [38] adapted the result of Krajíček to show that theories below PV_1, such as VNC^1, cannot prove that SAT is hard for p-size circuits, unless standard hardness assumptions break. Our results expand on [26] and [38] by making them unconditional at the price of strengthening the lower bound which is shown to be unprovable.

[2] Krajíček's result was formulated as an unprovability of a statement expressing that some string is outside the range of a suitable Nisan-Wigderson generator, but it yields the claimed unprovability using observations from [38] or the present paper.

In more detail, we first show that the methods from [26, 38] can be adapted to the case of nondeterministic lower bounds where they achieve a 'fixpoint' yielding an unconditional result. Next, we strengthen [26, 38] by employing hardness amplification within NP. We then employ methods from pseudorandomness to establish unprovability of a version of Rudich's conjecture. We also obtain the unprovability for theory $\mathsf{T}^0_{\mathsf{APC}_1}$ which is stronger than theories considered in [26, 38]. This required a proof of the KPT theorem for $\mathsf{T}^0_{\mathsf{APC}_1}$, a theory which is not universal, cf. Section 3.1.

There are also results based on incompleteness. It follows from the second incompleteness theorem that a sufficiently strong consistent theory, e.g. ZFC if we assume its consistency, cannot prove even a superlinear lower bound on lengths of proofs in itself (seen as a proof system for proving tautologies). Otherwise, ZFC would prove its own consistency. Our unprovability result is incomparable to the impossibility of proving super-linear lower bounds on systems such as ZFC inside ZFC. This is because the nondeterministic lower bounds we show to be unprovable do not ask for a proof of consistency of any proof system, i.e. the nondeterministic lower bounds do not postulate soundness of any specific proof system, they state only that if a given proof sytem is sound then it does not have short proofs.

A similar more involved result of Krajíček [25, Corollary 3.2] says that Buss's theory of bounded arithmetic S^1_2, which formalizes p-time reasoning, cannot prove a superlinear lower bound on lengths of proofs in the so called implicit Extended Frege system. The implicit Extended Frege system is defined as an extension of the well-known Extended Frege system EF which can conclude that a formula ϕ is a tautology even if EF proves that a circuit encodes an EF-proof of ϕ.

A difference between our setting and the setting above is that in the setting above, lower bound statements pertain to *individual* formulas, while our analogous lower bound statement is in an explicit formulation and talks about the entire *language* of satisfiable formulas. In our formalization the theory 'knows' if a given propositional formula is a tautology or not. More formally, in Theorem 1 the whole truth-table of the language of satisfiable formulas of size n is encoded by a string of length 2^n, where 2^n is a length of a number[3]. This means that the theory $\mathsf{T}^0_{\mathsf{APC}_1}$ is able to reason properly about p-time concepts with respect to the input length 2^n, i.e. about $2^{O(n)}$-time concepts. Therefore, it is significantly harder to show unprovability in our setting.

Finally, it is possible to consider the provability of circuit upper bounds as well. This has been investigated systematically by Cook-Krajíček [12] and in a sequence of more recent works by Bydžovský, Krajíček, Müller and Oliveira [6, 7, 30].

1.3 Subsequent Work

After an initial version of the present paper had been circulated, Krajíček [29] answered a question we asked about the plausibility of the KPT interpolation for strong proof systems. As it turns out, the general form of the KPT interpolation defined in Section 4 fails essentially in all proof systems where the standard feasible interpolation fails. It remains possible that

[3]Later in the paper we denote this by $n \in LogLog$

a restricted version of the KPT interpolation which holds only for the formulas considered in Corollary 1 holds.

2 Preliminaries

2.1 Bounded arithmetic

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

The first theory of bounded arithmetic formalizing p-time reasoning was introduced by Cook [11] as an equational theory PV. The definition of PV is very long so we provide only a high-level description and refer to [22] for full details. The language of PV, denoted PV as well, consists of symbols for all polynomial-time algorithms introduced inductively using Cobham's characterization of polynomial-time functions, cf. [10]. Axioms and derivations of the theory are introduced simultaneously with the symbols from the language. The theory is equational, i.e. its statements assert only that two terms are equal. We work with a conservative extension of PV, denoted PV_1, which is an ordinary first-order theory in the language PV, cf. [32]. Axioms of PV_1 are universal sentences and contain all equations provable in PV. In addition, PV_1 contains axioms replacing the induction axiom for open formulas.

A PV-formula is a first-order formula in the language PV. Σ^b_0 ($=\Pi^b_0$) 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. We often write $\exists x \leq t$ instead of $\exists x, x \leq t$. Similarly for the universal quantifier.

Buss [4] introduced a theory S^1_2 extending PV_1 with the 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 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 [23], Jeřábek [21] systematically developed a theory APC_1 capturing probabilistic p-time reasoning by means of approximate counting.[4] 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

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

[4] Krajíček [23] introduced a theory BT defined as $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$ and proposed it as a theory for probabilistic p-time reasoning. The notation APC_1 comes from Buss-Kołodziejczyk-Thapen [5]. Approximate counting in bounded arithmetic was also studied already by Paris and Wilkie [36].

where f is a PV-function, which might possibly involve other parameters not explicitly shown. We will also consider a version of $dWPHP(\mathsf{PV})$ for PV-functions f which are not allowed to involve other parameters than u. We denote the resulting principle $dWPHP'(\mathsf{PV})$.

PV_1 vs. APC_1. While the theory PV_1 is already very strong and proves sophisticated results such as the PCP theorem [39], it is not known how to formalize in PV_1 some standard probabilistic arguments available in APC_1. In particular, APC_1 proves the existence of a hard Boolean function [20] which is not known to be provable in PV_1. This difference is manifested also on the provability of more involved results from complexity theory. For example, APC_1 proves $\mathsf{AC}^0, \mathsf{AC}^0[p]$ and monotone circuit lower bounds for explicit Boolean functions [34] but it is not known if these are provable in PV_1. Here, we are referring to the formalization of circuit lower bounds which does not assume $n \in LogLog$, i.e. that the whole truth table of the hard Boolean function is a length of a number. Otherwise, the mentioned lower bounds are provable already in PV_1 [42].

$dWPHP(\mathsf{PV})$ vs $dWPHP'(\mathsf{PV})$. $dWPHP(\mathsf{PV})$ and $dWPHP'(\mathsf{PV})$ are equivalent over S^1_2, cf. [47], but to the best of our knowledge the equivalence is not known to hold over PV_1. Nevertheless, $\mathsf{PV}_1 + dWPHP'(\mathsf{PV})$ still supports a substantial fragment of the machinery of approximate counting developed in APC_1. For example, $\mathsf{PV}_1 + dWPHP'(\mathsf{PV})$ proves the existence of hard Boolean functions and the main theorem of approximate counting [21, Theorem 2.7] formalizing the construction of Nisan-Wigderson generators. In fact, $\mathsf{S}^1_2 + dWPHP(\mathsf{PV})$ is $\forall\Sigma^b_1$-conservative over $\mathsf{PV}_1 + dWPHP'(\mathsf{PV})$, which means that whenever we can prove a $\forall\Sigma^b_1$-statement in $\mathsf{S}^1_2 + dWPHP(\mathsf{PV}_1)$ (or in APC_1), we can prove it also in $\mathsf{PV}_1 + dWPHP'(\mathsf{PV})$, cf. [48, 19].

Theory	Theorem	Reference
PV_1	Cook-Levin's theorem	folklore [39]
	Resolution lower bounds	[13]
	the PCP theorem	[39]
	Hardness amplification	[20]
	AC^0, $\mathsf{AC}^0[p]$ and monotone circuit lower bounds (if $n \in LogLog$)	[42]
APC_1	AC^0, $\mathsf{AC}^0[p]$ and monotone circuit lower bounds	[34]
HARD^A	Nisan-Wigderson's derandomization	[21]
	Impagliazzo-Wigderson's derandomization	[20]
	Goldreich-Levin's theorem	[15]
	Natural proofs barrier	[34]

Table 1: A list of formalizations. The theory HARD^A is essentially APC_1. More precisely, it is a conservative extension of APC_1 with a function symbol for approximate counting which allows to formulate e.g. a difference of two probabilities, cf. [21].

3 Unprovability of lower bounds in $\mathsf{T}^0_{\mathsf{APC}_1}$

3.1 Theories $\mathsf{T}_{\mathsf{APC}_1}$, $\mathsf{T}^0_{\mathsf{APC}_1}$ and the KPT theorem

The theory $\mathsf{T}_{\mathsf{APC}_1}$ is defined as $\mathsf{T}_{\mathsf{PV}} + dWPHP(PV)$. Here, T_{PV} is the true universal theory of natural numbers in the language of PV_1, i.e. the theory of all universal statements which are true in the natural numbers and use the language consisting of function symbols for all p-time algorithms. The theory $\mathsf{T}^0_{\mathsf{APC}_1}$ is defined as $\mathsf{T}_{\mathsf{PV}} + dWPHP'(PV)$, i.e. using dual weak pigeonhole principle which does not allow parameters.

T_{PV} is much stronger than Cook's theory PV_1. For example, T_{PV} proves the consistency of Peano arithmetic while Gödel's incompleteness theorem shows that PV_1 cannot prove it. Further, T_{PV} proves the reflection principle for all propositional proof systems.

One of the main properties of T_{PV}, and the only one we will use, stems from the KPT theorem, which is a form of Herbrand's theorem.

Theorem 3 (KPT [32])**.** *Let T be a universal theory over a language L which contains at least one constant or function symbol. Let $\phi(x, y, z)$ be an open (i.e. quantifier-free) L-formula and suppose that $T \vdash \forall x \exists y \forall z \phi(x, y, z)$. Then there are finitely many L-terms $t_1(x), t_2(x, z_1), \ldots, t_l(x, z_1, \ldots, z_{l-1})$ (containing only the displayed variables) such that*

$$T \vdash \forall x, z_1, \ldots, z_l;\ \phi(x, t_1(x), z_1) \vee \phi(x, t_2(x, z_1), z_2) \vee \cdots \vee \phi(x, t_l(x, z_1, \ldots, z_{l-1}), z_l).$$

If $T = \mathsf{T}_{\mathsf{PV}}$, then the terms $t_1, \ldots, t_l$ from the KPT theorem are p-time functions. The KPT theorem can be applied in the case of $\mathsf{T}_{\mathsf{APC}_1}$ as well.[5]

Theorem 4. *Assume that $\mathsf{T}_{\mathsf{APC}_1} \vdash \forall x \exists y \forall z \phi(x, y, z)$ for an open PV-formula $\phi(x, y, z)$. Then, there is a constant l and l randomized p-time functions $f_1(x), f_2(x, z_1), \ldots, f_l(x, z_1, \ldots, z_{l-1})$ such that for every $x, z_1, \ldots, z_l$, with probability $\geq 1/poly(|x|, |z_1|, \ldots, |z_{l-1}|)$ over the internal randomness of $f_1, \ldots, f_l$,*

$$\phi(x, f_1(x), z_1) \vee \phi(x, f_2(x, z_1), z_2) \vee \cdots \vee \phi(x, f_l(x, z_1, \ldots, z_{l-1}), z_l).$$

Moreover, if $\mathsf{T}^0_{\mathsf{APC}_1} \vdash \forall x \exists y \forall z \phi(x, y, z)$ for an open PV-formula $\phi(x, y, z)$, then there is a constant l' and l' functions $f'_1(x), f'_2(x, z_1), \ldots, f'_{l'}(x, z_1, \ldots, z_{l'-1})$ computable by p-size circuits such that for every $x, z_1, \ldots, z_{l'}$,

$$\phi(x, f'_1(x), z_1) \vee \phi(x, f'_2(x, z_1), z_2) \vee \cdots \vee \phi(x, f'_{l'}(x, z_1, \ldots, z_{l'-1}), z_{l'}).$$

Proof. We skolemize $\mathsf{T}_{\mathsf{APC}_1}$ and then apply the original KPT theorem.

Define a theory T as T_{PV} extended with a witnessed version of $dWPHP(\mathsf{PV})$,

$$x > 0 \to h(x, y) < x(|y| + 1) \wedge \forall u < x|y|,\ e(u) \neq h_e(x, y), \tag{3.1}$$

where h_e is a new function symbol corresponding to PV-function e.

[5] It was pointed out to us by Krajíček that a version of KPT theorem for APC_1 can be obtained via skolemization.

If $\mathsf{T}_{\mathsf{APC}_1}$ proves $\forall x \exists y \forall z \phi(x,y,z)$, then T proves the formula too. Since T is a universal theory, by Theorem 3, there are finitely many terms $t_1, \ldots, t_l$ (in the language of T) such that

$$\forall x, z_1, \ldots, z_l;\ \phi(x, t_1(x), z_1) \vee \phi(x, t_2(x, z_1), z_2) \vee \cdots \vee \phi(x, t_l(x, z_1, \ldots, z_{l-1}), z_l). \tag{3.2}$$

Each term t_i is obtained by a composition of finitely many PV-function symbols and finitely many function symbols h_e.[6] That is, each t_i defines a p-time function g_i querying oracles h_e a constant k number of times. Consider a computation of g_i on a given input. Replace each query $h_e(a,b)$ by random bits representing a number $< a(|b|+1)$. This defines partial functions h'_e. The resulting computation defines a computation of a randomized p-time function f_i. The probability that random $r < x(|y|+1)$ satisfies $\forall u < x|y|, e(u) \neq r$ is $\geq 1/(|y|+1)$. Thus, the probability that all partial functions h'_e can be extended to functions h_e satisfying $dWPHP(e)$ is at least $1/poly(|x|, |z_1|, \ldots, |z_{l-1}|)^k$. Since terms $t_1, \ldots, t_l$ satisfy (3.2) for all h_e satisfying $dWPHP(e)$, it follows that for each $x, z_1, \ldots, z_l$, the probability that

$$\phi(x, f_1(x), z_1) \vee \phi(x, f_2(x, z_1), z_2) \vee \cdots \vee \phi(x, f_l(x, z_1, \ldots, z_{l-1}), z_l)$$

holds, is at least $1/poly(|x|, |z_1|, \ldots, |z_{l-1}|)^{kl}$.

The 'moreover' part is proved by replacing queries to h_e by strings outside the range of e, which are provided as a nonuniform advice. A complication is that for each input a, b of h_e we need a priori a different advice. Fortunately, it turns out that a single advice can serve well for many inputs a, b so that in the end we need only polynomially many different strings. We will, however, need an additional assumption that h_e does not have other parameters (i.e. inputs) than the ones displayed in definition (3.1). This is guaranteed by replacing $\mathsf{T}_{\mathsf{APC}_1}$ by $\mathsf{T}^0_{\mathsf{APC}_1}$.

Specifically, if we consider g_i on a fixed input length n, there is an absolute polynomial p such that all queries of g_i (on input length n) to h have length $\leq p(n)$. For each b such that $|b| \leq p(n)$, partition the interval $[0, 2^{p(n)+1}|b|]$ into the interval $[0, 2|b||b|]$ and $poly(n)$ intervals of the form $[c|b|, (c + c/|b| - 1)|b|]$ for $c \geq 2|b|$. To see that this is possible, note that $c + c/|b| - 1 \geq c + c/2|b|$ for $c \geq 2|b|$ and since $(1 + 1/2|b|)^{2|b|} \geq 2$, $poly(n)$ intervals $[c|b|, (c + c/|b| - 1)|b|]$ suffice to cover interval $[c|b|, 2c|b|]$. Now, given a query to h on input a, b we will respond with a fixed advice $h(a,b) < a(|b|+1)$ assigned to the interval $[c|b|, (c + c/|b| - 1)|b|]$ containing $a|b|$. By the pigeonhole principle there exists a single advice $< c(|b|+1)$ working for each a, b such that $a|b| \in [c|b|, (c + c/|b| - 1)|b|]$. For each a, b such that $a|b| \in [0, 2|b||b|]$ we will use a new advice $h(a,b)$. As there are $\leq p(n)$ possible values $|b|$ on input length n and for each of them only $poly(n)$ intervals covering the whole range $[0, 2^{p(n)+1}|b|]$ of possible values of $a|b|$, the total number of strings of advice we need is $poly(n)$. They can be hardwired into a p-size circuit simulating the computation of g_i by accessing efficiently the right string for each a, b.

Note that the argument would not go through if h_e contained additional parameters besides a and b, corresponding to x and y in (3.1), since for each such input we would need to get a different set of advice. □

[6] Importantly, terms of T are not constructed using limited recursion on notation as in Cobham's definition of p-time which is used to define function symbols of PV_1 (and, hence, a subset of function symbols of T). This would result in terms using h_e a polynomial number of times and make the subsequent estimations of probabilities that $f_1, \ldots, f_l$ work exponentially small.

3.2 Formalizing non-deterministic lower bounds

Let M be a nondeterministic Turing machine running in time $s \leq 2^{O(n)}$. Abusing the notation, we denote by $M(x, y)$ the $poly(s, n)$-time predicate representing the computation of M on input $x \in \{0,1\}^n$ with nondeterministic bits $y \in \{0,1\}^s$.

Given a constant n_0, rational $\delta \in (0,1)$, and a $2^{O(n)}$-time constructible function $m(n) \leq 2^n$, we can write down a $\forall\Sigma^b_2$ formula $\mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}(M, 2^{n^\delta}, m)$ stating that for all $n > n_0$ every co-nondeterministic circuits of size 2^{n^δ} makes at least m errors when computing the problem defined by M. More precisely, $\mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}(M, 2^{n^\delta}, m)$ states the following.

For each $n_0 < n \in LogLog$, and each co-nondeterministic circuit $D(x, z)$ with n-bit input x, nondeterministic inputs z, and size 2^{n^δ}, there are m tuples x_i, y_i, z_i, $i \in [m]$ witnessing that at least m n-bit inputs x satisfy $M(x) \neq D(x)$. That is, for every $i \in [m]$,

$$\forall y'_i, z'_i \ (M(x_i, y_i) = 1 \wedge D(x_i, z_i) = 0) \vee (M(x_i, y'_i) = 0 \wedge D(x_i, z'_i) = 1)$$

Note that the displayed formula is equivalent to the formula in which the universal quantifiers $\forall y'_i, z'_i$ are moved inside the disjunction so that they appear in front of the second disjunct. The notation $n \in LogLog$ is a shortcut for '$n = ||u||$ for some u', where $||\cdot||$ is a double application of the length function $|\cdot|$, i.e. $n \in LogLog$ stands for '2^n is a length of some u'. Since the length of all quantified numbers in the formula above are bounded by $2^{O(n)}$, the assumption $n \in LogLog$ implies that these quantifiers are bounded (even if we do not display these bounds above) and $\mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}(M, 2^{n^\delta}, m)$ is indeed a $\forall\Sigma^b_2$ formula, resp. in the language of T_{PV} it is a formula of the form $\forall\exists\forall\phi$ for an open formula ϕ representing a p-time predicate. The choice of n_0 will be always clear from the context so we do not display it in $\mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}$. The subscript tt refers to the fact that $n \in LogLog$ and so the whole truth-table of M is a feasible object from the perspective of the theory in which we are working with formula $\mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}$.

Note also that we do not specify what is $|z_i|, |z'_i|$ and $|y_i|, |y'_i|$, these can be arbitrary and are bounded just by the size of D and M respectively. Further, the assumption $n \in LogLog$, which means that 2^n is a length of some number, also implies that p-time function symbols from the language of PV_1 given such a number of length 2^n as input can express properties such as $M(x, y) = 1$ even though M runs in exponential time in n.

3.3 Unprovability of strong co-nondeterministic lower bounds

First, we show the unprovability of strong average-case co-nondeterministic lower bounds for nondeterministic predicates.

Theorem 5. *Let $n_0, \delta \in (0,1)$ be rational constants and M be a nondeterministic Turing machine running in time $2^{n^{o(1)}}$. Then,*

$$\mathsf{T}^0_{\mathsf{APC}_1} \nvdash \mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}(M, 2^{n^\delta}, 2^n/2 - \tfrac{2^n}{2^{n^\delta}}).$$

Proof. To prove the theorem we build on the conditional unprovability results of [38, 26]. We, however, obtain an unconditional unprovability result by first showing that the truth of the

statement implies its unprovability. By the soundness of $\mathsf{T}^0_{\mathsf{APC}_1}$, the statement is unprovable if it is false, hence the unprovability of the statement follows unconditionally.

For the sake of contradiction assume $\mathsf{T}^0_{\mathsf{APC}_1} \vdash \mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}(M, 2^{n^\delta}, m)$ with $m = 2^n/2 - 2^n/2^{n^\delta}$. By Theorem 4, there are finitely many functions $f_1, \ldots, f_l$ computable by circuits of size $2^{O(n)}$ such that for each $n > n_0$ and each co-nondeterministic circuit D of size 2^{n^δ}, either $f_1(D)$ outputs x^1, y^1, z^1 satisfying formula $Err(x^1, y^1, z^1)$ defined as

$$\forall y'^1, z'^1 \ (M(x^1, y^1) = 1 \wedge D(x^1, z^1) = 0) \vee (M(x^1, y'^1) = 0 \wedge D(x^1, z'^1) = 1)$$

or there are counterexamples y'^1, z'^1 witnessing $\neg Err(x^1, y^1, z^1)$. In the latter case, $f_2(D, y'^1_1, z'^1_1)$ generates x^2, y^2, z^2 potentially satisfying $Err(x^2, y^2, z^2)$ and the protocol continues in the same way, but after $t \leq l$ many rounds some f_t will succeed in finding the right x^t, y^t, z^t. Note that then $M(x^t) \neq D(x^t)$.

We will apply the witnessing functions $f_1, \ldots, f_l$ on co-nondeterministic circuits defining a Nisan-Wigderson generator based on (the negation of) M to show that they allow us to compute M by subexponential-size deterministic circuits. Specifically, for $c \geq 4$ and $n^c \leq m' \leq 2n^c$, let $A = \{a_{i,j}\}^{i=1,\ldots,2^n}_{j=1,\ldots,m'}$ be $2^n \times m'$ 0-1 matrix with $n^{c/2}$ ones per row and $J_i(A) := \{j \in \{1, \ldots, m'\}; a_{i,j} = 1\}$. Then define an NW-generator, $NW_{\neg M} : \{0,1\}^{m'} \mapsto \{0,1\}^{2^n}$ as

$$(NW_{\neg M}(w))_i = \neg M(w|J_i(A))$$

where $w|J_i(A)$ are w_j's such that $j \in J_i(A)$.

For any $c \geq 4$, Nisan and Wigderson [35] constructed $2^n \times m'$ 0-1 matrix A with $n^{c/2}$ ones per row and $n^c \leq m' \leq 2n^c$ which is also an $(n, n^{c/2})$-design meaning that for each $i \neq j$, $|J_i(A) \cap J_j(A)| \leq n$. Moreover, there are $n^{9c/2}$-size circuits which given $i \in \{0,1\}^n$ and $w \in \{0,1\}^{m'}$ output $w|J_i(A)$, cf. [8]. Therefore, if M is computable by 2^{n^ϵ}-size nondeterministic circuits, then for each $w \in \{0,1\}^{m'}$, $(NW_{\neg M}(w))_x$ is a function on n inputs x computable by co-nondeterministic circuits of size $2^{n^{2c\epsilon}}$, which is $< 2^{n^\delta}$ if $\epsilon < \delta/2c$.

For $u \in \{0,1\}^{n^{c/2}}$ and $v \in \{0,1\}^{m'-n^{c/2}}$ define $r_x(u,v) \in \{0,1\}^{m'}$ by putting bits of u into positions $J_x(A)$ and filling the remaining bits by v (in the natural order). Let $f^C_i(D, \ldots)$ be the first of the three strings generated by $f_i(D, \ldots)$, i.e. x^i. For all $w \in \{0,1\}^{m'}$ and co-nondeterministic circuits $D(x) = NW_{\neg M}(w)_x$ there is a trace $tr(w) = f^C_1(D), \ldots, f^C_t(D, \ldots), t \leq l$ of outputs of functions f_l such that $M(f^C_t(D, \ldots)) \neq NW_{\neg M}(w)_{f^C_t(D,\ldots)}$. That is, $tr(w)$ is a sequence of inputs of circuit D on which D is expected to fail to compute M. The trace is defined w.r.t. a fixed 'helpful' adversary Y providing counterexamples to the claims that $Err(x, y, z)$ holds. If there is no such counterexample, Y provides 0. The advice of Y depends only on x and $w|J_x(A)$.

Claim 3.1. *There is a trace $Tr = X_1, \ldots, X_t, t \leq l$ and $a \in \{0,1\}^{m'-n^{c/2}}$ such that for $s \geq 1/(n3^{t-1}2^{nt})$ fraction of all u's trace $tr(r_{X_t}(u,a))$ starts with Tr and for at least $(2/3 - 1/n)s$ fraction of all u's trace $tr(r_{X_t}(u,a))$ is exactly Tr, i.e. $tr(r_{X_t}(u,a)) = Tr$.*

Tr and a can be constructed inductively. There are at most 2^n strings X_j, hence there is X_1 such that for $\geq 1/2^n$ fraction of all w's trace $tr(w)$ begins with X_1. Assume we have a trace

$X_1, \ldots, X_i$ such that for $s'' \geq 1/(3^{i-1}(2^n)^i)$ fraction of all w's trace $tr(w)$ begins with $X_1, \ldots, X_i$. Either for at least $2s''/3$ of all w's trace $tr(w) = X_1, \ldots, X_i$ or for at least $s''/3$ of w's $tr(w)$ extends $X_1, \ldots, X_i$. In the latter case, there is X_{i+1} such that for $s' \geq 1/(3^i(2^n)^{i+1})$ fraction of all w's trace $tr(w)$ begins with $X_1, \ldots, X_{i+1}$. Since the trace is finite, there is $X_1, \ldots, X_t$, $t \leq l$ such that for $s''' \geq 1/(3^{t-1}2^{nt})$ of all w's trace $tr(w)$ begins with $X_1, \ldots, X_t$ and for $\geq 2s'''/3$ of w's $tr(w) = X_1, \ldots, X_t$. Therefore, by an averaging argument, there is $a \in \{0,1\}^{m'-n^{c/2}}$ such that for at least $s \geq s'''/n$ of all u's $tr(r_{X_t}(u,a))$ starts with $X_1, \ldots, X_t$ (possibly extending it) and $tr(r_{X_t}(u,a)) = X_1, \ldots, X_t$ for $\geq (2/3 - 1/n)s2^{n^{c/2}}$ of u's. This proves the claim.

Fix now Tr and a from the previous claim.

Let $r_{X_t}(\cdot, a)$ be the bits of a in the positions of $[m']\backslash J_{X_t}(A)$. Since A is $(n, n^{c/2})$-design, for any row $x \neq X_t$ at most n bits of $r_{X_t}(\cdot,a)|J_x(A)$ are not set by a. Let $Y_x, x \neq X_t$ be the set of all counterexamples y'_x, z'_x provided by Y on x and $r_{X_t}(u,a)|J_x(A)$ for all u. The size of each set Y_x is $2^{O(n)}$.

Now we define a circuit D' that approximates $\neg M$ using as advice $Tr, r_{X_t}(\cdot,a)$ and $t-1$ sets $Y_{X_1}, \ldots, Y_{X_{t-1}}$. For each $u \in \{0,1\}^{n^{c/2}}$ produce $r_{X_t}(u,a)$. Let V be the set of those inputs u for which $tr(r_{X_t}(u,a))$ either is Tr or starts as Tr and let U be the complement of V. Define d_0 to be the majority value of $\neg M$ on U, this will be another bit of a nonuniform advice. Then use f_1 to produce x^1. If X_1 from Tr is not x^1, output d_0. Otherwise, use the advice in Y_{X_1} to generate a counterexample against the claim that $M(x^1) \neq NW_{\neg M}(r_{X_t}(u,a))_{x^1}$. If the advice does not work, output d_0. Otherwise use the resulting counterexample in f_2 and continue in the same manner until f_t produces x^t. If X_t from Tr is not x^t, output d_0. Otherwise, output 0 iff $M(X_t) = 1$. Since X_t is fixed, $M(X_t)$ can be decided by a final single bit of a nonuniform advice.

D' is a circuit with $n^{c/2}$ inputs and size $2^{O(n)}$.

By the choice of Tr, for at least $(2/3-1/n)s$ fraction of all $u \in \{0,1\}^{n^{c/2}}$, D' will successfully predict $\neg M(u)$. Moreover, at most $(1/3+1/n)s$ of all traces $tr(r_{X_t}(u,a))$ extend Tr. Because $s \geq 1/(n3^{t-1}2^{nt})$ and d_0 is the correct value on at least half of $u \in U$, $P_u[D'(u) = \neg M(u)] \geq 1/2 + (1/6 - 1/n)/(n3^{t-1}2^{nt})$. This contradicts the assumption that every circuits of size 2^{m^δ} errs in computing M on $\geq 2^m/2 - 2^m/2^{m^\delta}$ inputs of length m, assuming c is sufficiently big. $\square$

We now weaken the average-case lower bound which is shown to be unprovable in Theorem 5 by employing hardness amplification of Healy, Vadhan and Viola [17]. The following lemma combines Corollary 10.1 in [17] with a simple padding argument that enables the hardness amplification to work almost everywhere.

Lemma 1 (Healy-Vadhan-Viola [17])**.** *If $f : \{0,1\}^n \mapsto \{0,1\}$ is a Boolean function computable by p-size nondeterministic circuits such that each circuit of size $s(n)$ fails to compute f on $\geq 1/poly(n)$ inputs for all large enough n, then there is a function $f' : \{0,1\}^m \mapsto \{0,1\}$ computable by p-size nondeterministic circuits such that each circuit of size $s(m^{1/4})^{\Omega(1)}$ fails to compute f' on $\geq 1/2 - 1/s(m^{1/4})^{\Omega(1)}$ inputs for all large enough m.*

Theorem 6. *Let $n_0, \delta \in (0,1)$ be rational constants and M be a p-time nondeterministic Turing machine. Then,*

$$\mathsf{T}^0_{\mathsf{APC}_1} \not\vdash \mathsf{LB}^{\mathsf{coN}}_{\mathsf{tt}}(M, 2^{n^\delta}, \tfrac{2^n}{n}).$$

Proof. We proceed as in the proof of Theorem 5 but instead of basing the Nisan-Wigderson generator on $\neg M$ we use amplified version of $\neg M$ from Lemma 1. More precisely, given M, Lemma 1 produces a function $Amp(M)$ computable by p-size nondeterministic circuits. Using $\neg Amp(M)$ instead of $\neg M$ in the proof of Theorem 5, we then obtain deterministic circuits of size $2^{O(m)}$ computing $Amp(M)$ on inifinitely many $m^{c/2}$-bit inputs with probability $\geq 1/2+1/2^{O(m)}$. If c is sufficiently big, Lemma 1 implies that there are $2^{O(n^4)}$-size circuits computing M on infinitely many inputs of length $n^{c/2}$ with probability $\geq 1 - 1/poly(n)$, which contradicts the assumption. □

3.4 Unprovability of the existence of super-bits

In [46] Rudich conjectured the existence of so called super-bits and showed that the existence of super-bits refutes the existence of NP-natural proofs against $\mathsf{P/poly}$.

Definition 1 (Rudich's super-bit [46]). *A sequence of functions $g_n : \{0,1\}^n \to \{0,1\}^{n+1}$ computable by polynomial-size circuits is a super-bit if there exists $\epsilon > 0$ such that for every nondeterministic circuit C with $n+1$ inputs and size $S \leq 2^{n^\epsilon}$,*

$$\Pr_{y\in\{0,1\}^{n+1}}[C(y) = 1] - \Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] < 1/S.$$

The statement that a map $g : \{0,1\}^n \mapsto \{0,1\}^{n+1}$ is a super-bit can be also formalized by a $\forall\Sigma^b_2$-formula assuming g is computable not just in $\mathsf{P/poly}$ but in p-time. Specifically, we say that a theory T in the language PV proves that a sequence of functions $g_n : \{0,1\}^n \mapsto \{0,1\}^{n+1}$ computable in p-time is a super-bit if there exists $\epsilon > 0$ such that T proves that for each $n \in LogLog$, $N = 2^n$ and each nondeterministic circuit C with $n+1$ inputs and size $S \leq 2^{n^\epsilon}$,

$$\Pr_{y\in\{0,1\}^{n+1}}[C(y) = 1] - \Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] < 1/S, \tag{3.3}$$

where the inequality (3.3) is expressed by the following Σ^b_2-formula: there exist rational numbers $p_1, p_2 \in [0,1]$ and S-bit strings $z_1^{p_1,1}, \ldots, z_{2N}^{p_1,1}$, $z_1^{p_2,1}, \ldots, z_N^{p_2,1}$ such that for all S-bit strings $z_1^{p_1,0}, \ldots, z_{2N}^{p_1,0}$, $z_1^{p_2,0}, \ldots, z_N^{p_2,0}$ the following holds

- $\Pr_{y\in\{0,1\}^{n+1}}[C(y) = 1] = p_1$ is witnessed by strings $z_y^{p_1,1}, z_y^{p_1,0}$, i.e. $C'(y, z_y^{p_1.1}) = 1$ for $2p_1N$ strings $y \in \{0,1\}^{n+1}$ and $C'(y, z_y^{p_1,0}) = 0$ for $2(1-p_1)N$ strings $y \in \{0,1\}^{n+1}$ such that $C'(y, z_y^{p_1,1}) = 0$, where $C'(a,b)$ denotes $C(a)$ with nondeterministic bits b.
- $\Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] = p_2$ is witnessed by strings $z_x^{p_2,1}, z_x^{p_2,0}$, i.e. $C'(g_n(x), z_x^{p_2,1}) = 1$ for p_2N strings $x \in \{0,1\}^n$ and $C'(g_n(x), z_x^{p_2,0}) = 0$ for $(1-p_2)N$ strings $x \in \{0,1\}^n$ such that $C'(g_n(x), z_x^{p_2,1}) = 0$,
- $p_1 - p_2 < 1/S$.

The rational numbers p_1, p_2 have the form a/b, where $a, b \in [2N]$.

On the choice of formalization. An alternative formalization of the statement that g_n is a super-bit would say that for each p_1, p_2, if there are strings $z_i^{p_1,1}, z_i^{p_2,1}$ such that for all strings $z_i^{p_1,0}, z_i^{p_2,0}$ the strings witness $\Pr_{y\in\{0,1\}^{n+1}}[C(y) = 1] = p_1$ and $\Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] = p_2$, then $p_1 - p_2 < 1/S$. It is, however, unclear if a theory such as $\mathsf{T}^1_{\mathsf{APC}_1}$ can prove that these formalizations are equivalent. This is because the concepts involved in this claim are defined by nondeterministic circuits and $\mathsf{T}^1_{\mathsf{APC}_1}$ is a priori not well-equiped for reasoning about them. A concrete obstacle towards establishing the equivalence in $\mathsf{T}^1_{\mathsf{APC}_1}$ is the so called sharply bounded collection scheme for PV-formulas, which has the form

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

where B is an open PV-formula and $[w]_i$ is the i-th element of the sequence of strings coded by w. While S^1_2 proves the scheme, Cook and Thapen [14] showed that PV_1 does not prove it unless factoring is easy. We do not use the alternative formalization in the rest of the paper.

We will show that there is no sequence of p-time functions g_n such that $\mathsf{T}^0_{\mathsf{APC}_1}$ proves that g_n is a super-bit satisfying $\Pr_{y\in\{0,1\}^{n+1}}[y \in Rng(g_n)] = 1/2$, i.e. there is no g_n such that $\mathsf{T}^0_{\mathsf{APC}_1}$ proves that g_n is an injective super-bit. The proof will proceed by a reduction to Theorem 5. Let $R := Rng(g_n)$ and $\overline{R} = \{0,1\}^{n+1}\backslash R$. The crucial observation is that if g_n is such a super-bit then $\overline{R}$, which is a coNP language, cannot be computed by any nondeterministic circuit C of size $S \leq 2^{n^\epsilon}$ with probability $\geq 1/2 + 1/S$. To see that, note that

$$\begin{aligned}
&\Pr_{y\in\{0,1\}^{n+1}}[C(y) = 1] - \Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] = \\
&\quad = \Pr_y[C(y) = 1 \wedge y \in R] + \Pr_y[C(y) = 1 \wedge y \notin R] - 2\Pr_y[C(y) = 1 \wedge y \in R] \\
&\quad = 1/2 - \Pr_y[C(y) = 0 \wedge y \notin R] - \Pr_y[C(y) = 1 \wedge y \in R] \\
&\quad = 1/2 - \Pr[C(y) = R(y)] \geq 1/S
\end{aligned}$$

where the first two equalities use $|R| = 1/2$ and the last inequality assumes $\Pr[C(y) = \overline{R}(y)] \geq 1/2 + 1/S$.

Theorem 7 (Breaking super-bits in $\mathsf{T}^0_{\mathsf{APC}_1}$)**.** *There is no p-time $g_n : \{0,1\}^n \mapsto \{0,1\}^{n+1}$ and $\epsilon > 0$ such that $\mathsf{T}^0_{\mathsf{APC}_1}$ proves that g_n is a super-bit satisfying $\Pr_{y\in\{0,1\}^{n+1}}[y \in Rng(g_n)] = 1/2$.*

Proof. For the sake of contradiction assume that this is not the case. We claim that then for some $\epsilon > 0$, $\mathsf{T}^0_{\mathsf{APC}_1}$ proves that for each sufficiently big $n \in LogLog$ for each nondeterministic circuit C of size $S \leq 2^{n^\epsilon}$ there exists y such that $C(y) \neq \overline{R}(y)$. Otherwise, we could conclude in $\mathsf{T}^0_{\mathsf{APC}_1}$ that for some S-size C, $\Pr_{x\in\{0,1\}^n}[C(g_n(x)) = 1] = 0$ and since $\Pr_{y\in\{0,1\}^{n+1}}[y \in R] = 1/2$ also $\Pr_{y\in\{0,1\}^{n+1}}[C(y) = 1] = 1/2$, which would contradict the assumption that g_n is a super-bit. However, since the claim involves reasoning with nondeterministic circuits we need to be more careful when showing that it is indeed formalizable in $\mathsf{T}^0_{\mathsf{APC}_1}$.

In more detail, we perform the following argument in $\mathsf{T}^0_{\mathsf{APC}_1}$. Assume that the conclusion does not hold for some C, i.e. $\forall y\ [\exists z\ C'(y,z)=1 \leftrightarrow y \notin R]$, which we express as

$$\forall y \forall z_1,\ldots,z_{2N} \exists z'_1,\ldots,z'_{2N}\ [(C'(y,z_y)=0 \vee y \notin R) \wedge (C'(y,z'_y)=1 \vee y \in R)]. \tag{3.4}$$

Then, for each $p_2 \geq 1/N$ for all strings $z_1^{p_2,1},\ldots,z_N^{p_2,1}$ there are strings $z_1^{p_2,0},\ldots,z_N^{p_2,0}$ satisfying $C'(g_n(x),z_x^{p_2,1}) \neq 1$ for all x, and hence falsifying $\Pr_{x\in\{0,1\}^n}[C(g_n(x))=1]=p_2$,. Next, note that the probability $\Pr_{y\in\{0,1\}^{n+1}}[y \in R]=1/2$ can be counted exactly in $\mathsf{T}^0_{\mathsf{APC}_1}$ because $n \in \mathit{LogLog}$ and the predicate $y \in R$ is in $2^{O(n)}$-time. This can be done, for example, by listing all y's satisfying $y \in R$. We now claim that, for each $p_1 \neq 1/2$ for all strings $z_1^{p_1,1},\ldots,z_{2N}^{p_1,1}$ there are strings $z_1^{p_1,0},\ldots,z_{2N}^{p_1,0}$ refuting $\Pr_{y\in\{0,1\}^{n+1}}[C(y)=1]=p_1$, i.e. either not satisfying $C'(y,z_y^{p_1,1})=1$ for $2p_1N$ y's or not satisfying $C'(y,z_y^{p_1,0})=0$ for the remaining y's. If the claim holds, this contradicts the assumption that g_n is a super-bit and proves our original claim.

To see that the claim is true, suppose it is false. There are two cases. In the first case, $p_1 > 1/2$ and there are strings $z_1^{p_1,1},\ldots,z_{2N}^{p_1,1}$ such that for all strings $z_1^{p_1,0},\ldots,z_{2N}^{p_1,0}$, $C'(y,z_y^{p_1,1})=1$ for $>1/2$ of y's, while by (3.4) for all strings $z_1^{p_1,1},\ldots,z_{2N}^{p_1,1}$ there are strings $z_1^{p_1,0},\ldots,z_{2N}^{p_1,0}$ such that $C'(y,z_y^{p_1,1})=0$ for $\geq 1/2$ of y's. This is a contradiction. In the second case, $p_1 < 1/2$ and there are strings $z_1^{p_1,1},\ldots,z_{2N}^{p_1,1}$ such that for all strings $z_1^{p_1,0},\ldots,z_{2N}^{p_1,0}$, $C'(y,z_y^{p_1,0})=0$ for $>1/2$ of y's, while by (3.4) for all strings $z_1^{p_1,1},\ldots,z_{2N}^{p_1,1}$ there are strings $z_1^{p_1,0},\ldots,z_{2N}^{p_1,0}$ such that $C'(y,z_y^{p_1,0})=1$ for $\geq 1/2$ of y's. This is again a contradiction, which proves our claim.

Given the $\mathsf{T}^0_{\mathsf{APC}_1}$-provability of S-size nondeterministic lower bound for $\overline{R}$ expressed with the negation of formula (3.4), we can follow the proof of Theorem 5 adapted to the symmetric case where M is co-nondeterministic and D nondeterministic and derive that for infinitely many n, we can approximate $\overline{R}$. More precisely, choosing a sufficiently big c, there is a circuit of size $2^{O(n^{\epsilon/2})}$ computing $\overline{R}$ with probability $\geq 1/2+1/2^{O(n^{\epsilon/2})}$, which by the argument presented before Theorem 7 contradicts the assumption that g_n is a super-bit. □

Note that in the proof of Theorem 7 it is crucial that we avoided using the sharply bounded collection scheme. In particular, it is not clear how to adapt the proof of Theorem 7 to the case of the alternative formalization of super-bits discussed earlier.

4 KPT-interpolation property

Abstracting on the notion of KPT witnessing which played a crucial role in the proof of Theorem 5 we define a generalized notion of feasible interpolation property, which we call KPT-interpolation property. We then point out that the lower bound method of Krajíček [26] yields conditional lower bounds for systems admitting KPT-interpolation. This strengthens a result from [37] showing that Razborov's conjecture [44] about a conditional hardness of Nisan-Wigderson generators for strong propositional proof systems holds for proof systems with feasible interpolation.

Definition 2. *We say that a propositional proof system P admits KPT-interpolation property, if there exists a constant k and p-time functions $f_1,\ldots,f_k$ such that whenever π is a P-proof*

of

$$A_1(x, y_1) \vee \cdots \vee A_m(x, y_m)$$

where y_i are disjoint tuples of variables and x are the only common variables of $A_1, \ldots, A_m$, for each x, either $f_1(x, \pi)$ outputs j_1 such that $A_{j_1}(x, y_{j_1})$ is a tautology, or if it is not then having a counterexample a_1 witnessing that $\neg A_{j_1}(x, a_1)$, $f_2(x, \pi, a_1)$ generates j_2 such that $A_{j_2}(x, y_{j_2})$ is a tautology, or there is a counterexample a_2 and the protocol proceeds with each f_ℓ having access to all counterexamples obtained so far, but one of the functions $f_1, \ldots, f_k$ definitely succeeds in finding a tautology among $A_1, \ldots, A_m$

The notion of feasible interpolation, one of the most powerful methods for proving proof complexity lower bounds, can be seen as a special case of KPT interpolation for $m = 2$ and $k = 1$. In fact, the ordinary feasible interpolation often yields p-time functions not only deciding which of the two given disjuncts is a tautology but even constructing its proof. In such case repeating the construction $\log m$ times, we can treat m-size disjunctions as well, without a need for counterexamples. The results deriving the impossibility of feasible interpolation for strong proof systems such as EF, cf. [31], do not go through directly with the KPT interpolation simply because witnessing a disjunction of length 2 is always easy for the interactive process from the KPT interpolation. We thus originally hoped that KPT interpolation could be used to derive lower bounds for some proof systems for which feasible interpolation fails. This would present an alternative to a similar weakening of feasible interpolation, the so called feasible disjunction property [28, Problem 17.9.1]. However, in light of the subsequent work of Krajíček [29] this possibility seems less likely. Nevertheless, it is still possible that the KPT interpolation could hold in a specific case needed to derive the conditional lower bound from Corollary 1 below, or that it can be further modified into a property which is not ruled out by [29].

Our introduction of the KPT-interpolation was motivated by the attempts to turn the unprovability result of Krajíček [26] into a lower bound for strong proof systems. As discussed in [26, 29], a direct attempt to get the lower bound out of the unprovability fails on the fact that theories such as PV_1 are unlikely to prove sharply bounded collection schemes discussed in Section 3.4. The KPT-interpolation is designed to avoid the obstacle.

We now want to prove a lower bound on lengths of proofs in proof systems with KPT interpolation by simulating the proof of Theorem 5 in propositional logic. A complication is that Theorem 5 speaks about unprovability of a Σ^b_2-statement and it is not clear how to translate it into propositional formulas.

In order to avoid this problem we will follow Krajíček [26] and focus on $\mathsf{NP} \cap \mathsf{coNP}$ lower bounds of a specific form given by a NW-generator. More precisely, we will consider propositional formulas $f \notin Rng(NW_h)$ expressing that a Boolean function f with n inputs represented by its truth table is outside of the range of NW generator based on a function $h \in \mathsf{NP} \cap \mathsf{coNP}$. The NW generator will have the same input/output ratio as the one in the proof of Theorem 5, i.e. n^c columns and 2^n rows, for $c \geq 4$, and it will be based on the same $(n, n^{c/2})$-combinatorial design. Since the function h is in $\mathsf{NP} \cap \mathsf{coNP}$, there are nondeterministic circuits $H_1(z, w), H_0(z, w)$ of polynomial-size with nondeterministic inputs w such that

$$h(z) = \epsilon \Leftrightarrow \exists w \; H_\epsilon(z, w).$$

We can thus express $f \notin Rng(NW_h)$ as a disjunction

$$\bigvee_{y \in [2^n]} f_y \neq (NW_h(x))_y$$

where $f_y \neq (NW_h(x))_y$ is $\neg H_1(x|J_y, w)$ if $f_y = 1$ and $\neg H_0(x|J_y, w)$ if $f_y = 0$. Here, $x|J_y$ is a restriction of x to the inputs specified by the combinatorial design. The total size of the formula $f \notin Rng(NW_h)$ is $2^{O(n)}$.

Corollary 1. *Let $h \in \mathsf{NP} \cap \mathsf{coNP}$ be hard to approximate by circuits of size $2^{O(n^{2/c})}$ with advantage $1/2^{O(n^{2/c})}$. Then, no propositional proof system P admitting KPT-interpolation has p-size proofs of propositional formulas $f \notin Rng(NW_h)$. This holds regardless of the choice of f.*

Proof. Corollary 1 follows directly from the main result of Krajíček [26] formulated in the setting of propositional proof systems with KPT interpolation. Assume the proof system P proves $f \notin Rng(NW_h)$ efficiently. We proceed analogously as in the proof of Theorem 5 while replacing the application of KPT theorem by the KPT-interpolation. Further, instead of using $\neg M$ as the base function of the NW-generator, we use h. The rest of the proof still works and gives us a circuit of subexponential size approximating h with subexponential advantage. □

Finally, we show that if constant-depth Frege admits KPT interpolation, a conditional lower bound on (unbounded) Frege follows. This is a consequence of a proof complexity magnification from [34].

Proof complexity magnification exploits the possibility of expressing circuit lower bounds succinctly. As proved by Lipton and Young [33], whenever $f : \{0,1\}^n \mapsto \{0,1\}$ is hard for circuits of size $poly(s(n))$, there is a set S_n of $poly(s(n))$ n-bit strings such that each $s(n)$-size circuit fails to compute f on some input from the 'anti-checking' set S_n. The $s(n)$-size circuit lower bound for f can be thus expressed by a $poly(s(n))$-size formula $\bigvee_{y \in S_n} f(y) \neq C(y)$ where the formula $f(y) \neq C(y)$ says that a circuit C represented by $poly(s)$ variables does not output $f(y)$ on input y.

This implication works for nondeterministic circuits as well, cf. also [9], i.e. whenever $f : \{0,1\}^n \mapsto \{0,1\}$ is hard for nondeterministic circuits of size $poly(s(n))$, there is a set S_n of $poly(s(n))$ n-bit strings such that each $s(n)$-size nondeterministic circuit fails to compute f on some input from the set S_n. In particular, this means that for every $f \notin \mathsf{NP/poly}$ we can express $f \notin Rng(NW_h)$ more succinctly by formulas $f \notin_{S_n} Rng(NW_h)$:

$$\bigvee_{y \in S_n} f_y \neq (NW_h(x))_y,$$

for an appropriate set S_n of size $poly(n)$.

Theorem 8. *Let $h \in \mathsf{NP} \cap \mathsf{coNP}$ be hard to approximate by circuits of size $2^{O(n^{2/c})}$ with advantage $1/2^{O(n^{2/c})}$ and $f \notin \mathsf{NP/poly}$. If constant-depth Frege admits KPT-interpolation, then there are no p-size Frege proofs of $f \notin_{S_n} Rng(NW_h)$. This holds for each set S_n of size $poly(n)$.*

Proof sketch. Assume Frege has p-size proofs of $f \notin_{S_n} Rng(NW_h)$ for some S_n. A generic collapse of Frege to constant-depth Frege by Filmus, Pitassi and Santhanam [16] implies that there is a constant K such that for any d, Frege proves $f \notin_{S_n} Rng(NW_h)$ by proofs of size $2^{O(dn^{K/d})}$ and depth $d+2$. Tautologies $f \notin Rng(NW_h)$ can be then derived by weakening which increases the size of the proof to $2^{O(n)}$. This contradicts Corollary 1. □

5 Concluding remarks and open problems

Worst-case lower bounds, APC_1 and S^1_2. Is it possible to improve our unprovability from Theorem 5 into unprovability of a worst-case lower bound? A step in this direction could be to show the unprovability of the nonexistence of natural proofs, which is a zero-error average-case lower bound. More precisely, can we show $\mathsf{T}^0_{\mathsf{APC}_1}$-unprovability of the nonexistence of NP-natural proofs against $\mathsf{P/poly}$? Another possible improvement of Theorem 5 would be to make it work for a stronger theory. In particular, can we prove it for the theory APC_1 or S^1_2?

Witnessing hard tautologies. The methods for proving our unprovability statements show that under certain hardness assumptions it is impossible to efficiently witness errors of computations of Nisan-Wigderson generators. For example, Krajíček [26, 27] used this to show that under the assumption of the existence of one-way permutations secure against circuits of exponential-size, there is no subexponential-time algorithm which given a nondeterministic circuit C generates a C-proof of an invalid formula or a tautology which is hard for C. It would be very interesting to achieve the non-existence of such witnessing under the promise that the given circuit C is sound. This question has been investigated already in [27].

Question 1. *Is there a p-time algorithm which for any nondeterministic circuit C of p-size accepting only tautologies outputs a tautology which has no C-proof?*

Acknowledgements

We would like to thank Jan Krajíček for helpful discussions, for comments on the draft of the paper and for pointing out Theorem 4 to us. We would also like to thank Erfan Khaniki and Emil Jeřábek for helpful comments and Igor Carboni Oliveira for pointing out a correction in the draft of the paper. Ján Pich was supported by Grant 19-05497S of GAČR. Rahul Santhanam was supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2014)/ERC Grant Agreement No. 615075. This project has received funding from the Eurpoean Union's Horizon 2020 research and innovation programme under the Marie Skłodovska-Curie grant agreement No 890220.

References

[1] Alekhnovich M., Ben-Sasson E., Razborov A.A., Wigderson A.; *Pseudorandom generators in propositional proof complexity*, SIAM Journal on Computing, 34(1):67-88, 2004.

[2] Aaronson, S., Wigderson, A.; *Algebrization: A New Barrier in Complexity Theory*, Transactions in Computation Theory, 2:1-54, 2009.

[3] Baker T., Gill J., Solovay R.; *Relativizations of the* $\mathsf{P} = \mathsf{NP}$ *Question*, SIAM Journal on Computing, 4(4), pp. 431-442, 1975.

[4] Buss S.; *Bounded Arithmetic*, Bibliopolis, 1986.

[5] Buss S., Kołodziejczyk L., Thapen N.; *Fragments of Approximate Counting*, Journal of Symbolic Logic, 49:496-525, 2014.

[6] Bydžovský J., Müller M.; *Polynomial time ultrapowers and the consistency of circuit lower bounds*, Arch. Math. Log., 59(1):127-147, 2020.

[7] Bydžovský J., Krajíček J., Oliveira I.C.; *Consistency of circuit lower bounds with bounded theories*, Logical Methods in Computer Science, 16(2:12), 2020.

[8] Carmosino M., Impagliazzo R., Kabanets V., Kolokolova A.; *Learning algorithms from natural proofs*, Conference on Computational Complexity, 10:1-24, 2016.

[9] Chen L., Hirahara S., Oliveira I.C., Pich J., Rajgopal N., Santhanam R.;*Beyond natural proofs: hardness magnification and locality*, Innovations in Theoretical Computer Science, 70:1-48, 2020.

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

[11] Cook S.A.; *Feasibly constructive proofs and the propositional calculus*, Symposium on the Theory of Computing, pp. 83-97, 1975.

[12] Cook S., Krajíček J.; *Consequences of the Provability of* $\mathsf{NP} \subseteq \mathsf{P/poly}$, Journal of Symbolic Logic, 72(4):1353-1371, 2007.

[13] Cook S.A., Pitassi T.; *A feasibly constructive lower bound for Resolution proofs*, Information Processing Letters, 34(2):81-85, 1990.

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

[15] Dai Tri Man Le; *Bounded arithmetic and formalizing probabilistic proofs*, Ph.D. thesis, University of Toronto, 2014.

[16] Filmus Y., Pitassi T., Santhanam R.; *Exponential lower bounds for* AC^0*-Frege imply superpolynomial Frege lower bounds*, International Colloquium on Automata, Languages and Programming, 6755:618-629, 2011.

[17] Healy A., Vadhan S., Viola E.; *Using nondeterminism to amplify hardness*, Symposium on the Theory of Computing, pp. 192-201, 2004.

[18] Huang H.; *Induced Subgraphs of Hypercubes and a Proof of the Sensitivity Conjecture*, Arxiv preprint, 2019.

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

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

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

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

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

[24] Krajíček J.; *Dual weak pigeonhole principle, pseudor-surjective functions, and provability of circuit lower bounds*, Journal of Symbolic Logic, 69(1):265-286, 2004.

[25] Krajíček J.; *Implicit proofs*, Journal of Symbolic Logic, 69(2):387-397, 2004.

[26] Krajíček J.; *On the proof complexity of the Nisan-Wigderson generator based on a hard* $\mathsf{NP} \cap \mathsf{coNP}$ *function*, Journal of Mathematical Logic, 11(1):11-27, 2011.

[27] Krajíček J.; *On the computational complexity of finding hard tautologies*, Bulleting of the London Mathematical Society, 46(1):111-125, 2014.

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

[29] Krajíček J.; *A limitation on the KPT interpolation*, Logical Methods in Computer Science, 16(3:9), 2020.

[30] Krajíček J., Oliveira I.C.; *Unprovability of circuit upper bounds in Cook's theory* PV_1, Logical Methods in Computer Science, 13(1), 2017.

[31] Krajíček J., Pudlák P.; *Some consequences of cryptographical conjectures for* S^1_2 *and* EF, Information and Computation, 140(1):82-94, 1998.

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

[33] Lipton R.J., Young N.E.; *Simple strategies for large zero-sum games with applications to complexity theory*, Symposium on Theory of Computing, pp. 734-740, 1994.

[34] Müller M., Pich J.; *Feasibly constructive proofs of succinct weak circuit lower bounds*, Annals of Pure and Applied Logic, 171(2), 2020.

[35] Nisan N., Wigderson A.; *Hardness vs. randomness*, J. Comput. Systems Sci., 49:149-167, 1994.

[36] Paris J., Wilkie A.; *Counting problems in bounded arithmetic*, Methods in Mathematical Logic, 1130:317-340, 1985.

[37] Pich J.; *Nisan-Wigderson generators in proof systems with forms of interpolation*, Mathematical Logic Quarterly, 57(4), 2011.

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

[39] Pich J.; *Logical Strength of Complexity Theory and a formalization of the PCP theorem in bounded arithmetic*, Logical Methods in Computer Science, 11(2), 2015.

[40] Pich J., Santhanam R.; *Why are proof complexity lower bounds hard?* Symposium on Foundations of Computer Science, pp. 1305-1324, 2019.

[41] Pudlák P.; *Lower Bounds for Resolution and Cutting Plane proofs and Monotone computations*; J. of Symb. Logic, 62(3):981-998, 1997.

[42] Razborov A.A.; *Bounded Arithmetic and Lower Bounds in Boolean Complexity*, Feasible Mathematics II, pp. 344-386, 1995.

[43] Razborov A.A.; *Unprovability of lower bounds on circuit size in certain fragments of bounded arithmetic*, Izvestiya of the Russian Academy of Science, 59:201-224, 1995.

[44] Razborov A.A.; *Pseudorandom generators hard for k-DNF resolution and polynomial calculus*, Annals of Mathematics, 181(2):415-472, 2015.

[45] Razborov A.A., Rudich S.; *Natural proofs*, Journal of Computer and System Sciences, 55(1):24-35, 1997.

[46] Rudich S.; *Super-bits, Demi-bits, and NP/qpoly-natural Proofs*, Journal of Computer and System Sciences, 55:204-213, 1997.

[47] Thapen N.; *A model-theoretic characterization of the weak pigeonhole principle*, Annals of Pure and Applied Logic, 118(1-2):175-195, 2002.

[48] Thapen N.; *The weak pigeonhole principle in models of bounded arithmetic*, Ph.D. thesis, Oxford University, 2002.