Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 144 (2017)

Feasibly constructive proofs of succinct weak circuit lower bounds[*]

Moritz Müller   Ján Pich
Kurt Gödel Research Center for Mathematical Logic
University of Vienna
Währinger Straße 25, 1090 Wien, Austria
{moritz.mueller,jan.pich}@univie.ac.at[†]

Abstract

We ask for feasibly constructive proofs of known circuit lower bounds for explicit functions on bit strings of length n. In 1995 Razborov showed that many can be proved in PV_1, a bounded arithmetic formalizing polynomial time reasoning. He formalized circuit lower bound statements for small n of doubly logarithmic order. It is open whether PV_1 proves known lower bounds in *succinct* formalizations for n of logarithmic order. We give such proofs in APC_1, an extension of PV_1 formalizing probabilistic polynomial time reasoning: for parity and AC^0, for mod q and $\mathsf{AC}^0[p]$ (only for n slightly smaller than logarithmic), and for k-clique and monotone circuits. We also formalize Razborov and Rudich's natural proof barrier.

We ask for short propositional proofs of circuit lower bounds expressed succinctly by propositional formulas of size $n^{O(1)}$ or at least much smaller than the $2^{O(n)}$ size of the common "truth table" formula. We discuss two such expressions: one via feasible functions witnessing errors of circuits, and one via the antichekers of Lipton and Young 1994. Our APC_1 formalizations yield conditional upper bounds for the succinct formulas obtained by witnessing: we get short Extended Frege proofs *from* general circuit lower bounds expressed by the common "truth-table" formulas. We also show how to construct in quasipolynomial time propositional proofs of quasipolynomial size tautologies expressing $\mathsf{AC}^0[p]$ quasipolynomial size lower bounds; these proofs are in Jeřábek's system WF.

Keywords: proof complexity, bounded arithmetic, circuit lower bounds, approximate counting, natural proofs

2000 MSC: 03F20, 03D15, 68Q17

[*] A preliminary preprint of this paper appeared at the Electronic Colloquium on Computational Complexity (ECCC) TR17-144. A different, unpublished version of this paper is [46].

[†] E-mail address of the corresponding author: jan.pich@univie.ac.at.

ISSN 1433-8092

Contents

1 Introduction

It comes as no surprise when a complexity theorist, being concerned with the algorithmic hardness of computational tasks, starts wondering whether the notorious conjectures in the field are in some sense 'hard' to prove. Can one show first that existing proofs of partial results are 'simple' in some sense and second that such 'simple' reasoning is insufficient to settle the conjecture under consideration?

It is unclear whether there exists a good general notion of simplicity of proofs, already Hilbert asked for it in his 24th problem [66]. From a complexity theoretic perspective, however, one would naturally like to grade the complexity of proofs by the computational complexity of the concepts and constructions appearing in them. This is the viewpoint of "Bounded Reverse Mathematics" taken in the monograph [23, p.xiv] on proof complexity. In particular, the bounded arithmetic theory[1] PV_1, going back to Cook [20], can be viewed as formalizing proofs restricted to polynomial time computable concepts and constructions. In Cook's own words, "if one believes that all feasibly constructive arguments can be formalized in PV_1, then it is worthwhile seeing which parts of mathematics can be so formalized" [20, p.96]. As it turns out, a large part of contemporary complexity theory can be carried out in PV_1 or slight extensions of it (see the table in Section 5).

An example of particular interest is the apparently difficult task of proving circuit lower bounds for explicit functions. We consider three seminal results in the area:

(a) The Switching Lemma and a size lower bound for bounded depth circuits computing the parity function [1, 25, 27].

(b) Razborov and Smolensky's method of approximations by low degree polynomials and a size lower bound for bounded depth circuits containing modulo p counting gates computing the modulo q counting function [57, 63].

(c) Razborov's method of approximations and a size lower bound for monotone circuits deciding the clique problem [56].

We refer to [5] or [3] for surveys. We give proofs of (a)-(c) that are in a certain sense feasibly constructive. This Introduction gives an informal description of and motivation for our upper bounds and, moreover, aims to compactly survey the area, including independence and lower bounds.

1.1 Circuit lower bounds in PV_1

We continue Razborov's search for the "right fragment [of arithmetic] capturing the kind of techniques existing in Boolean complexity at present" [59, p.344]. He argued[2] that PV_1

[1] All relevant technical notions in this Introduction are explained later; e.g. PV_1 in Section 2.1.

[2] More precisely, Razborov argued for the two-sorted theory V^1_1 instead of PV_1. If one translates to the one-sorted setting using the $RSUV$-isomorphism (see e.g. [34, Theorem 5.5.13]), then V^1_1 becomes S^1_2,

"is exactly the required theory. By this I mean in particular that it proves all lower bounds mentioned above *and, moreover, these formal proofs are obtained in a very natural and straightforward way*[3]" [59, p.376]. Indeed, proofs of (a)-(c) formalize in PV_1 and partly even below: (a) in a theory corresponding to NC via a now famous new proof of Håstad's Switching Lemma [27], and (c) in a theory corresponding to circuits of a certain sublinear depth. We refer to [59] for precise statements.

We want to talk about circuit lower bounds for computational problems like the satisfiability problem SAT, and therefore blur the distinction between an explicit function $\mathsf{Q} : \{0,1\}^* \to \{0,1\}$ and the computational problem $\{x \mid \mathsf{Q}(x) = 1\}$. It is not straightforward to formalize a size s circuit lower bound for Q

$$\textit{For every circuit } C \textit{ of size } s \textit{ there exists } y \in \{0,1\}^n \textit{ such that } C(y) \neq \mathsf{Q}(y) \tag{1}$$

in bounded arithmetic which lacks exponentiation. A typical model of a bounded arithmetic theory has a proper initial segment of *small* numbers, numbers n whose exponential 2^n exists in the model; these are the numbers n that equal the *length* $|N|$ of (the binary representation of) another number N in the model. Razborov's formalization of (1) assumes 2^n is small, i.e. 2^{2^n} exists. This allows to encode C by (the binary expansion of) a number even for s exponential in n, and, in fact, the whole `truth table` of Q on $\{0,1\}^n$ is encoded by a number. Denote such a formalization of (1) by $\mathsf{LB}_{\mathsf{tt}}[\mathsf{Q}]$.

In such a formalization (numbers coding) inputs are small while circuits are not necessarily small. In Razborov's words, this captures "the common practice in the area which tends to treat Boolean inputs and functions separately, as two different kinds of objects" [59, p.375]. In Krajíček's words, it "differs from the one usually accepted in bounded arithmetic [. . .] in which all combinatorial objects (inputs, circuits,...) are coded at the same level [. . .] while (Boolean) functions are identified with definable classes" [35, Section 8]. An according *succinct* formalization assumes only that n is small. Consequently, it cannot consider bigger than polynomial size bounds $s \leqslant n^k$ for some constant $k \in \mathbb{N}$. Denote such a formalization by $\mathsf{LB}[\mathsf{Q}]$. More precisely, we have a formula $\mathsf{LB}[\mathsf{C},\mathsf{Q}](C,s,n,N)$ expressing a size s lower bound for circuits C from the class C; it uses an auxiliary variable N to witness that n is small.

The formula $\mathsf{LB}_{\mathsf{tt}}[\mathsf{Q}]$ assumes that 2^n is small, which intuitively means that the whole truth-table of Q on $\{0,1\}^n$ is considered a feasible object. The succinct LB-formalization assumes only that n is small. Intuitively, this means that only the size $\leqslant n^k$ of the circuit is considered feasible. For size bound $s = n^k$, the theory PV_1 is in some sense exponentially stronger w.r.t. $\mathsf{LB}_{\mathsf{tt}}[\mathsf{Q}]$ than it is w.r.t. $\mathsf{LB}[\mathsf{Q}]$. We now ask again for the right fragment of arithmetic to capture circuit lower bounds, this time in the succinct formalization. This is the topic of the present paper.

which is conservative over PV_1 with respect to $\mathsf{LB}_{\mathsf{tt}}[\mathsf{Q}]$ (see below).

[3] Emphasis added by the authors. Additionally to our (a)-(c), Razborov refers to lower bounds for monotone formulas.

1.2 Succinct circuit lower bounds in APC_1

As a candidate we put forward Jeřábek's theory APC_1 of approximate counting [30] which is a slight extension of PV_1 by the (*dual* or) *surjective* weak pigeonhole principle for polynomial time functions. While PV_1 formalizes polynomial time reasoning, APC_1 formalizes probabilistic polynomial time reasoning. We remark that a proof of $\mathsf{LB}[\mathsf{C},\mathsf{Q}]$ in APC_1 gives a probabilistic polynomial time algorithm that witnesses errors of small C-circuits trying to decide Q (see Section 3.5). Recalling Razborov's quote, we aim at formalizations as close as possible to the original arguments. Some changes are, however, needed.

For (a) we formalize in APC_1 an argument close to Furst, Saxe and Sipser's [25] based on probabilistic reasoning with random restrictions. Probabilities are estimated using Jeřábek's notion of approximate counting, and doing so requires the construction of feasible surjections witnessing these estimations. That APC_1 proves the succinct formalization of (a) has already been shown by Krajíček [34, Theorem 15.2.3] formalizing Razborov's abovementioned alternative proof of Håstad's Switching Lemma. His proof is different and of independent interest.

Letting AC^0_d denote the set of circuits of depth $\leqslant d$, and PARITY denote the set of numbers whose binary expansion contains an odd number of ones, the formal statement reads as follows (see Theorem 3.7):

Theorem 1.1. *Let $d, k \in \mathbb{N}$. There is $n_0 \in \mathbb{N}$ such that the theory APC_1 proves*

$$n_0 \leqslant n \to \mathsf{LB}[\mathsf{AC}^0_d, \mathsf{PARITY}](C, n^k, n, N).$$

Razborov and Smolensky's method for (b) typically requires to consider exponentially large objects such as the ring of n-variate polynomials over some finite field. In order to simulate the argument in APC_1 we compromise slightly on our aspired succinctness and assume not only that n is small but even that a fixed quasi-polynomial function of n is small (formally expressed by "$\in \mathit{Log}$" below). As a consolation prize, this scaled down n allows to formulate and prove a lower bound for $s = n^{\log n}$ instead of just n^k. Secondly, polynomials approximating formulas are not constructed directly but instead we construct succinct descriptions of them by arithmetical circuits.

Letting $\mathsf{AC}^0_d[p]$ denote the set of circuits of depth $\leqslant d$ with MOD_p-gates, and MOD_q denote the set of numbers whose binary expansion contains a number of ones divisible by q, the formal statement reads as follows (see Theorem 3.14):

Theorem 1.2. *Let $d \in \mathbb{N}$ and $p \neq q$ be primes. There is $n_0 \in \mathbb{N}$ such that the theory APC_1 proves*

$$n_0 \leqslant n \wedge 2^{\log^{9d} n} \in \mathit{Log} \to \mathsf{LB}[\mathsf{AC}^0_d[p], \mathsf{MOD}_q](C, n^{\log n}, n, N).$$

The proof [5] of the monotone circuit lower bound (c) is formalizable in APC_1 without essential change. Letting MC denote the set of all monotone circuits, and k-CLIQUE the set of (numbers coding) graphs with a clique of size k, the formal statement reads:

Theorem 1.3. *There is a rational* $0 < \epsilon < 1$ *such that for all* $k \geqslant 2$ *there is* $n_0 \in \mathbb{N}$ *such that the theory* APC_1 *proves*

$$n_0 \leqslant n \to \mathsf{LB}[\mathsf{MC}, k\text{-}\mathsf{CLIQUE}](C, n^{\epsilon\sqrt{k}}, n, N).$$

Notably, for each k we get a different hard function. Actually, we prove a more general statement allowing for non-constant k – see Theorem 3.16.

1.3 Independence and natural proofs

Recall that, informally, PV_1 formalizes proofs working with polynomial time computable concepts and constructions, and the central problem is whether PV_1 is able to prove general circuit lower bounds such as $\mathsf{LB}_{\mathsf{tt}}[\mathsf{SAT}]$ for $s = n^k$.

As what can be seen as a partial negative answer Razborov and Rudich [62] observed that many lower bound proofs for an explicit function Q (e.g. (a) and (b)) do exhibit a feasible property of Q restricted to $\{0,1\}^n$ which is not shared by functions computed by the circuit class under consideration. Moreover, this property is after all not that special to Q but true for random functions on $\{0,1\}^n$ with non-negligible probability. Now, if strong pseudorandom generators exist, then such "natural proofs" for superpolynomial lower bounds against general circuits do not exist.

It has been suggested, amongst others by Razborov and Rudich themselves [62, Conclusions], that "the natural proof barrier should be regarded a hint, and not a barrier, to separating complexity classes" [18, p.1587] (see [17, 16, 49, 50] for proposals). In any case, the notion of naturality as a property of proofs is informal and it is questionable whether it could imply independence from PV_1. What Razborov [60] could show is that it rules out proofs in the theory $\mathsf{S}^2_2(\alpha)$ (see e.g. [34, p.98] for a definition, alternative proofs based on propositional feasible interpolation are given in [58, 35, 7]).

We shall formalize the natural proof barrier itself (Theorem 3.27). We work in APC_1^+, a variant of APC_1 from [12], which allows for a relatively smooth formalization of the underlying concepts.

The succinct lower bound $\mathsf{LB}[\mathsf{SAT}]$ for $s = n^k$ is shown in [52] to be unprovable in a theory formalizing NC^1 reasoning unless subexponential size formulas can approximate polynomial size circuits. Relatedly, $\mathsf{LB}[\mathsf{Q}]$ has been shown to be *consistent* with PV_1 for $\mathsf{Q} = \mathsf{SAT}$ in [21] (improving upon [36]) unless the polynomial hierarchy collapses to the Boolean hierarchy, and recently unconditionally for some $\mathsf{Q} \in \mathsf{P}$ [40, 13, 14].

1.4 Succinct tautologies

For every $n \in \mathbb{N}$ statement $\mathsf{LB}_{\mathsf{tt}}[\mathsf{Q}]$, say for $s = n^k$, translates to propositional formulas

$$\mathsf{tt}[\mathsf{Q}, n^k] := \bigvee\nolimits_{a \in \{0,1\}^n} \text{“}C(a) \neq \mathsf{Q}(a)\text{”}, \tag{2}$$

where "$C(a) \neq \mathsf{Q}(a)$" is a propositional formula with variables for the encoding of the circuit C and its computation on a. The formula has size $2^{O(n)}$ and is tautological if and only if the lower bound is true.

It is well-known [20] that PV_1 is simulated by the Extended Frege system EF (see Section 4.1 for a precise statement). In particular, Razborov's [59] PV_1-proofs of (a)-(c) translate to short EF-proofs of the corresponding tt-tautologies. 'Short' means polynomial in the size of the tautology, i.e. $2^{O(n)}$. Unprovability of $\mathsf{LB}_{\mathsf{tt}}[\mathsf{SAT}]$ for $s = n^k$ in PV_1 would be implied by $\mathsf{tt}[\mathsf{SAT}, n^k]$ not admitting short EF-proofs. Consistency of the succinct formula $\mathsf{LB}[\mathsf{SAT}]$ with PV_1 would be implied by lower bounds for EF *with constant advice* (see [21, Theorems 6.8, 3.4]).

The tt-formulas are a particular instance of so-called τ-formulas suggested as candidate hard tautologies independently by Alekhnovich et al. [2] and Krajíček [37], and in some sense the hardest among them (cf. [38]). Not too much is known concerning lower bounds though. The natural proof barrier rules out short proofs of $\mathsf{tt}[\mathsf{Q}, n^{\omega(1)}]$ for sufficiently strong systems with feasible interpolation (cf. [39, Theorem 29.2.3]). Some unconditional lower bounds are known for weak systems with suitably written $\mathsf{tt}[\mathsf{Q}, n^{\omega(1)}]$ (namely for Resolution by Raz [55], and Res(log log) by Razborov [61]). We refer to the Introduction of [61] for a short survey, or to [39, Chapters 27–30] for a more comprehensive one.

We ask whether it is possible to construct by a feasible algorithm propositional proofs of circuit lower bounds expressed succinctly. We study two ways to get such succinct formulas of size $n^{O(1)}$ or at least far smaller than $2^{O(n)}$.

The first is via the succinct formula $\mathsf{LB}[\mathsf{Q}]$ and has been discussed in [54]. Its quantifier complexity is too high to be canonically translated to tautologies, but if the existential quantifier on y in (1) could be witnessed by a polynomial time or P/poly function w, then it does translate to a tautology lb_w of size $n^{O(1)}$. Such a function produces given a circuit C an input string y such that $C(y) \neq \mathsf{Q}(y)$. Of course, the question whether such functions exist is of independent complexity theoretic interest. We observe that they do exist for $\mathsf{Q} = \mathsf{SAT}$ under plausible hardness assumptions (Proposition 4.8).

Our main result concerning lb_w-formulas is a general relative upper bound: we show that APC_1-proofs of succinct lower bounds give lb_w-formalizations such that there are short EF-proofs of lb_w assuming that some function is hard for a specific circuit of subexponential size. We refer to Theorem 4.10 for a precise statement.

The second way is via Lipton and Young's antichecker [43] which allow to move to a size $n^{O(1)}$ subdisjunction of (2) which is still tautological. Intuitively, such a formula should be even harder than the tt-formula because it has the same meaning but is exponentially more succinct. To support the intuition, we observe that hardness of the lavish tt-formulas for constant depth Frege implies hardness of the succinct tautologies for unrestricted Frege (Proposition 4.14). Recently, results of this type have been called *magnification results* [49, 50].

A non-uniform variant of the anti-checked formula has variables for the bits $\mathsf{Q}(a)$.

It expresses a circuit lower bound for a partial function given by a partial truth table. Based on Razborov and Rudich's naturalization of Smolensky's proof of (b) we exhibit a property of such partial truth tables such that the lower bound formulas are tautological whenever a partial function with this property is substituted. We observe that there are many such functions (MOD_q being one of them) and give a quasipolynomial time algorithm producing proofs of these tautologies in Jeřábek's proof system WF – it is to APC_1 as EF is to PV_1 [28]. We refer to Corollary 4.17 (and Remark 4.18) for a precise statement. In other words, we exhibit a succinct version of a natural property. Notably, this is also motivated by a generic learning task described in Section 4.5.

1.5 Overview of dependencies and notation

The material from the preliminary Section 2 is used throughout.

Section 3 on formalizations in APC_1 is based on a preliminary subsection 3.1 after which the following subsections are independent from each other, except that 3.5 states a corollary to the result from 3.4.

Section 4 on propositional formalizations is independent from section 3, except that 4.6 relies on 3.3.

The final section 5 depends on all others.

The following lists, for the convenience of the reader, the most important formalizations of circuit lower bound statements studied in this paper. The list omits some formulas appearing only locally. All formulas express a lower bound against a circuit class C for a function f or a decision problem Q. All formulas depend on PV-formulas defining C and Q, and this dependence is not reflected in the notations. As a rule, we omit C from the notations if it is the class of all circuits.

First-order formalizations Writing $\varphi(\bar{x})$ for a first-order formula means that its free variables are among $\bar{x}$.

- $\mathsf{LB}_{\mathsf{tt}}[\mathsf{C}](f, C, s, n, N)$

  This formula is defined in Section 2.2 and means "if C is a C-circuit of size s, then C does not compute $f : \{0,1\}^n \to \{0,1\}$". The index tt indicates that the function is given by its truth table, namely the binary expansion of f. The variable N is auxiliary and used to ensure that n is of *doubly logarithmic order.*

- $\mathsf{LB}[\mathsf{C}, \mathsf{Q}](C, s, n, N)$

  This formula is defined in Section 2.2 and means "if C is a C-circuit of size s, then C does not decide Q on $\{0,1\}^n$". The variable N is auxiliary and used to ensure that n is of *logarithmic order.*

- $\mathsf{LB}_{\mathsf{ptt}}[\mathsf{C}](f, C, s, n, N, \ell, L)$

  This formula is defined in Section 4.5 and means "if C is a C-circuit of size s, then C does not agree with the size ℓ partial function f". The index ptt stands for partial truth table. The variables N, L are auxiliary and used to ensure that n, ℓ are of *logarithmic order*.

Propositional formalizations All propositional formulas contain auxiliary propositional variables that are not mentioned in the descriptions given below. Their size is measured as a function of n, in particular, $k \in \mathbb{N}$ is treated as a constant.

- $\mathsf{tt}[\mathsf{C}, f, s(2^n)] = \bigvee_{a<2^n}$ "$C(a)\neq f(a)$"

  This formula is defined in Section 4.2 and means "if C is a C-circuit of size $s(2^n)$, then C does not compute $f : \{0,1\}^n \to \{0,1\}$". The function s is in PV. The formula has 2^n variables for the truth table of f and 2^n many variables for (a binary encoding of a circuit) C of size $s(2^n)$. It is a "truth table" formula of size $2^{O(n)}$.

  If one substitutes values of a particular truth table h for the variables corresponding to f, then the resulting formula is tautological if and only if h is not computed by C-circuits of size $s(2^n)$.

- $\bigvee_{a<2^m}$ "$\mathit{circ}(\bar{x}, \tilde{f}, \cdot)(a)\neq\tilde{f}(a)$"

  This formula is defined in Section 4.2 given some $0 < \epsilon < 1$ and means "if $\mathit{circ}(\bar{x}, \tilde{f}, \ldots)$ is a size $2^{\epsilon m}$ circuit, then it does not compute $\tilde{f} : \{0,1\}^m \to \{0,1\}$". The function circ is in PV. The formula has variables for the bits of $\bar{x}$ and 2^m variables for the truth table of $\tilde{f}$. It is a "truth table" formula of size at least 2^m.

  The notational change from f, n to $\tilde{f}, m$ is just to make it fit into the context of its use in Section 4.4. The dot $\cdot$ in "$\mathit{circ}(\bar{x}, \tilde{f}, \cdot)(a)\neq\tilde{f}(a)$" indicates that circ has more arguments which, however, do not correspond to propositional variables.

- $\mathsf{tt}[\mathsf{C}, \mathsf{Q}, n^k] = \bigvee_{a<2^n}$ "$C(a)\neq\mathsf{Q}(a)$"

  This formula is defined in Section 4.2 and means "if C is a C-circuit of size n^k, then C does not decide Q on $\{0,1\}^n$". It has n^{k+1} variables for C. It is a "truth table" formula of size $2^{O(n)}$. Every disjunct "$C(a)\neq\mathsf{Q}(a)$" has size $n^{O(1)}$.

- $\mathsf{lb}_w[\mathsf{C}, \mathsf{Q}, n^k]$

  This formula is defined in Section 4.3 and means "if C is a C-circuit of size n^k, then C fails to decide Q on the string $w(C, \ldots) \in \{0,1\}^n$". Here, $w(C, \ldots)$ is a witnessing function in PV. The formula has n^{k+1} variables for C. It is a "succinct" formula of size $n^{O(1)}$.

- $\mathsf{lb}_{w(\cdot,\bar{z})}[\mathsf{C},\mathsf{Q},n^k]$

  This formula, defined in Section 4.3, is a variant of the previous one where w has additional arguments $\bar{z}$. The dot $\cdot$ indicates that the function w has more arguments which, however, do not correspond to propositional variables.

- $\mathsf{lb}_A[\mathsf{C},\mathsf{Q},n^k] = \bigvee_{a\in A_n}$ "$C(a){\neq}\mathsf{Q}(a)$"

  This formula is defined in Section 4.5 and means "if C is a C-circuit of size n^k, then C fails to decide Q on some string from $A_n \subseteq \{0,1\}^n$". Here, $A = (A_n)_{n\in\mathbb{N}}$ is an antichecker. The formula has n^{k+1} variables for C. It is a "succinct" formula of size polynomial in $|A_n|\cdot n$.

- $\mathsf{ptt}[\mathsf{C},f,s(n),n,\ell] = \bigvee_{i<\ell}$ "$C(x_i) \neq b_i$"

  This formula is defined in Section 4.5 and means "if C is a C-circuit of size $s(n)$, then C does not agree with the partial function f that maps the strings $x_0,\dots,x_{\ell-1} \in \{0,1\}^n$ to the bits $b_0,\dots,b_{\ell-1} \in \{0,1\}$, respectively". The function s is in PV. The formula has $O(s(n)\log s(n))$ variables for C. It is a "succinct" formula of size polynomial in $s(n)\cdot\ell\cdot n$.

2 Preliminaries

2.1 The theory PV_1

The first theory formalizing polynomial time reasoning was introduced by Cook [20]. Its language PV contains $<$ and symbols for all polynomial time functions (over $\mathbb{N}$) introduced inductively according to Cobham's characterization [19, p.28]. We blur the distinction between the symbol and the function, that is, between the symbol and its interpretation in the *standard model* with universe $\mathbb{N}$.

Following [42], PV_1 is a universal theory in the language PV given by Cobham's equations and a scheme equivalent to *induction*

$$\varphi(0,\bar{x}) \wedge \forall y(\varphi(y,\bar{x}) \to (y+1,\bar{x})) \to \varphi(x,\bar{x})$$

for $\varphi(x,\bar{x})$ quantifier-free. We refer to [34, Section 5.3] for a definition. In fact, PV_1 proves induction for formulas in $\Sigma^b_0 = \Pi^b_0$, i.e. PV-formulas with only *sharply bounded* quantifiers $\exists x{<}|t|, \forall x{<}|t|$, where t is a PV-term without x and $|z|$ denotes the length of the binary representation of z (i.e. $|z| = \lceil\log(z+1)\rceil$ in the standard model). 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 quantification and $\exists x{<}t$ (resp. $\forall x{<}t$).

The theory $\mathsf{S}^1_2 = \mathsf{S}^1_2(\mathsf{PV})$ is obtained from PV_1 by adding *length induction*

$$\varphi(0,\bar{x}) \wedge \forall y(\varphi(y,\bar{x}) \to \varphi(y+1,\bar{x})) \to \varphi(|x|,\bar{x})$$

for $\varphi(x,\bar{x}) \in \Sigma^b_1$. It is Σ^b_1-conservative over PV_1 by [6]:

Theorem 2.1 (Buss' Witnessing)**.** *If S^1_2 proves $\exists y \varphi(y,\bar{x})$ for $\varphi(y,\bar{x}) \in \Sigma^b_1$, then PV_1 proves $\varphi(f(\bar{x}),\bar{x})$ for some function symbol $f(\bar{x})$ in PV.*

Let n, m, N be variables. We write $n \in \mathit{Log}$ for $\exists N\ n = |N|$, and $n \in \mathit{LogLog}$ for $\exists N\ n = ||N||$. In a context where $n = |N|$ we write 2^n for $1\#N$. We view numbers below 2^n as n-bit strings. There is $\mathit{eval} \in \mathsf{PV}$ denoting (in the standard model) the circuit evaluation function: for a circuit C with n inputs $C(x) := \mathit{eval}(C,x)$ for $x < 2^n$ is the value computed by C on x; if C has m outputs then this value is a number $< 2^m$. The *size* of a circuit is the number of inner (non-input) gates. The following is folklore.

Proposition 2.2. *For every $f \in \mathsf{PV}$ there are $\ell, k \in \mathbb{N}$ such that the theory PV_1 proves that for every $n \in \mathit{Log}$ there exists a size n^ℓ circuit C with n inputs and n^k outputs such that $f(x) = C(x)$ for all $x < 2^n$.*

Like 2^n we use similar suggestive notation for other fast growing functions when applied to arguments $n = |N|$ in Log. For example, for $f \in \mathsf{PV}$ we write $\sum_{i<n} f(i)$ for a PV-symbol $g(N)$ such that PV_1 proves $g(2N) = g(2N+1) = g(N) + f(|N|)$. Similarly for $\prod_{i<n} f(i)$. For example, PV_1 proves $N = \sum_{i<n} \mathit{bit}(i,N) \cdot 2^i$ for a suitable $\mathit{bit} \in \mathsf{PV}$; we understand that $\mathit{bit}(i,N) = 0$ for $i \geqslant n$. Rationals a/b are naturally coded by pairs and we use them freely in equations and inequalities. E.g. $a/b \in \mathit{Log}$ means $\exists c\ a/b \leqslant c \in \mathit{Log}$. This allows to formally use $n!$ and $\binom{n}{i}$ for $i \leqslant n$. For example, PV_1 proves $\sum_{i=0}^n \binom{n}{i} = 2^n$.

We shall need the following less trivial calculations in PV_1.

Proposition 2.3 (Stirling's bound, Jeřábek [28])**.** *There is a $c > 1$ such that PV_1 proves:*

$$0 < k < n \in \mathit{Log} \ \rightarrow\ \frac{1}{c}\binom{n}{k} < \frac{n^n}{k^k(n-k)^{n-k}} \cdot \Big(\Big\lfloor\sqrt{\frac{k(n-k)}{n}}\Big\rfloor + 1\Big)^{-1} < c\binom{n}{k}.$$

Proposition 2.4. *For every rational $\epsilon > 0$ there is an $n_0 \in \mathbb{N}$ such that PV_1 proves:*

$$n_0 < n \in \mathit{Log} \ \rightarrow \sum_{i=0}^{\lfloor n/2+n^{1/3}\rfloor} \binom{n}{i} < (1/2+\epsilon)\cdot 2^n.$$

Proof. Argue in PV_1. We have

$$\sum_{i=0}^{\lfloor n/2\rfloor -1} \binom{n}{i} = \frac{1}{2}\Big(\sum_{i=0}^{\lfloor n/2\rfloor -1} \binom{n}{i} + \sum_{i=0}^{\lfloor n/2\rfloor -1} \binom{n}{n-i}\Big) < 2^{n-1}$$

and by Stirling's bound, for some constant $c > 1$,

$$\sum_{i=\lfloor n/2\rfloor}^{\lfloor n/2+n^{1/3}\rfloor} \binom{n}{i} < (n^{1/3}+1)\binom{n}{\lfloor n/2\rfloor} < 2^n 4c\Big(\frac{n^{1/3}}{\lfloor n^{1/2}/2\rfloor} + \frac{1}{\lfloor n^{1/2}/2\rfloor}\Big),$$

where to verify the last inequality for odd n we also used $(1+a/b) \leqslant 4^{a/b}$ for $a, b \in \mathit{Log}$, $b > 0$ as shown in [28, Stirling's bound, Claim 1]. □

Proposition 2.5. PV_1 *proves:*

$$n < m \in \mathit{Log} \ \rightarrow\ (m-n)^m \leqslant m^m/2^n \wedge (1-n/m) \leqslant 2^{-n/m}.$$

Proof. Note the second conjunct of the conclusion follows from the first. To prove the first one, we proceed as in [28, Proposition A.2, Claim 2]. Specifically, the case $n = 0$ is trivial and for $n = 1$ we have $(m-1)^m \leqslant m^m/2$ obtained by expanding $(m-1)^m$. Then, assuming $n+1 < m \in \mathit{Log}$ and using the induction hypothesis for 1 and n, we get

$$(m-n-1)^{m-n} \leqslant (m-n)^{m-n}/2 \leqslant m^m/(2^{n+1}(m-n)^n) \leqslant m^m/(2^{n+1}(m-n-1)^n),$$

hence $(m-(n+1))^m \leqslant m^m/2^{n+1}$. □

2.2 Two formalizations of circuit lower bounds

As outlined in the introduction we give two PV-formulas expressing a size s lower bound for circuits from a class C computing a function $f : \mathbb{N} \rightarrow \{0,1\}$ on (numbers smaller than) 2^n which play the role of binary strings of length n.

We assume throughout that the class of circuits C is in polynomial time, and more precisely, that it is defined (in the standard model) by a Σ^b_0-formula. In particular,

$$\text{“}C \text{ is a } \mathsf{C}\text{-circuit of size} \leqslant s\text{”} \tag{3}$$

is a Σ^b_0-formula with free variables C and s.

The two formalizations use a dummy variable N which the formulas suppose to be either such that $2^n = |N|$ or such that $n = |N|$. In the intuitive mode of speech from the introduction, the different scalings used by the two formulas are thus made explicit.

The two formulas can be obtained following two ways of how to make up one's mind about the "a little bit annoying"[59, p.377] problem of what is meant by an *explicit* function f. The first is to assume $n \in \mathit{LogLog}$, so f restricted to (numbers smaller than) 2^n is given by a number whose binary expansion codes its truth table:

$$\mathsf{LB}_{\mathsf{tt}}[\mathsf{C}](f, C, s, n, N) := \exists y{<}|N|\ \mathsf{LB}^0_{\mathsf{tt}}[\mathsf{C}](f, C, s, n, N, y), \tag{4}$$

$$\begin{aligned}&\mathsf{LB}^0_{\mathsf{tt}}[\mathsf{C}](f, C, s, n, N, y) :=\\ &\qquad \Big(2^n = |N| \rightarrow \big(C \text{ is a } \mathsf{C}\text{-circuit of size} \leqslant s \rightarrow C(y) \neq \mathit{bit}(y, f)\big)\Big).\end{aligned} \tag{5}$$

Recall $C(y)$ abbreviates $\mathit{eval}(C, y)$. The antecedent $2^n = |N|$ defines a polynomial time relation between n and N and can thus be represented by a Σ^b_0-formula. Thus $\mathsf{LB}_{\mathsf{tt}}[\mathsf{C}](f, C, s, n, N)$ is Σ^b_0.

Somewhat less explicitly, one views f as the characteristic function of the computational problem $\mathsf{Q} := f^{-1}(1)$ and uses a formula defining Q. We denote this formula by $\mathsf{Q}(y)$. Such a formalization works supposing only $n \in \mathit{Log}$. More precisely, define

$$\begin{aligned} &\mathsf{LB}[\mathsf{C},\mathsf{Q}](C,s,n,N) := \exists y{<}1\#N\ \mathsf{LB}^0[\mathsf{C},\mathsf{Q}](C,s,n,N,y), && (6)\\ &\mathsf{LB}^0[\mathsf{C},\mathsf{Q}](C,s,n,N,y) := \\ &\qquad \Big(n = |N| \to \big(C \text{ is a } \mathsf{C}\text{-circuit of size} \leqslant s \to \big(C(y) = 1 \oplus \mathsf{Q}(y)\big)\big)\Big). && (7)\end{aligned}$$

Here, $\oplus$ denotes exclusive disjunction. Note that the existential quantifier on y is not sharply bounded anymore. If $\mathsf{Q} \in \mathsf{P}$ or $\mathsf{Q} \in \mathsf{NP}$, then the formula $\mathsf{Q}(x)$ can be chosen Σ^b_0 or Σ^b_1 respectively, and then $\mathsf{LB}[\mathsf{C},\mathsf{Q}](C,s,n,N)$ becomes Σ^b_1 and Σ^b_2 respectively.

We do not mention C if it is the class of all circuits, so the resulting formulas are denoted $\mathsf{LB}_{\mathsf{tt}}, \mathsf{LB}^0_{\mathsf{tt}}, \mathsf{LB}[\mathsf{Q}], \mathsf{LB}^0[\mathsf{Q}]$.

Remark 2.6. Corresponding to Razborov's formulas [59] mentioned in the introduction, a truth table formalization of a circuit lower bound for a fixed problem Q would read

$$\mathsf{LB}_{\mathsf{tt}}[\mathsf{C},\mathsf{Q}](C,s,n,N) := \mathsf{LB}[\mathsf{C},\mathsf{Q}](C,s,n,|N|) \tag{8}$$

in our formalism. We are not going to use these formulas.

Note a circuit of size s is coded by a number of length $O(s\cdot|s|)$, so formally quantifying over circuits of size $\leqslant s$ is meaningful only for $s \in \mathit{Log}$. In the $\mathsf{LB}_{\mathsf{tt}}$-formula this allows $s \leqslant 2^{(1-o(1))n}$ while the LB-formula allows only $s = n^{O(1)}$. We repeat the intuition from the introduction for $s \leqslant n^{O(1)}$. Choosing the scale of n means choosing the "feasible object". In the $\mathsf{LB}_{\mathsf{tt}}$-formulas $n \in \mathit{LogLog}$, so the truth-table (and everything polynomial in it) is feasible. The LB-formalization assumes just that $n \in \mathit{Log}$. This means that only the objects of polynomial-size in (n or) the size of the circuit are feasible. Likewise, a theory reasoning about the circuit lower bound becomes less resp. more powerful when working with LB resp. $\mathsf{LB}_{\mathsf{tt}}$.

2.3 The theory APC_1

We want to formally talk about the size of bounded definable sets $X = \{x < a \mid \varphi(x,\bar{x})\}$. These are not formal objects in our first-order language but a mode of speech: we let $x \in X$ stand for $(x < a \wedge \varphi(x,\bar{x}))$. We write $X \subseteq a$ instead of $X \subseteq [0,a)$. We often write a instead of $[0,a)$; for a rational a, this means $[0,\lfloor a\rfloor)$. With $X \subseteq a$, $Y \subseteq b$, also

$$\begin{aligned} X \mathbin{\dot\cup} Y &:= X \cup \{y + a \mid y \in Y\} \subseteq a + b,\\ X \times Y &:= \{bx + y \mid x \in X, y \in Y\} \subseteq ab,\end{aligned}$$

are definable; we write $\langle x,y\rangle$ for $bx+y$ in such a context.

In PV_1 'small' sets can be counted precisely in the sense that every definable $X \subseteq n$ for $n \in \mathit{Log}$ is coded by a number $\ulcorner X \urcorner$ and hence bijective via some coded bijection to a unique number $\mathit{Card}(\ulcorner X \urcorner)$ which we write as $\mathit{Card}(X)$ (see e.g. [34, Section 5.4]). Obviously, if $\mathsf{sWPHP}(\mathsf{PV})$ fails (see below), then there is no reasonable notion of size for 'large' definable sets $X \subseteq 2^n$, even quantifier free, i.e. *circuit definable*: $X = \{x < 2^n \mid C(x) = 1\}$ for a circuit C with n variables. Complexity theory in models of PV_1 where $\mathsf{sWPHP}(\mathsf{PV})$ fails is studied in [33]. Here, $\mathsf{sWPHP}(\mathsf{PV})$ is the *surjective weak pigeonhole principle* for PV-functions: the set containing the formula

$$\mathsf{sWPHP}(f) \ := \ (x > 0 \to \exists v{<}x(|y|+1)\ \forall u{<}x|y|\ f(u,\bar{x}) \neq v) \tag{9}$$

for each $f(u,\bar{x}) \in \mathsf{PV}$. Equivalently one can take the single formula obtained by replacing $f(u,\bar{x})$ with $C(u) = \mathit{eval}(C,u)$ (Proposition 2.2).

Following the notation of [12], we are led to consider

$$\mathsf{APC}_1 := \mathsf{PV}_1 + \mathsf{sWPHP}(\mathsf{PV}).$$

In the Introduction we informally referred to APC_1 as a "slight" extension of PV_1. One reason is that $\mathsf{sWPHP}(\mathsf{PV})$ is provable in T^2_2 [44], so APC_1 is quite low in the hierarchy of bounded arithmetics. But APC_1 appears to be considerably weaker than T^2_2 (see [11, 4] for recent results). In terms of witnessing the step from PV_1 to APC_1 is that from polynomial time to probabilistic polynomial time. This is is due to Wilkie and first published in [34, Theorem 7.3.7]. An alternative proof has been given by Thapen [65, Theorem 4.2], which, as observed in [28, Corollary 1.15], also yields the first statement in:

Theorem 2.7 (Wilkie's witnessing)**.** $\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV})$ *is* Σ^b_1*-conservative over* APC_1*. If one of these theories proves* $\exists y \varphi(y,\bar{x})$ *for* $\varphi(y,\bar{x}) \in \Sigma^b_1$*, then there exists a probabilistic polynomial time Turing machine which given a tuple* $\bar{n}$ *from* $\mathbb{N}$ *outputs with probability at least* $2/3$ *some* $m \in \mathbb{N}$ *such that* $\varphi(m,\bar{n})$ *is true in the standard model.*

The probability $2/3$ can be boosted and the probabilistic computation is definable in some suitable sense – see [28]. Formal approximate counting has been developed by Jeřábek in his PhD Thesis [29] and a sequence of papers [28, 30, 31, 32]. In particular, [30] showed that APC_1 supports a well-behaved notion of approximate cardinality.

Definition 2.8 (in PV_1)**.** Let $n \in \mathit{Log}$, and $X, Y \subseteq 2^n$ be definable. For a circuit C with n variables and m output gates, we write

$$C : X \twoheadrightarrow Y$$

for $\forall y \in Y\ \exists x \in X\ C(x) = y$. For $0 \leqslant \epsilon \leqslant 1$ define $Y \preccurlyeq_\epsilon X$ if and only if there exist a circuit C and $v \neq 0$ such that

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

We say C *witnesses* $Y \preccurlyeq_\epsilon X$. Further, $X \approx_\epsilon Y$ means $(X \preccurlyeq_\epsilon Y \wedge Y \preccurlyeq_\epsilon X)$.

One easily checks (in PV_1) that $X \subseteq Y$ implies $X \preccurlyeq_0 Y$, and that $(X \preccurlyeq_\epsilon Y \wedge Y \preccurlyeq_\delta Z)$ implies $X \preccurlyeq_{\epsilon+\delta} Z$. The main result of [30, Theorem 2.7] implies that in APC_1 every circuit definable set does have an approximate cardinality. Moreover, this is witnessed by invertible circuits. A circuit $C : a \to b$ is *invertible* if there is a circuit D such that $C \circ D$ is the identity on b, i.e.,

$$\forall z{<}b \; (D(z) < a \wedge C(D(z)) = z).$$

Theorem 2.9. *The theory APC_1 proves that for all $n, \epsilon^{-1} \in Log$ and every circuit definable $X \subseteq 2^n$ there exists $s \leqslant 2^n$ such that $X \approx_\epsilon s$. Moreover, both $X \preccurlyeq_\epsilon s$ and $X \succcurlyeq_\epsilon s$ are witnessed by invertible circuits.*

The proof uses the Nisan-Wigderson generator [48] to sample X and thus get an estimate of its size. It is for this "production of magic surjections" [32, p.842] why the "extra complication is necessary"[30, p.963] to make v copies in Definition 2.8. This theorem allows to show [30, Lemma 2.11]:

Proposition 2.10. *The theory APC_1 proves for all circuit definable $X, Y \subseteq 2^n$ and $s, t, u \leqslant 2^n$ and $\epsilon, \delta, \theta, \gamma < 1$ with $\gamma^{-1} \in Log$:*

(i) $X \preccurlyeq_\gamma Y$ or $Y \preccurlyeq_\gamma X$,

(ii) If $s \preccurlyeq_\epsilon X \preccurlyeq_\delta t$, then $s < t + (\epsilon + \delta + \gamma)2^n$,

(iii) If $X \preccurlyeq_\epsilon Y$, then $2^n \setminus Y \preccurlyeq_{\epsilon+\gamma} 2^n \setminus X$,

(iv) If $X \approx_\epsilon s$ and $Y \approx_\delta t$ and $X \cap Y \approx_\theta u$, then $X \cup Y \approx_{\epsilon+\delta+\theta+\gamma} s + t - u$.

The definition of $\preccurlyeq_\epsilon$ is an unbounded $\exists\Pi^b_2$-formula so cannot be used freely in bounded induction. Therefore Jeřábek introduces the theory HARD^A [28] in the language $\mathsf{PV}(\alpha)$ containing symbols for all polynomial time algorithms with oracle α. Roughly, the theory is given by the analogue of APC_1 in this language plus axioms stating that $\alpha(x)$ is the truth table of an average-case hard function with $||x||$ variables. This removes "the only non-uniformity" in the proof of Theorem 2.9, namely, "the choice of the hard function"[30, Section 3.1] fueling the Nisan-Wigderson generator. Jeřábek shows (see [30, Theorem 2.13, Lemma 2.14]):

Lemma 2.11. *The theory HARD^A is conservative over APC_1 and its language contains seven ternary function symbols such that it proves the following. For all $n, \epsilon^{-1} \in Log$ and every circuit C defining a set $X \subseteq 2^n$ the seven function symbols applied to $C, 2^n, \epsilon^{-1}$ yield circuits G_0, H_0, G_1, H_1 and numbers s, v, w such that $G_0 : v \times (X \mathbin{\dot\cup} \epsilon 2^n) \twoheadrightarrow v \cdot s$, $G_1 : w \cdot (s + \epsilon 2^n) \twoheadrightarrow w \times X$, $G_0 \circ H_0$ is the identity on $v \cdot s$, and $G_1 \circ H_1$ is the identity on $w \times X$.*

Having induction allows to prove [30, Proposition 2.15] and [30, Proposition 2.16] (the version with $\preccurlyeq$ replacing $\succcurlyeq$):

Proposition 2.12 (Disjoint union)**.** *The theory* APC_1 *proves for* $\epsilon, \delta \leqslant 1$ *and* $n, m, \delta^{-1} \in$ *Log and a sequence of circuits defining a sequence* $(X_i)_{i<m}$ *of subsets of* 2^n *and a sequence* $(s_i)_{i<m}$ *the following. If* $X_i \preccurlyeq_\epsilon s_i$ *for all* $i < m$*, then* $\bigcup_{i<m}(X_i \times \{i\}) \preccurlyeq_{\epsilon+\delta} \sum_{i<m} s_i$*.*

Here, the disjoint union $\bigcup_{i<m}(X_i \times \{i\})$ is viewed as a subset of $2^{n+|m|}$.

Proposition 2.13 (Averaging)**.** *The theory* APC_1 *proves for* $\epsilon, \delta, \gamma \leqslant 1$ *and* $n, m, \gamma^{-1} \in$ *Log and circuit definable* $Z \subseteq 2^n \times 2^m$ *and* $Y \subseteq 2^m$ *and all* a, b *the following. If* $Y \succcurlyeq_\epsilon b$ *and* $Z_y \succcurlyeq_\delta a$ *for all* $y \in Y$*, then* $Z \cap (2^n \times Y) \succcurlyeq_{\epsilon+\delta+\epsilon\delta+\gamma} ab$*.*

Here and below we write Z_y for $\{x < 2^n \mid \langle x, y\rangle \in Z\}$ where $Z \subseteq 2^n \times 2^m$ and $y < 2^m$. We illustrate a typical use of HARD^A by the following lemma.

Lemma 2.14. *The theory* HARD^A *proves for all* $n, m, \gamma^{-1} \in$ *Log and all circuit definable* $W, Z \subseteq 2^n \times 2^m$ *the following. If* $W_y \preccurlyeq_\gamma Z_y$ *for every* $y < 2^m$*, then* $W \preccurlyeq_{8\gamma} Z$*.*

Proof. Argue in HARD^A and fix $\gamma^{-1} \in$ *Log*. There are function symbols $s(y), t(y) \in \mathsf{PV}(\alpha)$ such that $W_y \approx_\gamma s(y)$ and $Z_y \approx_\gamma t(y)$ for all $y < 2^m$. Here and below we suppress the arguments γ^{-1} and 2^n of these functions. Moreover, there are $v(y), w(y) \in \mathsf{PV}(\alpha)$ and $f_0(y), g_0(y) \in \mathsf{PV}(\alpha)$, outputting circuits, such that

$$f_0(y) : v(y) \times (Z_y \dot\cup \gamma 2^n) \twoheadrightarrow v(y) \cdot t(y),$$
$$g_0(y) : w(y) \cdot (s(y) + \gamma 2^n) \twoheadrightarrow w(y) \times W_y.$$

By Proposition 2.10 (ii), $s(y) \leqslant t(y) + 4\gamma 2^n$. Modifying $f_0(y)$ we get $f_1(y) \in \mathsf{PV}(\alpha)$

$$f_1(y) : v(y) \times (Z_y \dot\cup 5\gamma 2^n) \twoheadrightarrow v(y) \cdot s(y).$$

We can replace every $v(y)$ and $w(y)$, for $y < 2^m$, by a sufficiently large upper bound u if we increase the error by γ. More precisely, there are $f_2(y), g_1(y) \in \mathsf{PV}(\alpha)$ such that

$$f_2(y) : u \times (Z_y \dot\cup 6\gamma 2^n) \twoheadrightarrow u \cdot s(y),$$
$$g_1(y) : u \cdot (s(y) + 2\gamma 2^n) \twoheadrightarrow u \times W_y.$$

Indeed, from $g_0(y)$ and any $u \geqslant w(y)$ construct a circuit computing a surjection from $\lceil u/w(y)\rceil \cdot w(y) \cdot (s(y) + \gamma 2^n)$ onto $(\lceil u/w(y)\rceil \cdot w(y)) \times W_y$ and hence onto $u \times W_y$; to get the domain down to u many copies (of $s(y) + \gamma 2^n \leqslant 2^{2n}$) we have to delete $< w(y)$ many of them; this is achieved by increasing the error by γ and mapping the $u\gamma 2^n$ many points gained onto the undesired copies; this is possible if $u\gamma 2^n \geqslant w(y) 2^{2n}$. The function f_2 is defined similarly, so it suffices to choose u above $\gamma^{-1} v(y) w(y) 2^n$ for all $y < 2^m$.

A suitable composition yields $h(y) \in \mathsf{PV}(\alpha)$ such that

$$h(y) : u \times (Z_y \dot\cup 8\gamma 2^n) \twoheadrightarrow u \times W_y.$$

From this construct a circuit C such that $C : u \times (Z \dot\cup 8\gamma 2^{n+m}) \twoheadrightarrow u \times W$, as desired. □

3 Succinct circuit lower bounds in APC_1

3.1 Approximate probabilistic reasoning

Approximate counting can be formulated as approximate probabilistic reasoning.

Definition 3.1 (in APC_1)**.** For circuit definable $X \subseteq 2^{|t|}$ and $Z \subseteq 2^{|t|} \times 2^{|s|}$ and $0 \leqslant \epsilon, p \leqslant 1$ define

$$\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p \iff \{x \in X \mid x < t\} \preccurlyeq_\epsilon pt,$$
$$\Pr_{\substack{x<t\\y<s}}[\langle x, y\rangle \in Z] \preccurlyeq_\epsilon p \iff \{\langle x, y\rangle \in Z \mid x < t, y < s\} \preccurlyeq_\epsilon pts$$

(recall $\langle x, y\rangle = x2^{|s|} + y$). We use similar notation for $\succcurlyeq_\epsilon$ and $\approx_\epsilon$.

We often use a probability bound p without checking that $p \leqslant 1$. This is needed for the notation to be defined. In such cases, we mean $\max\{p, 1\}$ instead of p.

The following lemma comprises the properties of approximate probabilities we are going to use.

Lemma 3.2. *The theory* APC_1 *proves the following statements for* $0 \leqslant \epsilon, \delta, \gamma, p, q \leqslant 1$, $m, \gamma^{-1} \in Log$*, circuit definable sets* $X, Y \subseteq 2^{|t|}$ *and* $Z \subseteq 2^{|t|} \times 2^{|s|}$*, a sequence* $(X_i)_{i<m}$ *of subsets of* $2^{|t|}$ *given by a sequence of circuits, and a sequence* $(p_i)_{i<m}$ *of rationals.*

(i) If $\Pr_{x<t}[x \in X] \preccurlyeq_{\epsilon+\delta} p$*, then* $\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p + 2\delta$*.*

If $\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p + \delta$*, then* $\Pr_{x<t}[x \in X] \preccurlyeq_{\epsilon+\delta} p$*.*

(ii) If $\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p$ *and* $\Pr_{x<t}[x \in Y] \preccurlyeq_\delta q$*, then* $\Pr_{x<t}[x \in X \cup Y] \preccurlyeq_{\epsilon+\delta} p + q$*.*

If $\Pr_{x<t}[x \in X_i] \preccurlyeq_\epsilon p_i$ *for all* $i < m$*, then* $\Pr_{x<t}\left[x \in \bigcup_{i<m} X_i\right] \preccurlyeq_{2\epsilon m+\gamma} \sum_{i<m} p_i$*.*

(iii) If $\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p$*, then* $\Pr_{x<t}[x \notin X] \succcurlyeq_{\epsilon+\gamma} 1 - p$*.*

If $\Pr_{x<t}[x \in X] \succcurlyeq_\epsilon p$*, then* $\Pr_{x<t}[x \notin X] \preccurlyeq_{\epsilon+\gamma} 1 - p$*.*

(iv) If $\Pr_{x<t,y<s}[\langle x, y\rangle \in Z] \preccurlyeq_\epsilon p$*, then* $\Pr_{x<t}[\langle x, y\rangle \in Z] \preccurlyeq_\epsilon p + 8\epsilon + \gamma$ *for some* $y < s$*.*

If $\Pr_{x<t,y<s}[\langle x, y\rangle \in Z] \succcurlyeq_\epsilon p$*, then* $\Pr_{x<t}[\langle x, y\rangle \in Z] \succcurlyeq_\epsilon p - 8\epsilon - \gamma$ *for some* $y < s$*.*

Proof. (i): note $\Pr_{x<t}[x \in X] \preccurlyeq_{\epsilon+\delta} p$ means there are $v > 0$ and a circuit computing a surjection from $v \times (pt + (\epsilon + \delta)2^{|t|})$ onto $v \times (X \cap t)$. But note the domain is a subset of $v \times (pt + 2\delta t + \epsilon 2^{|t|})$. The second statement is similar.

(ii): the first statement is easy. For the second statement, let $\gamma^{-1} \in Log$ be given. Recall that the disjoint union of the X_i's is viewed as a subset of $2^{|t|+|m|}$. For any $\delta^{-1} \in Log$, Proposition 2.12 gives a circuit C and $v > 0$ such that

$$C : v \times \left(\left(\sum_{i<m} p_i\right) \cdot t + (\epsilon + \delta) \cdot 2^{|t|+|m|}\right) \twoheadrightarrow v \times \bigcup_{i<m}((X_i \cap t) \times \{i\}).$$

This also gives a surjection onto $v \times \bigcup_{i<m}(X_i \cap t)$, with the same domain. Recall that $\bigcup_{i<m}(X_i \cap t)$ is viewed as a subset of $2^{|t|}$. Thus our claim follows noting

$$(\epsilon + \delta) \cdot 2^{|t|+|m|} \leqslant (2\epsilon m + 2\delta m) \cdot 2^{|t|},$$

and choosing $\delta := \gamma/(2m)$.

(iii): we only show the first statement. If $\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p$, then

$$(X \cap t) \cup [t, 2^{|t|}) \preccurlyeq_\epsilon \lfloor pt \rfloor \cup [t, 2^{|t|}).$$

Applying Proposition 2.10 (iii) yields

$$\lfloor (1-p)t \rfloor \preccurlyeq_0 [\lfloor pt \rfloor, t) = 2^{|t|} \setminus \big(\lfloor pt \rfloor \cup [t, 2^{|t|}) \big) \preccurlyeq_{\epsilon+\gamma} 2^{|t|} \setminus \big((X \cap t) \cup [t, 2^{|t|}) \big) = t \setminus X.$$

(iv): the second statement follows knowing the first and (iii) for all $\gamma^{-1} \in \mathit{Log}$. We prove the first statement only in the interesting case that $\gamma \cdot ts \geqslant 1$ (otherwise $ts \in \mathit{Log}$). Assume $\Pr_{x<t,y<s}[\langle x,y \rangle \in Z] \preccurlyeq_\epsilon p$ and note this means

$$\tilde{Z} := \{\langle x,y \rangle \in Z \mid x < t, y < s\} \preccurlyeq_\epsilon pts. \tag{10}$$

Appealing to (i), it suffices to show for arbitrary $\gamma^{-1} \in \mathit{Log}$ that

$$\{x \mid \langle x,y \rangle \in \tilde{Z}\} = \{x \mid \langle x,y \rangle \in Z\} \cap t \preccurlyeq_{\epsilon+\gamma} \tilde{p}t$$

for some $y < s$, where we abbreviate $\tilde{p} := p + (8\epsilon + 13\gamma)$. But if there is no such $y < s$, then $\{x \mid \langle x,y \rangle \in \tilde{Z}\} \succcurlyeq_{\epsilon+\gamma} \tilde{p}t$ for all $y < s$ by Proposition 2.10 (i). Applying Proposition 2.13 (with $Y := [0,s), a := \tilde{p}t, \epsilon := 0, \delta := \epsilon + \gamma, \gamma := \gamma$) yields

$$\tilde{Z} = \tilde{Z} \cap (2^{|t|} \times s) \succcurlyeq_{\epsilon+2\gamma} \tilde{p}ts. \tag{11}$$

Proposition 2.10 (ii) applied to (10) and (11) gives

$$\lfloor \tilde{p}ts \rfloor < \lfloor pts \rfloor + (2\epsilon + 3\gamma) \cdot 2^{|t|+|s|}.$$

But the r.h.s. is $\leqslant \lfloor pts \rfloor + (2\epsilon + 3\gamma) \cdot 2t \cdot 2s \leqslant \tilde{p}ts - \gamma \cdot ts$, a contradiction if $\gamma \cdot ts \geqslant 1$. □

Remark 3.3. Note that (i) and the first statement of (ii) do not require $\mathsf{sWPHP}(\mathsf{PV})$.

3.2 Parity lower bound for AC^0 circuits via random restrictions

By an AC^0_d-*circuit*, where $d \in \mathbb{N}$, we mean a depth $\leqslant d$ unbounded fan-in circuit with gates labeled $0, 1, \neg, \bigwedge, \bigvee$. The *depth* is the maximum length (number of edges) of a path from an input gate to an output gate. By the *size* of a circuit we mean the number of its inner gates. We formalize in APC_1 a lower bound for such circuits computing the parity function via a Switching Lemma which we prove by approximate probabilistic reasoning with random restrictions. Our argument is close to the one presented in [25]. We code restrictions as follows.

For $n \in Log$ and a (formal) rational $0 \leqslant a/b \leqslant 1$ we code a restriction of n propositional variables $x_1, \ldots, x_n$ by the number $\rho = \sum_{i=0}^{n-1} r_{i+1}(2b)^i, r_i < 2b$, and use the following suggestive notation that takes a, b understood from context: $\rho(x_i) = x_i$ means $r_i \in [0, 2a)$; $\rho(x_i) = 1$ means $r_i \in [2a, b+a)$, and $\rho(x_i) = 0$ means $r_i \in [b+a, 2b)$. If $\rho(x_i) = x_i$ we say ρ *leaves* x_i *unassigned*; note that for $a = 1$ this means $r_i < 2$.

The notation $\rho \sim R^n_{a/b}$ stands for $\rho < (2b)^n$. It is straightforward to construct, for $1 \leqslant i \leqslant n$, the circuits witnessing

$$\begin{aligned} \Pr_{\rho \sim R^n_{a/b}}[\rho(x_i) = x_i] &\approx_0 a/b, \\ \Pr_{\rho \sim R^n_{a/b}}[\rho(x_i) = 1] &\approx_0 \frac{1 - a/b}{2} \approx_0 \Pr_{\rho \sim R^n_{a/b}}[\rho(x_i) = 0]. \end{aligned}$$

If $C = C(x_1, \ldots, x_n)$ is a circuit in at most the variables listed, then $C{\restriction}\rho$ is the circuit $C(\rho(x_1), \ldots, \rho(x_n))$ obtained by relabeling input gates as indicated (without further simplifications). Given yet another restriction $\rho' \sim R^n_{a'/b'}$ we write $C{\restriction}\rho\rho'$ for $(C{\restriction}\rho){\restriction}\rho'$.

Definition 3.4. A DNF C *depends on* $> b$ *variables* if there does not exist a sequence of b (not necessarily distinct) variables with the property that every assignment to it either satisfies (all literals in) some disjunct or falsifies (at least one literal in) each disjunct. For CNFs this is analogously defined.

Note that for fixed standard $b \in \mathbb{N}$ the characteristic function of this property is in PV. This ensures the existence of circuits defining events involving this property, as required by approximate counting in $\mathsf{APC_1}$.

In the following we understand that irrational terms are rounded down on the innermost level unless specified otherwise, e.g. $(1/n^{1/2})^c$ is $(1/\lfloor n^{1/2} \rfloor)^c$ and $2 \log n$ is $2\lfloor \log n \rfloor$.

Lemma 3.5 (Switching Lemma)**.** *For every $k \in \mathbb{N}$ there are $b, n_0 \in \mathbb{N}$ such that $\mathsf{APC_1}$ proves: for every $n_0 < n, \epsilon^{-1} \in Log$ and DNF $D_n(x_1, \ldots, x_n)$ of size n^k:*

$$\Pr_{\substack{\rho_1 \sim R^n_{1/n^{1/2}} \\ \rho_2 \sim R^n_{1/n^{1/4}}}} [D_n{\restriction}\rho_1\rho_2 \textit{ depends on} > b \textit{ variables}] \preccurlyeq_\epsilon 1/n^{2k}.$$

The same holds for CNFs.

Proof. We prove the lemma for DNFs, the second statement follows from the first. We follow a familiar proof of the switching lemma estimating the probabilities that formulas simplify under random restrictions. The probabilities are approximated by $\preccurlyeq_\epsilon$. The extra work then boils down to the construction of surjections witnessing the inequalities $\preccurlyeq_\epsilon$. These constructions are postponed to the end of the proof.

Let n be sufficiently large and $n, \epsilon^{-1} \in Log$. Set $d := 12k$. Then

$$\begin{aligned}\Pr_{\rho_1}\Big[\rho_1 \text{ does not falsify all disjuncts in } D_n \text{ of size } \geqslant d\log n\Big] \\ \preccurlyeq_0 n^k \cdot \Big(1 - \frac{1-1/n^{1/2}}{2}\Big)^{d\log n} \leqslant n^k \cdot \Big(1-1/4\Big)^{d\log n} \leqslant 1/n^{3k}, \qquad (12)\end{aligned}$$

where we understand $\rho_1 \sim R^n_{1/n^{1/2}}$. Set $c := 12k + 3d = 48k$. Then

$$\begin{aligned}\Pr_{\rho_1}\Big[\rho_1 \text{ leaves } \geqslant c \text{ variables in some size } \leqslant d\log n \text{ disjunct of } D_n \text{ unassigned}\Big] \\ \preccurlyeq_0 n^k \cdot \Big(\frac{1}{n^{1/2}}\Big)^c \cdot 2^{d\log n} \leqslant 1/n^{3k} \qquad (13)\end{aligned}$$

where for simplicity we bound $\lfloor n^{1/2} \rfloor$ by $n^{1/3}$ when rounding.

Therefore, by the first statement of Lemma 3.2 (ii), the probability that $D_n\lceil\rho_1$ after a trivial simplification is not a c-DNF is $\preccurlyeq_0 2/n^{3k}$. Now it suffices to show:

Claim 3.6. For any $c' \leqslant c$, there are $b_{c'}, n_{c'} \in \mathbb{N}$ such that APC_1 proves: for every $n_{c'} \leqslant n, \epsilon^{-1} \in Log$ and c'-DNF $E(x_1, \ldots, x_n)$,

$$\Pr_{\rho_2}\Big[E\lceil\rho_2 \text{ depends on } > b_{c'} \text{ variables}\Big] \preccurlyeq_\epsilon 2/n^{3k}.$$

Here we understand as above that $\rho_2 \sim R^n_{1/n^{1/4}}$. However, we shall prove a slightly stronger claim, namely we claim APC_1 proves the statement for all $p \geqslant n^{1/4}$ and letting

$$\rho_2 \sim R^n_{1/p}.$$

This stronger claim helps a proof by induction on c'. The claim being trivial for $c' = 0$ we assume that $c' > 0$ and the claim holds for $(c'-1)$-DNFs.

We want to show that it holds for c'-DNFs. Let S be a sequence of conjunctions, namely E-disjuncts, with disjoint variables which is maximal in the sense that adding any other disjunct to S would break the disjointness property (we are not asking for a maximum length such sequence since finding one could be hard for APC_1).

We set $d' := 4^{c'+1}4k$ and distinguish two cases. In case S contains $\geqslant d'\log n$ conjunctions,

$$\begin{aligned}\Pr_{\rho_2}\Big[\rho_2 \text{ does not satisfy any conjunction in } S\Big] \\ \preccurlyeq_0 \Big(1 - \Big(\frac{1-1/p}{2}\Big)^{c'+1}\Big)^{d'\log n} \leqslant 2^{-d'\log n/4^{c'+1}} \leqslant 1/n^{3k}. \qquad (14)\end{aligned}$$

As said, we argue for $\preccurlyeq_0$ later. To see the first $\leqslant$ note $\frac{1-1/p}{2} \geqslant 1/4$ if n and hence p is large enough, and apply Proposition 2.5 (with $n := 1$ and $m := 4^{c'+1}$). The second $\leqslant$ is trivial with the choice of $d' = 4^{c'+1}4k$ instead of $d' = 4^{c'+1}3k$ again taking care of rounding.

In case S contains $< d' \log n$ conjunctions, then

$$\Pr_{\rho_2}\Big[\rho_2 \text{ leaves } > 15k \text{ variables in } S \text{ unassigned}\Big]$$
$$\preccurlyeq_0 (1/p)^{15k+1} \cdot \binom{c'd' \log n}{15k+1} \leqslant \frac{1}{n^{3k}}. \tag{15}$$

We set

$$b_{c'} := 15k + 2^{15k} \cdot b_{c'-1}$$

and aim to show

$$\Pr_{\rho_2}\big[\rho_2 \in A\big] \preccurlyeq_\epsilon 2/n^{3k}, \tag{16}$$

where A denotes the event that $E{\upharpoonright}\rho_2$ depends on $> b_{c'}$ variables. We do so assuming at various places that n is sufficiently large. It will be clear that $n \geqslant n_{c'}$ for a suitable constant $n_{c'} \in \mathbb{N}$ is sufficient.

Let σ range over assignments to the variables (appearing in the conjuncts) in S. Let ℓ be the number of these variables and note

$$\ell < c' \cdot d' \log n \in \mathit{LogLog}. \tag{17}$$

Observe that $E{\upharpoonright}\sigma$ is a $(c'-1)$-DNF because every E-disjunct shares at least one variable with some conjunction in S. If $(E{\upharpoonright}\sigma)'$ denotes $E{\upharpoonright}\sigma$ with variables renumbered to $x_1, \ldots, x_{n-\ell}$, then the induction hypothesis gives an APC_1-proof of

$$\Pr_{\rho_3}\big[(E{\upharpoonright}\sigma)'{\upharpoonright}\rho_3 \text{ depends on } > b_{c'-1} \text{ variables}\big] \preccurlyeq_\epsilon 2/(n-\ell)^{3k}, \tag{18}$$

where $\rho_3 \sim R^{n-\ell}_{1/p}$, i.e., $\rho_3 < (2p)^{n-\ell}$; the induction hypothesis applies because $n-\ell \geqslant n_{c'-1}$ if n is sufficiently large (by (17)).

Let X range over size $\leqslant 15k$ subsets of the variables in S, and $a < 2^{15k}$. Let B be the complement of the event in (15), and let B_X denote the event that X equals the set of variables in S left unassigned by ρ_2. For $\rho_2 \in B_X$ let $\sigma(\rho_2, a)$ be the assignment to the variables in S that assigns the variables in X according to the bits of a and all other variables in S as ρ_2.

If $\rho_2 \in A$, then $\rho_2 \notin B$ or there exist X, a, σ such that $\rho_2 \in C_{X,a,\sigma}$ where $C_{X,a,\sigma}$ is the event that $\rho_2 \in B_X$ and $\sigma(\rho_2, a) = \sigma$ and $(E{\upharpoonright}\sigma){\upharpoonright}\rho_2$ depends on $> b_{c'-1}$ variables. If σ disagrees with a on X, then there are no restrictions ρ_2 with $\rho_2 \in B_X$ and $\sigma(\rho_2, a) = \sigma$ and $C_{X,a,\sigma}$ is empty. Otherwise these restrictions are precisely those $\rho_2 \sim R^n_{1/p}$ that take certain prescribed values on the variables in S. If we renumber the variables outside S as above by $x_1, \ldots, x_{n-\ell}$, then there is an obvious bijection from these restrictions onto the restrictions $\rho_3 \sim R^{n-\ell}_{1/p}$. Moreover, if this bijection maps ρ_2 to ρ_3, then: $(E\upharpoonright\sigma)\upharpoonright\rho_2$ depends on $> b_{c'-1}$ variables if and only if $(E\upharpoonright\sigma)'\upharpoonright\rho_3$ depends on $> b_{c'-1}$ variables.

Hence, (18) implies

$$\Pr_{\rho_2}\left[\rho_2 \in C_{X,a,\sigma}\right] \preccurlyeq_{\epsilon\cdot\delta} \frac{2}{(n-\ell)^{3k}} \cdot \delta,$$
$$\text{where } \delta := \frac{(2p)^{n-\ell}}{(2p)^{n}} = (2p)^{-\ell}.$$

The number of triples (X, a, σ) is $\leqslant 2^{2\ell+15k}$. Lemma 3.2 (ii) implies that the union of the sets $C_{X,a,\sigma}$ has probability

$$\preccurlyeq_{2\cdot 2^{2\ell+15k}\cdot\epsilon\cdot\delta+\epsilon/2} \quad 2^{2\ell+15k} \cdot \frac{2}{(n-\ell)^{3k}} \cdot \delta.$$

For sufficiently large n we have $n-\ell \geqslant n/2$ (by (17)), and $\delta^{-1} \geqslant (2n^{1/4})^{\ell} \geqslant 2\cdot 2^{2\ell+15k}\cdot 2^{3k}$. Then the error above is $\leqslant \epsilon$ and the bound is $\leqslant 1/n^{3k}$. Using this and (15), the first statement of Lemma 3.2 (ii) yields (16).

It remains to describe circuits witnessing the estimations (12)-(15).

(12) We map every

$$z < n^k \cdot \left(1 - \frac{1-1/n^{1/2}}{2}\right)^{d\log n} \cdot (2n^{1/2})^n = n^k \cdot (n^{1/2}+1)^{d\log n} \cdot (2n^{1/2})^{n-d\log n}$$

to some $\rho_1 < (2n^{1/2})^n$ in such a way that every ρ_1 which does not falsify all size $\geqslant d\log n$ conjunctions in D_n is in the image of the mapping. A given such z determines a triple (s, p, r) with

$$\begin{array}{rcll} s & < & n^k, & \\ p & = & \sum_{i<d\log n} \epsilon_i \cdot (n^{1/2}+1)^i & \text{with } \epsilon_i < n^{1/2}+1, \\ r & = & \sum_{i<n-d\log n} r_i \cdot (2n^{1/2})^i & \text{with } r_i < 2n^{1/2}. \end{array}$$

Output the restriction ρ_1 that assigns the first $d\log n$ variables in the s-th disjunct of D_n according to $\epsilon_0, \ldots, \epsilon_{d\log n-1}$ so that the disjunct is not falsified and the rest according to $r_0, \ldots, r_{n-d\log n-1}$.

(13) A given $z < n^{k-c/2}2^{d\log n}(2n^{1/2})^n$ determines (s, t, p, r) with $s < n^k, t < 2^c, p < 2^{d\log n}$ and $r < (2n^{1/2})^{n-c}$. Output the restriction ρ_1 that assigns, for the maximal c_0 possible, the first $c_0 \leqslant c$ variables in the s-th disjunct of D_n on the positions specified by p according to t (these variables are left unassigned by ρ_1), and the rest of variables according to r together with the unused part of t.

(14) Let T be a conjunction of literals in $t \leqslant c'$ variables, and let $\rho_3 \sim R^t_{1/p}$. The probability that such a ρ_3 satisfies T is $\approx_0 \left(\frac{1-1/p}{2}\right)^t \geqslant \left(\frac{1-1/p}{2}\right)^{c'}$. By Lemma 3.2 (iii) and (i)

$$\Pr_{\rho_3}\left[\rho_3 \text{ does not satisfy } T\right] \preccurlyeq_0 1 - \left(\frac{1-1/p}{2}\right)^{c'+1}.$$

Let C_T be a circuit witnessing this inequality, that is, for some $v_T > 0$

$$C_T : v_T \cdot \left(1 - \left(\frac{1-1/p}{2}\right)^{c'+1}\right) \cdot (2p)^t \twoheadrightarrow v_T \times \left\{\rho_3 \mid \rho_3 \text{ does not satisfy } T\right\}.$$

Note that there are only standard finitely many conjunctions T over a fixed set of c' many variables. We can thus assume without loss of generality that all v_T equal some $v > 0$, e.g. we can replace each v_T by the product $v := \prod_T v_T$ and accordingly modify C_T. We shall assume for simplicity of the following exposition that $v = 1$; it is straightforward to adapt the construction to the general case $v > 0$.

To prove (14) we have to map numbers

$$z < \left(1 - \left(\frac{1-1/p}{2}\right)^{c'+1}\right)^{d' \log n} \cdot (2p)^n$$

to $\rho_2 \sim R^n_{1/p}$ such that all restrictions that do not satisfy any conjunction in S are hit. Assume for notational simplicity that S contains exactly $d' \log n$ conjunctions and let j range over numbers between 1 and $d' \log n$. View a given z as a pair of a sequence $(z_j)_j$ and r where

$$\begin{aligned} z_j &< \left(1 - \left(\frac{1-1/p}{2}\right)^{c'+1}\right) \cdot (2p)^{t_j}, \\ r &< (2p)^{n - \sum_j t_j}. \end{aligned}$$

Here, t_j is the number of variables appearing in the j-th disjunct in S. Output ρ_2 which sets the variables not occurring in S according to r; to set a variable occurring in S, say, the i-th variable in the j-th conjunction of S (hence $1 \leqslant i \leqslant t_j \leqslant c'$), first choose a conjunction T from the finite list of conjunctions considered above such that the j-th conjunct is a suitable variable substitution of T; then assign the given variable as the restriction $C_T(z_j)$ assigns its i-th variable.

(15) Given z coding the triple (s, t, r) with $s < 2^{15k+1}$, $t < \binom{c'd' \log n}{15k+1}$ and $r < (2p)^{n-15k-1}$, output the restriction ρ_2 assigning, for the maximal c_0 possible, the first $c_0 \leqslant 15k+1$ variables in S specified by the t-th $(15k+1)$-size subset of $c'd' \log n$ according to s (these variables are left unassigned) and the rest according to r together with the unused part of s. □

Theorem 3.7. *For all $k, d \in \mathbb{N}$ there is $n_0 \in \mathbb{N}$ such that APC_1 proves: for all $n_0 < n \in Log$ and every AC^0_d-circuit C_n of size n^k with n variables there is $y < 2^n$ such that $C_n(y) \neq \sum_{i=1}^{n} bit(i-1, y) \mod 2$.*

Proof. There is a PV-function transforming any n^k-size circuit C_n of depth d into an equivalent C'_n circuit of size $n^k + n - 1 \leqslant n^{2k}$, depth d and with negations appearing only at the variables. The equivalence is proven in PV_1 for each fixed assignment by Σ^b_0-induction on the number of gates in C_n.

By Lemma 3.5 there is a (standard) $b \in \mathbb{N}$ such that for any DNF or CNF C at the bottom level of C'_n we have that $C\lceil\rho_1\rho_2$ depends on $> b$ variables with probability $\preccurlyeq_\epsilon 1/n^{4k}$; here, $\rho_1 \sim R^n_{1/n^{1/2}}$ and $\rho_2 \sim R^n_{1/n^{1/4}}$ and ϵ is chosen 'small enough' with inverse in Log. By Lemma 3.2 (ii), this event happens for *some* bottom level DNF or CNF only with probability $\preccurlyeq_{(2n^{2k}+1)\epsilon} 1/n^{2k}$.

We further claim, understanding $\rho_1 \sim R^n_{1/n^{1/2}}$ and $\rho_2 \sim R^n_{1/n^{1/4}}$,

$$\Pr_{\rho_1,\rho_2}\Big[\text{there are } < n^{1/8} \text{ variables left unassigned by both } \rho_1 \text{ and } \rho_2\Big]$$
$$\preccurlyeq_0 n^{n^{1/8}} \cdot \left(1 - \frac{1}{n^{3/4}}\right)^{n-n^{1/8}} \leqslant n^{n^{1/8}} \cdot 2^{\frac{-(n-n^{1/8})}{n^{3/4}}} \leqslant n^{n^{1/8}} \cdot 2^{1-n^{1/4}} \leqslant 1/n^{2k}.$$

The first $\leqslant$ uses Proposition 2.5. To witness $\preccurlyeq_0$ we map

$$z < n^{n^{1/8}} \cdot \left(1 - \frac{1}{n^{3/4}}\right)^{n-n^{1/8}} \cdot (2n^{1/2})^n \cdot (2n^{1/4})^n = n^{n^{1/8}} \cdot (4n^{3/4} - 4)^{n-n^{1/8}} \cdot (4n^{3/4})^{n^{1/8}}$$

coding (s, p, r) with $s = \sum_{i<n^{1/8}} s_i n^i$, $s_i < n$, and $p < (4n^{3/4} - 4)^{n-n^{1/8}}$ and $r < (4n^{3/4})^{n^{1/8}}$ to the following pair $\langle\rho_1, \rho_2\rangle$ of restrictions: the variables $x_{s_i+1}, i < n^{1/8}$, are set according to r (in particular, these variables might be left unassigned by ρ_1, ρ_2); the number p can be used to determine the value pair of ρ_1 and ρ_2 on every other variable such that not both are 'unassigned'.

By Lemma 3.2 (ii), (iii), with probability $\succcurlyeq_{(2n^{2k}+2)\epsilon} 1 - 2/n^{2k}$ we have that ρ_1, ρ_2 leave at least $n^{1/8}$ variables unassigned and simplify all CNFs and DNFs at the bottom: all these CNFs and DNFs do not depend on $> b$ variables, and thus are (PV_1-provably) equivalent to both CNFs and DNFs of size $\leqslant (b+1)2^b + 1$. For ϵ chosen small enough, Proposition 2.10 (ii) implies that such restrictions ρ_1, ρ_2 exist.

In case $d = 2$ we get a contradiction assuming n is large enough so that $n^{1/8} > b$: if C'_n computed parity, then it depends on all its variables.

In case $d > 2$, the circuit $C'_n\lceil\rho_1\rho_2$ is equivalent to a circuit with $\geqslant n^{1/8}$ variables, depth $d-1$ and size $\leqslant ((b+1)2^b+1)n^{2k}$. If C'_n computed parity on 2^n then from $C'_n\lceil\rho_1\rho_2$ we get a circuit $C_{n'}$ computing parity or its negation on $2^{n'}, n' := \lceil n^{1/8}\rceil$. This circuit has depth $d-1$ and size $(n')^{k'}$ for suitably large k'. Arguing by induction on $d \geqslant 2$, we can assume to have already refuted the existence of such a circuit. □

Remark 3.8. We point out which steps in the proof presented rely on $\mathsf{sWPHP}(\mathsf{PV})$. In the proof of Lemma 3.5 it is the use of Lemma 3.2 (ii) in the proof of Claim 3.6 and Lemma 3.2 (iii) in the verification of (14). Theorem 3.7 uses the union bound Lemma 3.2 (ii) to bound the probability that all bottom level DNFs simplify. Note that the frequent uses of the first statement of this lemma do not require $\mathsf{sWPHP}(\mathsf{PV})$. Lemma 3.2 (iii) is used to argue that restrictions are good with probability $\succcurlyeq_{(2n^{2k}+2)\epsilon} 1 - 2/n^{2k}$, and then Proposition 2.10 (ii) is used to infer that good restrictions exist.

3.3 Razborov and Smolensky's lower bound for $\mathsf{AC}^0[p]$ circuits

Let $d, p \in \mathbb{N}, p > 0$. An $\mathsf{AC}^0_d[p]$-*circuit* is defined like an AC^0_d-circuit but we additionally allow unbounded fan-in gates labeled MOD_p; such a gate returns 1 or 0 depending on whether it receives a number of ones divisible by p or not. Recall that, by the *size* of a circuit we mean the number of its inner gates.

In a first step (Theorem 3.9), for prime p, we want to approximate a given $\mathsf{AC}^0_d[p]$ circuit by a low degree polynomial over the finite field $\mathbb{F}_p$. Unfortunately, the sequence of coefficients coding such a polynomial can be infeasible. For this reason, we represent polynomials by *arithmetical* $\mathbb{F}_p$-*circuits*: these have unbounded fan-in multiplication and addition gates labeled $\times$ and $+$ and input gates labeled by variables or constants from $\mathbb{F}_p$. Instead of the degree of the polynomial computed we use an easily computable upper bound: the *syntactic degree* of an arithmetical $\mathbb{F}_p$-circuit (with one output) is the number it computes (in the obvious sense) when we replace $\mathbb{F}_p$-constants by 0, variables by 1, $+$ by max, and $\times$ by $+$.

Recall that the sharply bounded collection scheme $\mathsf{BB}(\Pi^b_1)$ contains

$$\forall i{\leqslant}|x|\ \exists y{\leqslant}z\ \varphi(i, y, \bar{x}) \to \exists w\ \forall i{\leqslant}|x|\ \varphi(i, (w)_i, \bar{x})$$

for all $\varphi \in \Pi^b_1$; here, $(w)_i$ is some standard sequence coding (see [34, Section 5.4]).

Theorem 3.9 (Low-degree approximation)**.** *For all $d, p \in \mathbb{N}$ with p prime the theory*

$$\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV}) + \mathsf{BB}(\Pi^b_1)$$

proves: for $\ell \in LogLog$ and $n, s, \epsilon^{-1} \in Log$ and every $\mathsf{AC}^0_d[p]$-circuit C of size $\leqslant s$ with n variables, there is an arithmetical $\mathbb{F}_p$-circuit P of syntactic degree $\leqslant ((p-1)\ell)^d$ such that

$$\Pr_{x<2^n}[P(x) \neq C(x)] \preccurlyeq_0 s/2^\ell + \epsilon.$$

Proof. For a gate g of C let C_g be the subcircuit with output gate g. We prove in APC_1:

Claim 3.10. Let g be an inner gate of C and let $g_1, \ldots, g_m$ list the gates wired into g. Then there exists an arithmetical $\mathbb{F}_p$-circuit P_g with variables $X_1, \ldots, X_m$ and syntactic degree $\leqslant (p-1)\ell$ such that

$$\begin{gathered}\Pr_{x<2^n}\big[x \in \mathit{Error}_g\big] \preccurlyeq_0 1/2^\ell + \epsilon \\ \text{where } \mathit{Error}_g := \big\{x < 2^n \mid P_g(C_{g_1}(x), \ldots, C_{g_m}(x)) \neq C_g(x)\big\}.\end{gathered} \tag{19}$$

If g is labeled MOD_p, then set $P_g := 1 - \big(\sum_{i<m} X_i\big)^{p-1}$, and, if g is labeled $\neg$ (and $m = 1$), then set $P_g := 1 - X_1$. Note $\mathit{Error}_g = \emptyset$ in both cases. The $\bigwedge$-case being dual, the case that g is labeled $\bigvee$ is the only interesting one.

Observe first that $\Pr_{S \subseteq m}\big[\sum_{i\in S} y_i = 0 \mod p\big] \preccurlyeq_0 1/2$ for every fixed $0 < y < 2^m$, where we write $y_i := \mathit{bit}(i, y)$. This implies

$$\begin{gathered}\Pr_{\substack{x<2^n \\ S_0,\ldots,S_{\ell-1}\subseteq m}} [C_g(x) \neq P_{\vec{S}}(C_{g_1}(x), \ldots, C_{g_m}(x))] \preccurlyeq_0 1/2^\ell, \\ \text{where } P_{\vec{S}} := 1 - \textstyle\prod_{i<\ell}\big(1 - \big(\sum_{j\in S_i} X_j\big)^{p-1}\big).\end{gathered}$$

A formally precise notation would replace the index $S_0, \ldots, S_{\ell-1} \subseteq m$ by $s < 2^{m\cdot\ell}$ and S_i, in the event description, should be a suitable PV-term $t(s, i)$. By Lemma 3.2 (iv) we can fix $S_0, \ldots, S_{\ell-1} \subseteq m$ such that (19) holds with $P_g := P_{\vec{S}}$. This proves the claim.

We intend to define P by replacing every inner gate g of C by P_g. To do so we need the sequence $(P_g)_g$ where g ranges over the inner gates of C. It is not obvious that this sequence exists because (19) is an unbounded $\exists\Pi_2$-formula. Theorem 2.9 allows to bring the quantifier complexity down to Π_1^b as follows.

First choose s_g such that $s_g \approx_\epsilon \mathit{Error}_g$ and by Claim 3.10 and Proposition 2.10 (ii)

$$s_g \leqslant (1/2^\ell + 3\epsilon) \cdot 2^n. \tag{20}$$

Theorem 2.9 additionally gives a number v_g and circuits G_g, H_g such that

$$\begin{aligned}&\forall z < v_g \cdot (s_g + \epsilon \cdot 2^n)\ \big(G_g(z) \in v_g \times \mathit{Error}_g\big) \\ &\wedge\ \forall z \in v_g \times \mathit{Error}_g\ \big(H_g(z) < v_g \cdot (s_g + \epsilon \cdot 2^n) \wedge G_g(H_g(z)) = z\big).\end{aligned} \tag{21}$$

Thus, APC_1 proves that for every g there exists a (code of a) tuple $\langle P_g, s_g, v_g, G_g, H_g\rangle$ such that (20) and (21) hold. By Parikh's theorem [51] (see [9, Theorem 1.2.7.1]) the code of such a tuple can be bounded by a suitable term $t(C)$. Now, Π_1^b-collection gives a (code of a) sequence $(\langle P_g, s_g, v_g, G_g, H_g\rangle)_g$ such that (20) and (21) and hence also

$$\Pr_{x<2^n}[x \in \mathit{Error}_g] \preccurlyeq_0 1/2^\ell + 4\epsilon.$$

hold for all g. Given this sequence define P by replacing each inner gate g of C by P_g. By induction, P has syntactical degree $\leqslant ((p-1)\ell)^d$. Also by induction one sees that if

$P(x) \neq C(x)$ then there exists g (which is 'first' such that the computations differ and hence) such that $x \in \mathit{Error}_g$. Applying Lemma 3.2 (ii) we conclude $\Pr_{x<2^n}[P(x) \neq C(x)]$ is $\preccurlyeq_\epsilon s \cdot (1/2^\ell + 4\epsilon)$, so $\preccurlyeq_0 s \cdot (1/2^\ell + 4\epsilon) + 2\epsilon$ by Lemma 3.2 (i). As ϵ was arbitrary with inverse in Log and $s \in \mathit{Log}$, the theorem follows. □

Remark 3.11. The above theorem holds true more generally for $p \in \mathit{Log}$ instead of only for standard primes $p \in \mathbb{N}$. Jeřábek [29, Section 4.3] formalized basic properties of finite fields in bounded arithmetic, and shows in particular, that, for $p \in \mathit{Log}$ prime, PV_1 can construct $\mathbb{F}_p$ and prove $a^{p-1} = 1$ for $a \in \mathbb{F}_p \setminus \{0\}$ [29, Lemma 4.3.11].

To derive an $\mathsf{AC}^0[p]$ lower bound, one usually proceeds further by showing that any polynomial computing MOD_q with high probability must have degree $\Omega(n^{1/2})$. The simplest proof of this compares the number of all functions on n variables to the number of low-degree polynomials. As this argument is infeasible, we reproduce it on functions with only $\log^{O(1)} n$ arguments. This results in a weaker degree lower bound which, however, still suffices to derive an $\mathsf{AC}^0[p]$ lower bound.

Theorem 3.12 (Degree lower bound)**.** *For any $d \in \mathbb{N}$ and primes $p \neq q$, there is $n_0 \in \mathbb{N}$ such that APC_1 proves: if $n_0 < 2^{\log^{3d} n} \in \mathit{Log}$ and $n_0 < \epsilon^{-1} \in \mathit{Log}$, then every arithmetical $\mathbb{F}_p$-circuit P with n variables such that*

$$\Pr_{x<2^n}[P(x) \neq \mathit{MOD}_q(x_1, \ldots, x_n)] \preccurlyeq_\epsilon \frac{1}{5q2^q}$$

has syntactic degree bigger than $\log^d n$; here, $x_i := \mathit{bit}(i-1, x)$ for all $1 \leqslant i \leqslant n$.

Proof. Assume for contradiction that P is an arithmetical $\mathbb{F}_p$-circuit of syntactic degree $\leqslant \log^d n$ which differs from MOD_q with probability $\preccurlyeq_\epsilon 1/(5q2^q)$.

For $i < q$, let MOD_q^i be the Boolean function that outputs 1 if the number of ones in its input string equals i modulo q, and outputs 0 otherwise. Let $P_i(x_1, \ldots, x_{n-q})$ be obtained from $P(x_1, \ldots, x_n)$ by substituting i many zeros and $q - i$ many ones for the variables $x_{n-q+1}, \ldots, x_n$. We claim that for every $i < q$

$$\begin{aligned} &\Pr_{x<2^{n-q}}\big[x \in X_i\big] \preccurlyeq_\epsilon 1/(4q). \\ &\text{where } X_i := \big\{x < 2^{n-q} \mid P_i(x) \neq \mathit{MOD}_q^i(x)\big\}. \end{aligned} \tag{22}$$

Otherwise the probability is $\succcurlyeq_\epsilon 1/(4q)$ by Proposition 2.10 (i), that is, there are $v > 0$ and a circuit C that computes a surjection from $v \times (X_i \cup \epsilon 2^{n-q})$ onto $v \cdot 2^{n-q}/(4q)$. Note each $x \in X_i$ can be "prolonged" by q bits (as in the definition of P_i) to an element y_x of Y, the set of $y < 2^n$ with $\mathit{MOD}_q(y) \neq P(y)$. Mapping such y_x to $C(x)$ and all other y to 0, defines a surjection from $v \times (Y \cup \epsilon 2^n)$ (where we increase 2^{n-q} to 2^n) onto

$v \cdot 2^{n-q}/(4q) = v \cdot 2^n/(4q2^q)$. This implies $Y \succcurlyeq_\epsilon 2^n/(4q2^q)$. Since also $Y \preccurlyeq_\epsilon 2^n/(5q2^q)$, Proposition 2.10 (ii) gives a contradiction for small enough ϵ. This proves (22).

We now consider P and the P_i's as circuits over $\mathbb{F}_{p^{q-1}}$. This (constant size) field contains a q-th root of unity $\omega \neq 1$.

Using the substitution $y = \frac{x-1}{\omega-1}$ (which maps $\omega \mapsto 1$ and $1 \mapsto 0$) we get from every P_i, $i < q$, an arithmetical $\mathbb{F}_{p^{q-1}}$-circuit $P'_i(x_1, \dots, x_{n-q})$ of syntactic degree $\leqslant \log^d n$ such that $P'_i(x) = 1$ if $\prod_{j=1}^{n-q} x_j = \omega^i$ and $P'_i(x) = 0$ otherwise, for all except $\preccurlyeq_\epsilon 1/(4q)$ many $x \in \{\omega, 1\}^{n-q}$. More precisely, $x \in \{\omega, 1\}^{n-q}$ should read $y < 2^{n-q}$ where such y codes the tuple x, and x_j abbreviates a PV-term denoting its j-th component.

The circuit

$$P'(x_1, \dots, x_{n-q}) := \textstyle\sum_{i<q} P'_i \cdot \omega^i$$

then has syntactic degree $\leqslant \log^d n$ and satisfies

$$\Pr_{x \in \{\omega,1\}^{n-q}} \left[P'(x) \neq \textstyle\prod_{i=1}^{n-q} x_i\right] \preccurlyeq_{q\epsilon} 1/4,$$

by $(q-1)$ applications of the first statement of Lemma 3.2 (ii).

Let

$$m := \log^{3d} n.$$

Rewrite the above event as a set of pairs $\langle x, a\rangle \in \{\omega, 1\}^m \times \{\omega, 1\}^{n-q-m}$ and apply Lemma 3.2 (iv) to fix $a \in \{\omega, 1\}^{n-q-m}$ such that

$$\Pr_{x \in \{\omega,1\}^m}[X] \preccurlyeq_{q\epsilon} 1/4 + 9q\epsilon,$$
$$\text{where } X := \left\{x \in \{\omega, 1\}^m \mid P'(x, a) \neq \textstyle\prod_{i=1}^m x_i \prod_{i=1}^{n-q-m} a_i\right\}.$$

By the assumption of Theorem 3.12, $2^{\log^{3d} n} = 2^m \in \mathit{Log}$, so the set X can be counted precisely in PV_1 (cf. Section 2.3). In particular, $\mathit{Card}(X) \leqslant 1/3 \cdot 2^m$ if ϵ is sufficiently small. Define the circuit

$$P''(x) := P'(x, a) \cdot (\textstyle\prod_{i=1}^{n-q-m} a_i)^{-1}.$$

Now, consider an arbitrary function $f : \{\omega, 1\}^m \to \mathbb{F}_{p^{q-1}}$. For $c, b \in \{1, \omega\}$ observe

$$\frac{2cb - (1+\omega)(c+b) + 1 + \omega^2}{(1-\omega)^2} = \begin{cases} 1 & \text{if } c = b \\ 0 & \text{else.} \end{cases}$$

We can thus express f as

$$f(x) = \sum_{b \in \{\omega,1\}^m} f(b) \cdot \prod_{i=1}^m \frac{2x_ib_i - (1+\omega)(x_i+b_i) + 1 + \omega^2}{(1-\omega)^2} = \sum_{b \in \{\omega,1\}^m} f(b) \cdot \prod_{i=1}^m \frac{x_it_{i,1} + t_{i,2}}{(1-\omega)^2}$$

where $t_{i,1} := 2b_i - (1+\omega)$ and $t_{i,2} := -(1+\omega)b_i + 1 + \omega^2$. For $x \notin X$ we know $P''(x) = \prod_{i=1}^m x_i$, and thus can write

$$\prod_{i=1}^m (x_i t_{i,1} + t_{i,2}) = \sum_{\substack{T\subseteq[m]\\ |T|\leqslant m/2}} \prod_{i\in T} x_i t_{i,1} \prod_{i\in[m]\setminus T} t_{i,2} + P''(x)\cdot \sum_{\substack{T\subseteq[m]\\ |T|>m/2}} \prod_{i\in T} t_{i,1} \prod_{i\in[m]\setminus T} t_{i,2} x_i^{q-1},$$

where we use that $x_i^q = 1$. Since

$$x_i^{q-1} = \sum_{z\in\{\omega,1\}} z^{q-1} \frac{2x_i z - (1+\omega)(x_i+z) + 1 + \omega^2}{(1-\omega)^2},$$

we conclude that, outside X, the function f is computed by a circuit of syntactic degree $\lfloor \frac{m}{2} \rfloor + m^{1/3} + 1$. Note that the circuit $P''(x)$ can be expanded to the sum of $\leqslant 2^m \in$ *Log* monomials so the polynomial representing f can be coded by the sequence of its coefficients. By Proposition 2.4, the number of such polynomials is

$$\preccurlyeq_0 (p^{q-1})^{\sum_{i=0}^{\lfloor m/2+m^{1/3}\rfloor+1}\binom{m}{i}} < (p^{q-1})^{(5/9)2^m}$$

while the number of all functions $f : \{1,\omega\}^m \setminus X \to \mathbb{F}_{p^{q-1}}$ is $\succcurlyeq_0 (p^{q-1})^{(2/3)2^m}$. This contradicts Proposition 2.10 (ii). □

For later use we note that the proof above shows the following.

Remark 3.13. Every function $f : \{\omega,1\}^m \to \mathbb{F}_{p^{q-1}}$ is a linear combination of polynomials of the form $(\prod_{i=1}^m x_i)\cdot Q$ or Q or 1, where Q ranges over low degree multilinear monomials, i.e., products of at most $m/2$ variables.

Theorem 3.14. *For any $d \in \mathbb{N}$ and primes $p \neq q$, there is $n_0 \in \mathbb{N}$ such that* APC_1 *proves: if $n_0 < 2^{\log^{9d} n} \in Log$, then for every size $\leqslant n^{\log n}$ $\mathsf{AC}^0_d[p]$-circuit C with n variables there is $x < 2^n$ such that $C(x) \neq MOD_q(x_1,\ldots,x_n)$; here, $x_i := bit(i-1,x)$ for all $1 \leqslant i \leqslant n$.*

Proof. It suffices to give the proof in the theory of Theorem 3.9. Indeed, by [28, Corollary 4.12] this theory is Σ^b_2-conservative over $\mathsf{S}^1_2 + \mathsf{sWPHP(PV)}$ which in turn is Σ^b_1-conservative over APC_1 by Theorem 2.7. In particular, we are free to use Theorem 3.9. We apply this theorem to a given $\mathsf{AC}^0_d[p]$-circuit C of size $s \in Log$ with $\epsilon := 1/(10q2^q)$ and $\ell := \lceil \log(10q2^q s)\rceil \in LogLog$. This yields an arithmetical $\mathbb{F}_p$-circuit P of syntactic degree $\leqslant (\lceil \log(10q2^q s)\rceil\,(p-1))^d$ such that

$$\Pr_{x<2^n}[P(x) \neq C(x)] \preccurlyeq_0 1/(5q2^q).$$

If C computes MOD_q, then $(\lceil \log(10q2^q s)\rceil\,(p-1))^d \geqslant \log^{3d} n$ by Theorem 3.12, and hence $s > n^{\log n}$ as claimed. □

Remark 3.15. We point out which steps in the proof presented rely on $\mathsf{sWPHP}(\mathsf{PV})$. The proof of Theorem 3.9 heavily relies on the $\mathsf{sWPHP}(\mathsf{PV})$, namely first in the averaging argument Lemma 3.2 (iv) in the proof of Claim 3.10, then in the use of Theorem 2.9 preparing the application of the collection scheme, and then in the final union bound Lemma 3.2 (ii). In the proof of Theorem 3.12 we have Proposition 2.10 (i) and (ii) in the construction of polynomials P_i and the averaging argument Lemma 3.2 (iv) in the construction of the polynomial P'' approximating the iterated product. The final contradiction relies also on Proposition 2.10 (ii).

3.4 Razborov's lower bound for monotone circuits

We view numbers $G < 2^{\binom{n}{2}}$ as graphs on $[0, n)$ in the natural way. By a monotone circuit we mean a circuit without $\neg$-gates and all inner gates of fan-in 2. If it has $\binom{n}{2}$ variables we write them as $x_{\{i,j\}}$ for $i, j < n, i \neq j$, indicating presence of an edge between i and j in an input graph G.

Theorem 3.16. *There are $\epsilon > 0$ and $n_0 \in \mathbb{N}$ such that APC_1 proves: for all $n > n_0$ and $2 \leqslant k \leqslant n^{1/4}$ such that $n^k \in Log$, no monotone circuit of size $n^{\epsilon\sqrt{k}}$ with $\binom{n}{2}$ variables accepts exactly the n-vertex graphs containing a clique of size k.*

Proof. We follow the presentation in [5, Section 4.2] (cf. also [3]). Let C be a monotone circuit with $\binom{n}{2}$ variables and size s and set

$$\ell := \sqrt{k}, \quad p := \ell \cdot \lceil \log n \rceil, \quad m := (p-1)^\ell \cdot \ell!. \tag{23}$$

Observe that all these numbers are in *Log*. For $\tilde{m} \in Log$ we naturally code length $\tilde{m}$ sequences $\vec{X} = \langle X_0, \ldots, X_{\tilde{m}-1} \rangle$ of size $\leqslant \ell$ subsets $X_i \subseteq n$ by a number $< n^{\ell \cdot \tilde{m}}$. In the following we understand that $\vec{X}, \vec{Y}, \ldots$ range over such sequences of different lengths.

We aim to approximate C by an "approximator circuit" $C[\vec{X}] : 2^{\binom{n}{2}} \to 2$ where $\vec{X}$ has length $< m$: $C[\vec{X}]$ maps $G < 2^{\binom{n}{2}}$ to 1 or 0 depending on whether there is $i < m$ such that G restricted to X_i is a clique. The approximation is measured with respect to "test graphs": the "positive" ones are the graphs P_i, for $i < \binom{n}{k}$, containing a clique on the i-th size k subset of n and no other edges; the "negative" ones are the graphs N_c, for $c < (k-1)^n$, having an edge between j and j' if and only if $c_j \neq c_{j'}$ where we write $c = \sum_{i<n} c_i (k-1)^i$ with $c_i < k - 1$.

Claim 3.17 (Sunflower lemma)**.** If $\vec{X}$, say, of length $\tilde{m}$ contains $\geqslant m$ distinct sets, then it contains a *sunflower*, i.e. a set $F \subseteq \tilde{m}$ of p pairwise distinct indices such that for some *center* $X \subseteq n$ we have $X_j \neq X_j \cap X_{j'} = X$ for all $j, j' \in F, j \neq j'$.

The usual proof (e.g. [5, Lemma 4.1]) formalizes without change in PV_1 because all sets appearing in it have bounds in *Log*, so PV_1 can count them precisely (recall Section 2.3).

There is a function *plucking* $\in \mathsf{PV}$ which provably in PV_1 maps $\vec{X}$ to itself if it contains $< m$ many pairwise distinct sets, and otherwise to a sequence

$$\langle \langle F^1, \vec{X}^1 \rangle, \ldots, \langle F^u, \vec{X}^u \rangle \rangle$$

for some $u \geqslant 1$ such that we have for all $1 \leqslant i < u$:

- $\vec{X}^i$ contains at least m pairwise distinct sets,
- F^i is a sunflower in $\vec{X}^{i-1}$ (we understand $\vec{X}^0 := \vec{X}$), say, with center X,
- $\vec{X}^i$ is obtained from $\vec{X}^{i-1}$ by replacing entries X^{i-1}_j with $j \in F^i$ by X,
- $\vec{X}^u$ contains $< m$ many pairwise distinct sets.

The function *plucked* takes $\vec{X}$ to $\vec{Z}$ obtained from $\vec{X}^u$ above by deleting repetitions, i.e. deleting any entry equal to an earlier one.

The functions *plucking* and *plucked* take two arguments, denoted $\vec{X}$ and m above. Notationally, we supress m and write e.g. *plucked*$(\vec{X})$.

Given $\vec{X}, \vec{Y}$ of lengths $m', m'' < m$ respectively, we define

$$\vec{X} \sqcup \vec{Y} := \mathit{plucked}(\vec{Z})$$

where $\vec{Z}$ is the concatenation of $\vec{X}$ and $\vec{Y}$, that is, is the length $m' + m''$ sequence with $Z_i = X_i$ for $i < m'$ and $Z_i = Y_i$ for $m' \leqslant i < m''$. Similarly define

$$\vec{X} \sqcap \vec{Y} := \mathit{plucked}(\vec{Z})$$

where $\vec{Z}$ is obtained from $\vec{X} \times \vec{Y}$ by deleting all entries of size $> \ell$ where "size" is *Card* (cf. Section 2.3). The sequence $\vec{X} \times \vec{Y}$ is defined as the length $m' \cdot m'' = m' \times m''$ sequence with $\langle i, j \rangle$-th entry $X_i \cup Y_j$.

The following claim states that $\sqcup, \sqcap$ approximate $\vee, \wedge$ with respect to positive and negative test graphs. Note that positive test graphs form a probability space in *Log*, so events can be counted precisely using *Card*:

Claim 3.18. Let $\vec{X}, \vec{Y}$ have lengths $m', m'' < m$ respectively and let $\gamma^{-1} \in \mathit{Log}$. Then

$$\mathit{Card}\Big(\big\{ x < \tbinom{n}{k} \mid C[\vec{X} \sqcup \vec{Y}](P_x) < (C[\vec{X}] \vee C[\vec{Y}])(P_x) \big\} \Big) / \tbinom{n}{k} \quad = \quad 0 \tag{24}$$

$$\mathit{Card}\Big(\big\{ x < \tbinom{n}{k} \mid C[\vec{X} \sqcap \vec{Y}](P_x) < (C[\vec{X}] \wedge C[\vec{Y}])(P_x) \big\} \Big) / \tbinom{n}{k} \quad \leqslant \quad m^2 \cdot (k/n)^{\ell+1} \tag{25}$$

$$\Pr_{c < (k-1)^n} \Big[C[\vec{X} \sqcup \vec{Y}](N_c) > (C[\vec{X}] \vee C[\vec{Y}])(N_c) \Big] \quad \preccurlyeq_{\gamma} \quad m \cdot 1/2^p \tag{26}$$

$$\Pr_{c < (k-1)^n} \Big[C[\vec{X} \sqcap \vec{Y}](N_c) > (C[\vec{X}] \wedge C[\vec{Y}])(N_c) \Big] \quad \preccurlyeq_{\gamma} \quad m^2 \cdot 1/2^p \tag{27}$$

The event in (24) is empty since $C[\mathit{plucked}(\vec{Z})](G) \geqslant C[\vec{Z}](G)$ for all $\vec{Z}$ and $G < 2^{\binom{n}{2}}$. For the same reason, for every $x < \binom{n}{k}$ in the event in (25) there are $i < m', j < m''$ such that $X_i \cup Y_j$ has size $> \ell$ and is contained in the x-th size k subset of n; for every such i, j this has probability $\leqslant \binom{n-\ell-1}{k-\ell-1}/\binom{n}{k} \leqslant (k/n)^{\ell+1}$ and (25) follows from the union bound.

To see (26) let $\mathit{plucking}(\vec{Z}) = \langle\langle F^1, \vec{Z}^1\rangle, \ldots, \langle F^u, \vec{Z}^u\rangle\rangle$ for $\vec{Z}$ the concatenation of $\vec{X}$ and $\vec{Y}$, and note $u < m$. If $c < (k-1)^n$ is such that $C[\vec{X} \sqcup \vec{Y}](N_c) > (C[\vec{X}] \vee C[\vec{Y}])(N_c)$ then there is $1 \leqslant i \leqslant u$ such that $C[\vec{Z}^{i-1}](N_c) = 0$ and $C[\vec{Z}^i](N_c) = 1$ (again $\vec{Z}^0 := \vec{Z}$). Then c, viewed as a function from n to $k-1$, is injective on the center X of the sunflower F^i but contains a collision on each of the p many *petals* $X_j \setminus X, j \in F^i$. Since the petals are disjoint such collisions happen with probability $\preccurlyeq_0 (\binom{\ell}{2}/(k-1))^p < 1/2^p$. We leave it to the reader to witness $\preccurlyeq_0$ by a circuit: note $(k-1)^\ell \in \mathit{Log}$, so given a petal PV_1 can list all $\leqslant \binom{\ell}{2}/(k-1) \cdot (k-1)^\ell$ many functions with a collision on it. Now (26) follows from Lemma 3.2 (ii).

To see (27) let $\vec{X} \sqcap \vec{Y} = \mathit{plucked}(\vec{Z})$ for $\vec{Z}$ obtained from $\vec{X} \times \vec{Y}$ as described. Observe $C[\vec{Z}](G) \leqslant C[\vec{X} \times \vec{Y}](G)$ for all $G < 2^{\binom{n}{2}}$, so $C[\mathit{plucked}(\vec{Z})](N_c) > C[\vec{Z}](N_c)$ contains the event under consideration. Its probability is estimated as above, now with $u \leqslant m^2$.

This finishes the proof of Claim 3.18.

Claim 3.19. Let $\gamma^{-1} \in \mathit{Log}$. There is a length $< m$ sequence $\vec{X}$ such that

$$\begin{aligned} \mathit{Card}\big(\{x < \tbinom{n}{k} \mid C[\vec{X}](P_x) < C(P_x)\}\big)/\tbinom{n}{k} &\leqslant s \cdot m^2 \cdot (k/n)^{\ell+1}, & (28)\\ \Pr_{c<(k-1)^n}\big[C[\vec{X}](N_c) > C(N_c)\big] &\preccurlyeq_\gamma s \cdot m^2 \cdot 1/2^p. & (29) \end{aligned}$$

To prove the claim, first note that there is a function in PV that maps every gate g of C to a length $< m$ sequence $\vec{X}^g$ such that PV_1 proves:

- If g is labeled with a variable $x_{\{i,j\}}$, then $\vec{X}^g$ is the length 1 sequence $\langle\{i, j\}\rangle$;
- If g is labeled 1 or 0, then $\vec{X}^g$ is $\langle\emptyset\rangle$ or the empty sequence respectively;
- If g is labeled $\vee$ or $\wedge$, then $\vec{X}^g$ is obtained by applying $\sqcup$ or $\sqcap$ to the sequences computed for the gates wired into g.

We verify the claim for $\vec{X} := \vec{X}^g$ for the output gate g of C. To see (28) note for any x in the event there is a first gate g_x of C such that $C[\vec{X}^{g_x}](P_x) = 0$ while in C gate g_x computes 1 on P_x; here we refer to an enumeration of the gates of C such that any gate appears before the gates it is wired into. Since $C[\vec{X}^g]$ agrees with g if g is an input gate, g_x is an inner gate. Thus x is in the event of (24) or (25) with $\vec{X}, \vec{Y}$ denoting the sequences computed for the gates wired into g_x. Hence, (28) follows by a union bound.

For (29) we argue analogously, the final union bound being done by Lemma 3.2 (ii) causing the error $2\gamma s + \gamma$ for approximate counting. As γ is arbitrary with inverse in Log, this implies our claim. The lemma is applied to the the sequence $(E_g)_g$ of error sets where g runs over the gates of C. More precisely, E_g is the event in (26) or (27) for $\vec{X}, \vec{Y}$ the sequences computed for the gates wired into g.

Now assume C has size $s \leqslant n^{\epsilon \cdot \ell}$ and accepts all P_x and rejects all N_c. Choosing $\vec{X}$ according to Claim 3.19 we get a contradiction by distinguishing two cases.

First suppose that $\vec{X}$ is the empty sequence, so $C[\vec{X}]$ is identically 0. Then the event in (28) is trivial so the l.h.s. equals 1. Recalling (23) and the assumption $k \leqslant n^{1/4}$ we have

$$sm^2 < s(\ell p)^{2\ell} = s(k\lceil \log n \rceil)^{2\ell} < n^{(\epsilon+2/3)\ell},$$

and $(k/n)^{\ell+1} < n^{-3\ell/4}$, so the r.h.s in (28) is $< n^{(2/3+\epsilon)\ell - 3\ell/4} < 1$ (for ϵ small enough).

So suppose $\vec{X} = \langle X_1, \ldots \rangle$ is not empty. Then $C[\vec{X}](N_c) = 1$ if c does not have a collision on X_1; denote this event by Y. Then

$$1/2 \cdot (k-1)^n \preccurlyeq_{1/13} Y \preccurlyeq_{1/13} sm^2 \cdot 1/2^p \cdot (k-1)^n \leqslant n^{(\epsilon+2/3)\ell} \cdot n^{-\ell} \cdot (k-1)^n,$$

where the first $\preccurlyeq_{1/13}$ follows from Lemma 3.2 (iii): recall $\mathit{Card}(X_1) \leqslant \ell$ and we already noted that a collision has probability $\preccurlyeq_0 \binom{\ell}{2}/(k-1) \leqslant 1/2$. Proposition 2.10 (ii) gives

$$1/2 \cdot (k-1)^n < n^{(\epsilon-1/3)\ell} \cdot (k-1)^n + 3/13 \cdot 2^{|(k-1)^n|} \leqslant (n^{(\epsilon-1/3)\ell} + 6/13) \cdot (k-1)^n,$$

and this is wrong if ϵ is small enough and n is large enough. □

Remark 3.20. We point out which steps in the proof presented rely on $\mathsf{sWPHP}(\mathsf{PV})$. The proofs of (26), (27) and (29) use the union bound Lemma 3.2 (ii). The final contradiction uses Lemma 3.2 (iii) and Proposition 2.10 (ii).

3.5 Probabilistic witnessing

We find it worthwhile to point out explicitly the following complexity theoretic benefit of succinct circuit lower bound proofs in APC_1. It is a direct application of Wilkie's Witnessing Theorem 2.7.

Proposition 3.21. *Let $k, n_0 \in \mathbb{N}$ and $\mathsf{Q} \in \mathsf{P}$. If APC_1 proves*

$$n_0 \leqslant n \rightarrow \mathsf{LB}[\mathsf{C}, \mathsf{Q}](C, n^k, n, N),$$

then there exists a probabilistic polynomial time Turing machine which given $n \geqslant n_0$ in unary and a C-circuit C of size $\leqslant n^k$, outputs with probability at least $2/3$ some $y < 2^n$ such that C does not decide Q on y, that is, $C(y) = 1, y \notin \mathsf{Q}$ or $C(y) = 0, y \in \mathsf{Q}$.

For example, from Theorem 1.3 we get:

Corollary 3.22. *There is a rational $0 < \epsilon < 1$ such that for all $k \geqslant 2$ there is $n_0 \in \mathbb{N}$ and a probabilistic polynomial time Turing machine which given $n \geqslant n_0$ in unary and a monotone circuit C of size $\leqslant n^{\epsilon\sqrt{k}}$, outputs with probability at least $2/3$ a graph G on n vertices such that C does not decide k-CLIQUE on G.*

In fact, the probabilistic witnessing is definable and provable in PV_1 and APC_1 in appropriate senses. We refer the interested reader to [28, Proposition 1.16].

3.6 Razborov and Rudich's natural proof barrier

The definitions of natural properties and pseudorandom generators both require to count the sizes of certain sets quite precisely, namely up to certain inverse polynomial factors. Formalizing these concepts in APC_1 thus requires careful quantification of the error in approximate counting. Cleaner definitions of these concepts can be given in the theory APC_1^+ of Buss et al. [12]: relativize APC_1 to a new binary function symbol Sz, i.e. take $\mathsf{PV}_1(Sz) + \mathsf{sWPHP}(\mathsf{PV}_1(Sz))$, and add the axiom

$$n, \epsilon^{-1} \in Log \wedge C \text{ is a circuit with } n \text{ variables } \to \{x < 2^n \mid C(x) = 1\} \approx_\epsilon Sz(C, 2^n). \quad (30)$$

Intuitively, APC_1^+ adds to APC_1 approximate cardinalities with error smaller than all inverse polynomial factors simultaneously but does not add any reasoning power. More precisely, the following is [12, Proposition 27]. Its proof builds on Jeřábek's theory HARD^A mentioned in Section 2.3.

Let Σ^b_∞ denote the set of all bounded PV-formulas.

Theorem 3.23. *The theory APC_1^+ is Σ^b_∞-conservative over APC_1.*

For $X \subseteq 2^n$ defined by circuit C we write $Sz(X)$ for $Sz(C, 2^n)$.

Definition 3.24 (in APC_1^+)**.** For circuit definable $X \subseteq 2^{|t|}$ set

$$\Pr^+_{x<t}[x \in X] := Sz\big(\{x \in X \mid x < t\}\big)/t.$$

Of course, approximate probabilities in APC_1^+ and APC_1 are approximately the same:

Lemma 3.25. *The theory APC_1^+ proves for all t, circuit definable $X \subseteq 2^{|t|}$ and $0 \leqslant p, \epsilon, \gamma \leqslant 1$ with $\gamma^{-1} \in Log$:*

(i) if $\Pr_{x<t}[x \in X] \succcurlyeq_\epsilon p$, then $\Pr^+_{x<t}[x \in X] \geqslant p - (2\epsilon + \gamma)$;
if $\Pr_{x<t}[x \in X] \preccurlyeq_\epsilon p$, then $\Pr^+_{x<t}[x \in X] \leqslant p + (2\epsilon + \gamma)$;

(ii) if $\Pr^+_{x<t}[x \in X] \geqslant p$, then $\Pr_{x<t}[x \in X] \succcurlyeq_\gamma p$;
if $\Pr^+_{x<t}[x \in X] \leqslant p$, then $\Pr_{x<t}[x \in X] \preccurlyeq_\gamma p$.

Proof. (i): we only show the first statement. If $\Pr_{x<t}[x \in X] \succsim_\epsilon p$, then by (30)

$$\Pr^+_{x<t}[x \in X] \cdot t = Sz(\{x \in X \mid x < t\}) \approx_{\gamma/4} \{x \in X \mid x < t\} \succsim_\epsilon pt.$$

This implies $\Pr^+_{x<t}[x \in X] \geqslant p - (2\epsilon + \gamma)$ via Proposition 2.10 (ii):

$$pt \leqslant \Pr^+_{x<t}[x \in X] \cdot t + (\epsilon + \gamma/4 + \gamma/4) \cdot 2^{|t|} \leqslant \Pr^+_{x<t}[x \in X] \cdot t + (\epsilon + \gamma/2) \cdot 2t.$$

(ii): again, we only show the first statement. If $\Pr^+_{x<t}[x \in X] \geqslant p$, then by (30)

$$\{x \in X \mid x < t\} \approx_{\gamma/4} Sz(\{x \in X \mid x < t\}) \geqslant pt,$$

so $\{x \in X \mid x < t\} \succsim_\gamma pt$, i.e. $\Pr_{x<t}[x \in X] \succsim_\gamma p$. □

Definition 3.26 (in APC_1^+)**.** Let $s \in Log$. A circuit G with k variables and $2k$ outputs is an *s-secure pseudorandom generator* if for all circuits C with $2k$ variables and size $\leqslant s$:

$$\left| \Pr^+_{y<2^{2k}} \left[C(y) = 1\right] - \Pr^+_{x<2^k} \left[C(G(x)) = 1\right] \right| < 1/s.$$

As Chow [17, Theorem 1] we present Razborov and Rudich's naturalization barrier as one to proving a fixed polynomial circuit lower bound. Since approximate counting incurs inverse polynomial errors we use Razborov and Rudich's largeness parameter 2^{-dm} instead of Chow's 2^{-m^d}.

Theorem 3.27 (Natural proof barrier)**.** *For all $c, d \in \mathbb{N}$ and $0 < \delta < 1$ there is $k_0 \in \mathbb{N}$ such that APC_1^+ proves for all $k \geqslant k_0$ with $k^\delta \in LogLog$ and $m := \lceil k^{\delta/2} \rceil$: if*

(Constructivity) *C is a circuit with 2^m variables and size $\leqslant 2^{dm}$,*

(Largeness) $\Pr^+_{f<2^{2^m}}[C(f) = 1] \geqslant 1/2^{dm}$*,*

(Usefulness) *C accepts only functions of circuit complexity $> (c+4)m^{(1+2c/\delta)}$, i.e.*

$$\forall f, D, M \; \big(C(f) = 1 \to \mathsf{LB}_{\mathsf{tt}}(f, D, (c+4)m^{(1+2c/\delta)}, m, M)\big)$$

then 2^{k^δ}-secure pseudorandom generators with k variables and size $\leqslant ck^c$ do not exist.

Proof. Argue in APC_1^+. Assume k is large enough and G is a size $\leqslant ck^c$ circuit with k variables and $2k$ outputs. Assuming further that C is as stated we show G is not a 2^{k^δ}-secure pseudorandom generator.

Let $G' : 2^k \times 2 \to 2^k$ be a size $\leqslant 4k + ck^c$ circuit that maps $\langle x, 0 \rangle$ and $\langle x, 1 \rangle$ respectively to the first and the last k bits of $G(x)$. For $b < 2$ we write $G^b(x) := G'(\langle x, b \rangle)$. For $y < 2^m$ write y_i for $bit(i, y)$. Consider a circuit $G'' : 2^k \times 2^m \to 2$ that maps $\langle x, y \rangle$ to

$$G''(\langle x, y \rangle) := bit(0, G^{y_{m-1}} \circ \cdots \circ G^{y_0}(x)).$$

Such a circuit is constructed using m copies of G' so has size $\leqslant (c{+}4)m^{(1+2c/\delta)}$. Hardwiring some fixed $x < 2^k$ into G'' computes the function $y \mapsto G''(\langle x, y\rangle)$. Let $G_x < 2^{2^m}$ be its truth table, i.e. $\mathit{bit}(y, G_x) = G''(\langle x, y\rangle)$ for all $y < 2^m$.

Consider now the binary tree T of height m and let $M := 2^{m+1} - 1$ be the number of its nodes. List its internal nodes $t_1, \dots, t_{2^m-1}$ so that $i < j$ whenever t_i is a child of t_j. Identify its leaves with $[0, 2^m)$. For $i < 2^m$ let T_i be the union of subtrees of T whose nodes are $\{t_1, \dots, t_i\}$ along with all the leaves. For a leaf $y < 2^m$, let $r_i(y)$ be the root of the subtree in T_i containing y, and let $h(i, y)$ denote its height. In particular, T_0 is the set of leaves and $r_0(y) = y$ and $h(0, y) = 0$.

Let a range over $[0, 2^{kM})$ and view it as an assignment mapping nodes t of T to $a(t) < 2^k$. Given such a and $i < 2^m$ define for $y < 2^m$

$$G_i^a(y) := \mathit{bit}(0, G^{y_{m-1}} \circ \cdots \circ G^{y_{m-h(i,y)}}(a(r_i(y)))). \tag{31}$$

Intuitively, $G_i^a(y)$ is obtained from the label $a(r_i(y)) < 2^k$ by iteratively labeling the path in T from $r_i(y)$ to the leaf y, each time applying either G^0 or G^1 to the previous label. Which one is applied depends on the corresponding bit of y. Pictorially speaking, take G^0 if the path turns left, and G^1 if it turns right. Note $G_{i+1}^a(y) = G_i^a(y)$ unless $r_{i+1}(y) = t_{i+1}$. If $r_{i+1}(y) = t_{i+1}$, then $r_i(y)$ is either the left child t_{i+1}^0 or the right child t_{i+1}^1 of t_{i+1}, say the left. Then $G_{i+1}^a(y)$ is given by the r.h.s. of (31) with label $a(r_i(y))$ replaced by $G^0(a(r_{i+1}(y)))$. More generally, assume $a' < 2^{kM}$ is an assignment that agrees with a on $T_i \setminus \{t_{i+1}^0, t_{i+1}^1\}$ and satisfies $a'(t_{i+1}^0) = G^0(a(t_{i+1}))$ and $a'(t_{i+1}^1) = G^1(a(t_{i+1}))$; then for all $y < 2^m$

$$G_i^{a'}(y) = G_{i+1}^a(y). \tag{32}$$

We blur the distinction between the function G_i^a and its truth table, and write

$$p_i := \mathrm{Pr}_a^+[C(G_i^a) = 1]$$

for $i < 2^m$. For r the root of T we have $G_{2^m-1}^a(y) = G''(\langle a(r), y\rangle)$ for all $y < 2^m$, that is, $G_{2^m-1}^a = G_{a(r)}$. Further, $G_0^a(y) = \mathit{bit}(0, a(y))$ for all $y < 2^m$.

We have $\mathrm{Pr}_a[C(G_{2^m-1}^a) = 1] \preccurlyeq_0 0$ by (Usefulness), so $p_{2^m-1} \leqslant 1/2^{dm+1}$ by Lemma 3.25 (i). By (Largeness) and Lemma 3.25 (ii), $\mathrm{Pr}_{f<2^{2^m}}[C(f) = 1] \succcurlyeq_\gamma 1/2^{dm}$ for any $\gamma^{-1} \in \mathit{Log}$. It is easy to see that this implies $\mathrm{Pr}_a[C(G_0^a) = 1] \succcurlyeq_\gamma 1/2^{dm}$ and hence $p_0 \geqslant 1/2^{dm} - 3\gamma$ by Lemma 3.25 (i). Thus,

$$p_0 - p_{2^m-1} \geqslant 1/2^{dm+1} - 3\gamma.$$

Setting $\gamma := 1/(6 \cdot 2^{dm+1})$ the r.h.s. is $\geqslant 1/2^{dm+2}$. The l.h.s. is $\leqslant \sum_{i<2^m-1} |p_i - p_{i+1}|$, so $|p_j - p_{j+1}| \geqslant 1/2^{(d+1)m+2}$ for some $j < 2^m - 1$. For simplicity assume $p_j \geqslant p_{j+1}$, so

$$p_j - p_{j+1} \geqslant 1/2^{(d+1)m+2}. \tag{33}$$

Rewrite the event $\{a \mid C(G_j^a) = 1\}$ as a set of pairs $\langle b_0, b_1\rangle$ with $b_0 < 2^{2k}$ determining $a(t_{j+1}^0)$ and $a(t_{j+1}^1)$ for the children t_{j+1}^0, t_{j+1}^1 of t_{j+1} in T, and $b_1 < 2^{k(M-2)}$ determining the rest of a. Accordingly write $G_j^{\langle b_0, b_1\rangle}$ for G_j^a.

Claim 3.28. There exists $b_1 < 2^{k(M-2)}$ such that

$$\Pr^+_{b_0<2^{2k}}[C(G_j^{\langle b_0,b_1\rangle}) = 1] - \Pr^+_{a_0<2^k}[C(G_j^{\langle G(a_0),b_1\rangle}) = 1] \geqslant 1/2^{(d+1)m+3}. \tag{34}$$

This claim implies the theorem. Indeed, for large enough standard e the function $b_0 \mapsto G_j^{\langle b_0,b_1\rangle}$ can be computed by circuits of size $\leqslant 2^{em}$ applying (31) for all leaves $y < 2^m$ above t^0_{j+1}, t^1_{j+1}. Thus, the events in (34) are defined by circuits of size $\leqslant 2^{em+1}$. Since $2^{(d+1)m+3}, 2^{em+1} \leqslant 2^{k^\delta}$ for large enough k, (34) means that G is not 2^{k^δ}-secure.

We are left to prove Claim 3.28. Assume for the sake of contradiction that there does not exists b_1 as claimed. Let $q_0(b_1)$ and $q_1(b_1)$ denote the two probabilities in (34), so failure of (34) means $q_0(b_1) - q_1(b_1) < 1/2^{(d+1)m+3}$.

Let $\gamma^{-1} \in \mathit{Log}$ be arbitrary. By Lemma 3.25 (ii), $\{b_0 \mid C(G_j^{\langle b_0,b_1\rangle}) = 1\} \preccurlyeq_\gamma q_0(b_1)2^{2k}$ and $q_1(b_1)2^k \preccurlyeq_\gamma \{a_0 \mid C(G_j^{\langle G(a_0),b_1\rangle}) = 1\}$. The latter implies

$$\big(q_1(b_1) + 1/2^{(d+1)m+3}\big) \cdot 2^{2k} \preccurlyeq_\gamma 2^k \times \big\{a_0 \mid C(G_j^{\langle G(a_0),b_1\rangle}) = 1\big\} \,\dot\cup\, 2^{2k}/2^{(d+1)m+3}.$$

By failure of (34), the l.h.s. is $> q_0(b_1)2^{2k}$. Hence,

$$\big\{b_0 \mid C(G_j^{\langle b_0,b_1\rangle}) = 1\big\} \preccurlyeq_{2\gamma} \big\{\langle a_0', a_0\rangle \mid C(G_j^{\langle G(a_0),b_1\rangle}) = 1\big\} \,\dot\cup\, 2^{2k}/2^{(d+1)m+3},$$

where a_0' ranges over $[0, 2^k)$. We combine these estimations, one for each $b_1 < 2^{k(M-2)}$, using Lemmas 2.11 and 2.14:

$$\big\{\langle b_0, b_1\rangle \mid C(G_j^{\langle b_0,b_1\rangle}) = 1\big\} \preccurlyeq_{16\gamma} \big\{\langle\langle a_0', a_0\rangle, b_1\rangle \mid C(G_j^{\langle G(a_0),b_1\rangle}) = 1\big\} \,\dot\cup\, 2^{kM}/2^{(d+1)m+3}. \tag{35}$$

Consider the bijection from the set of $\langle\langle a_0', a_0\rangle, b_1\rangle$ with $a_0', a_0 < 2^k$ onto $[0, 2^{kM})$ that maps such a triple to the $a < 2^{kM}$ that assigns the nodes $t_{j+1}, t^0_{j+1}, t^1_{j+1}$ to $a_0, a_0', b_1(t_{j+1})$, respectively, and all other nodes t to $b_1(t)$. Then, by (32),

$$G_j^{\langle G(a_0),b_1\rangle} = G^a_{j+1}.$$

Thus, our bijection maps $\big\{\langle\langle a_0', a_0\rangle, b_1\rangle \mid C(G_j^{\langle G(a_0),b_1\rangle}) = 1\big\}$ onto $\{a \mid C(G^a_{j+1}) = 1\}$.

Hence, the r.h.s. of (35) is $\approx_\gamma (p_{j+1} + 1/2^{(d+1)m+3})2^{kM}$. The l.h.s. of (35) is $\approx_\gamma p_j 2^{kM}$. By Proposition 2.10 (ii), $p_j 2^{kM} \leqslant (p_{j+1} + 1/2^{(d+1)m+3})2^{kM} + 19\gamma 2^{kM}$. But this contradicts (33) if we set $\gamma := 1/20 \cdot 1/2^{(d+1)m+3}$. □

4 Propositional proof complexity

4.1 Propositional translation

To fix some notation we briefly recall the propositional simulation of PV_1 by EF going back to Cook [20]. We choose a particular variant of the propositional translation from the literature and use it to define the propositional tt-formulas (2) from the Introduction. This is for definiteness. The reader's favorite versions of the definitions of the translation and the tt-formulas can be used for the results in Sections 4.3 and 4.4 provided there are short EF-proofs of equivalence to our versions (which is not much to ask for).

We write propositional formulas in de Morgan language $\wedge, \vee, \neg, 0, 1$. Fix some standard propositional proof system given by finitely many (axiom schemes and) inference rules; we refer to its proofs as *Frege proofs*. *Extended Frege* EF additionally allows to abbreviate formulas by atoms during the proof. The *depth* of a Frege proof is the minimal d such that every formula (viewed as a circuit) appearing in it has depth $\leqslant d$. We refer to [34, Sections 4.4, 4.5] for definitions.

The propositional translation $[\![\varphi]\!]^{\bar{n}}$ is defined for a Σ^b_0-formula $\varphi(x_1, \ldots, x_k)$ and *length bounds* $\bar{n} = (n_1, \ldots, n_k) \in \mathbb{N}^k$ *associated to* its free variables. Its size is polynomial in $\bar{n}$. It has n_i propositional variables *corresponding* to x_i plus some auxiliary variables. A tuple $(a_1, \ldots, a_k) \in \prod_{i=1}^k [0, 2^{n_i})$ satisfies φ in the standard model if and only if

$$[\![\varphi]\!]^{\bar{n}}\ [a_1/x_1, \ldots, a_k/x_k]$$

is tautological. Here we allow ourselves some convenient but nonstandard notation: by $[a_1/x_1, \ldots, a_k/x_k]$ we mean the substitution that for all $1 \leqslant i \leqslant k$ substitutes the Boolean constants $\mathit{bit}(0, a_i), \ldots, \mathit{bit}(n_i - 1, a_i)$ for the n_i many variables corresponding to x_i.

We fix some *bounding polynomials* p_t for terms $t(\bar{x})$ once and for all: $t(\bar{x})$ takes values of length $\leqslant p_t(\bar{n})$ on arguments of lengths $\bar{n}$. We assume that variables x have the identity as bounding polynomial p_x. The translation is defined by induction on the logical complexity of φ with straightforward inductive clauses. For example,

$$[\![\exists y{<}|t(\bar{x})|\ \varphi(\bar{x}, y)]\!]^{\bar{n}} := \textstyle\bigvee_{a<p_t(\bar{n})} [\![y \leqslant |t(\bar{x})| \wedge \varphi(\bar{x}, y)]\!]^{\bar{n}, |p_t(\bar{n})|}\ [a/y]. \tag{36}$$

More precisely, we should write $t(\bar{x}')$ for the subtuple $\bar{x}'$ of variables from $\bar{x}$ that actually occur in t. We refer to [28, Section 2] for more details.

Theorem 4.1 (Simulation, Cook 1975)**.** *If* S^1_2 *proves* $\varphi(\bar{x}) \in \Sigma^b_0$*, then there is a polynomial time algorithm that, given a tuple* $\bar{n}$ *of naturals in unary, computes an* EF*-proof of* $[\![\varphi(\bar{x})]\!]^{\bar{n}}$.

In [28, Section 2] Jeřábek introduced the propositional proof system WF. Intuitively, WF extends EF by an axiom scheme encoding the surjective weak pigeonhole principle for circuits. We refer to [28, section 2] for the definition. All we need to know about this system is the following simulation result [28, Proposition 2.12]:

Theorem 4.2 (Simulation, Jeřábek 2004)**.** *If* $\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV})$ *proves* $\varphi(\bar{x}) \in \Sigma^b_0$*, then there is a polynomial time algorithm that, given a tuple* $\bar{n}$ *of naturals in unary, computes a* WF*-proof of* $[\![\varphi(\bar{x})]\!]^{\bar{n}}$.

Remark 4.3. We comment on variants of Theorem 4.1 appearing in the literature and motivate our choice [28]. As for some minor differences, the original source [20] uses Tseitin's [68] Extended Resolution and translates only quantifier-free PV-formulas, [34, Section 9.2] uses the QBF system G_1, [7, 10] use EF but translate only formulas in Buss' language instead of PV. In contrast to [28] the various translations [20, 34, 7, 41] all use only a single length bound n associated to all variables. Such translations are with respect to a bounding polynomial that works for all terms appearing in the formula. This has the unpleasant property that the translation of a formula can vary when considered a subformula of another. Another unpleasant property is that proofs of analogues of Theorem 4.1 in [34] and [7] need to choose a bounding polynomial that works for all formulas in the simulated PV_1-proof, so the translation depends on this proof instead of only the formula proved – see the statements of [34, Theorem 9.2.5, Corollaries 9.2.6, 9.2.7]. The statements in [7, Theorem 30] and the underlying lecture notes [10, p.10-6] should be rephrased accordingly.

4.2 Propositional formalizations of circuit lower bounds

Note that our propositional translation is not applicable to the succinct formula $\mathsf{LB}[\mathsf{C},\mathsf{Q}]$ (see (6) in Section 2.2) because its quantifier complexity is too high. As announced in the Introduction we discuss two ways how to obtain succinct size $n^{O(1)}$ propositional formulas expressing circuit lower bounds in Sections 4.3 and 4.5. Here, we define "truth table" formulas of exponential size.

First consider the formula $\mathsf{LB}_{\mathsf{tt}}[\mathsf{C}]$, see (4) in Section 2.2. We use variable x instead of n to avoid a double use of this letter, and substitute for the 'size' variable s a PV-term $s(N)$. Thus we consider the formula $\mathsf{LB}_{\mathsf{tt}}[\mathsf{C}](f, C, s(N), x, N)$ with free variables f, C, x, N. We omit superscripts in the translations and understand that f, C, x, N, y have associated length bounds $2^n, 2^n, |n|, 2^n, n$ respectively.

We define

$$\mathsf{tt}[\mathsf{C}, f, s(2^n)] := \llbracket \mathsf{LB}_{\mathsf{tt}}[\mathsf{C}](f, C, s(N), x, N) \rrbracket \; [2^{2^n} - 1/N, n/x]. \tag{37}$$

We omit C if it is the class of all circuits, thus writing $\mathsf{tt}[f, s(2^n)]$. Here and below, note $[2^{2^n} - 1/N]$ substitutes 2^n many Boolean constants 1 for the 2^n variables corresponding to N. Apart from some auxiliary variables this formula has 2^n many variables for the bits of f and 2^n many variables for the bits of C. It has size $2^{O(n)}$.

Recalling $\mathsf{LB}^0_{\mathsf{tt}}[\mathsf{C}]$ from (5) in Section 2.2, we see that our formula has the desired form (2) from the Introduction:

$$\begin{aligned} &\mathsf{tt}[\mathsf{C}, f, s(2^n)] = \textstyle\bigvee_{a<2^n} \text{“}C(a)\neq f(a)\text{”} \\ &\qquad \text{with “}C(a)\neq f(a)\text{”} := \llbracket \mathsf{LB}^0_{\mathsf{tt}}[\mathsf{C}](f, C, s(N), x, N, y) \rrbracket \; [2^{2^n} - 1/N, n/x, a/y]. \end{aligned} \tag{38}$$

We shall need two additional formulas, one obtained by specifying the circuit C and another by fixing the function f.

Specifying C: Recall Wilkie's Witnessing Theorem 2.7 states that a Σ^b_1-consequence $\exists y \varphi(y, \bar{x})$ of APC_1 are witnessed by probabilistic polynomial time functions. One can also get a win-win type of witnessing by a deterministic polynomial time function as follows: additionally to $\bar{x}$ the function takes as input a truth table of a function $\tilde{f} : \{0,1\}^m \to \{0,1\}$ and computes either a witness y to $\varphi(y, \bar{x})$ as desired or a size $2^{\epsilon m}$ circuit computing $\tilde{f}$. Here $0 < \epsilon < 1$ is an arbitrary but fixed rational and the parameter m can be chosen polylogarithmic in the other inputs. We refer to Lemma 4.9 for a precise statement.

We now define propositional formulas expressing that a function given by a truth table is not computed by a specific size $2^{\epsilon x}$ circuit which is computed by a PV-function. As the formulas are to be used in the context above we define them using a copy $\tilde{f}, \tilde{C}, \tilde{x}, \tilde{N}, \tilde{y}$ of the variables f, C, x, N, y and substitute a function $\mathit{circ}(\bar{x}, \tilde{f}, \tilde{x}, \tilde{N})$ for $\tilde{C}$:

$$\bigvee\nolimits_{a<2^m} \text{“}\mathit{circ}(\bar{x}, \tilde{f}, \cdot)(a) {\neq} \tilde{f}(a)\text{”}. \tag{39}$$

This formula will have propositional variables for the bits of $\tilde{f}$ (i.e. the truth table) and $\bar{x}$ plus auxiliary variables. The dot $\cdot$ indicates that circ has more arguments that are not displayed because they do not have corresponding propositional variables in the formula. The definition assumes that the length bounds associated with the variables $\tilde{f}, \tilde{x}, \tilde{N}, \tilde{y}$ are $2^m, |m|, 2^m, m$, and those associated with $\bar{x}$ are given by context:

$$\text{“}\mathit{circ}(\bar{x}, \tilde{f}, \cdot)(a) {\neq} \tilde{f}(a)\text{”} := [\![\mathsf{LB}^0_{\mathsf{tt}}(\tilde{f}, \mathit{circ}(\bar{x}, \tilde{f}, \tilde{x}, \tilde{N}), \tilde{x}, |\tilde{N}|^\epsilon, \tilde{N}, \tilde{y})]\!]\ [2^{2^m} - 1/\tilde{N}, m/\tilde{x}, a/\tilde{y}].$$

Specifying f: A straightforward way to define propositional formulas expressing circuit lower bounds for a particular problem Q would be to substitute the truth table of Q restricted to $y < 2^n$ for the variables corresponding to f in $\mathsf{tt}[\mathsf{C}, f, s(2^n)]$.

For example, the formula $\mathsf{tt}[\mathsf{SAT}, s(2^n)]$ from the Introduction could be defined as $\mathsf{tt}[\mathsf{C}, f, 2^{\epsilon n}]\ [\mathit{sat}/f]$ where C is the class of all circuits and $\mathit{sat} < 2^{2^n}$ is the number whose bits give the truth table of SAT restricted to $y < 2^n$.

Remark 4.4. As mentioned in the Introduction, circuit lower bounds yield candidate hard tautologies for EF or Frege: for a rational $0 < \epsilon < 1$ one asks whether all infinite subsets of $\left\{\mathsf{tt}[f, 2^{\epsilon n}]\ [h/f] \mid h < 2^{2^n}, n \in \mathbb{N}\right\}$ are hard for EF or Frege.

We shall however not use these formulas but proceed differently. The reason is that a disjunct “$C(a){\neq}f(a)$”$[\mathit{sat}/f]$ of $\mathsf{tt}[\mathsf{SAT}, n^k]$ has size $2^{O(n)}$ while the same is easily expressed in size $n^{O(1)}$. One way is to translate the formula $\mathsf{LB}^0[\mathsf{C}, \mathsf{Q}]$ (see (7) in Section 2.2).

Let $k \in \mathbb{N}$ and assume the defining formula $\mathsf{Q}(y)$ of Q is Σ^b_0 (i.e. $\mathsf{Q} \in \mathsf{P}$). In this case our translation applies to the formula $\mathsf{LB}^0[\mathsf{C}, \mathsf{Q}](C, x^k, x, N, y)$. We agree that the free variables C, x, N, y have associated length bounds $n^{k+1}, |n|, n, n$. Note that a size $s \geqslant n$

circuit with n variables is naturally coded by $O(s\cdot|s|)$ bits, so, if n is large enough, the n^{k+1} variables corresponding to C are enough to hold an encoding of a size n^k circuit C.

We define

$$\begin{aligned}&\mathsf{tt}[\mathsf{C},\mathsf{Q},n^k] := \textstyle\bigvee_{a<2^n} \text{“}C(a){\neq}\mathsf{Q}(a)\text{”} \\ &\qquad \text{with “}C(a){\neq}\mathsf{Q}(a)\text{” } := [\![\mathsf{LB}^0[\mathsf{C},\mathsf{Q}](C,x^k,x,N,y)]\!]\ [2^n-1/N, n/x, a/y].\end{aligned} \tag{40}$$

Note that for every $a < 2^n$ the subformula “$C(a){\neq}\mathsf{Q}(a)$” has size $n^{O(1)}$. We do not mention C if it is the class of all circuits, thus writing $\mathsf{tt}[\mathsf{Q},n^k]$.

4.3 Succinct tautologies via witnessing

In case the existential quantifier $\exists y{<}1\#N$ in the formula $\mathsf{LB}[\mathsf{C},\mathsf{Q}]$ can be witnessed by a polynomial time algorithm, we get a Σ^b_0-formula whose propositional translation is a succinct size $n^{O(1)}$ expression of a circuit lower bound:

Definition 4.5. Let $\mathsf{Q} \subseteq \mathbb{N}$ be Σ^b_0-defined. For ternary $w \in \mathsf{PV}$ define

$$\mathsf{lb}_w[\mathsf{C},\mathsf{Q},n^k] := [\![\mathsf{LB}^0[\mathsf{C},\mathsf{Q}](C,x^k,x,N,w(C,x,N))]\!]\ [2^n-1/N, n/x]. \tag{41}$$

We define $\mathsf{lb}_{w(\cdot,\bar z)}[\mathsf{C},\mathsf{Q},n^k]$ similarly for $w(C,x,N,\bar z)$ having additional arguments $\bar z$ which we refer to as *parameters of* w. The notation is explained only in contexts associating length bounds to $\bar z$; in particular, when applying a substitution $\mathsf{lb}_{w(\cdot,\bar z)}[\mathsf{C},\mathsf{Q},n^k]\ [\bar a/\bar z]$ for a tuple $\bar a$ from $\mathbb{N}$, we understand that these length bounds are the lengths of the numbers in $\bar a$. Again, we shall omit C from these notations if it is the set of all circuits.

Remark 4.6. Continuing Remark 4.4 a suggestive notation would be $\mathsf{lb}^k_{\mathsf{P/poly}}[\mathsf{C},\mathsf{Q}]$ for the set of formulas $\mathsf{lb}_{w(\cdot,\bar z)}[\mathsf{C},\mathsf{Q},n^k]\ [\bar a/\bar z]$ for all $w \in \mathsf{PV}$ and all tuples $\bar a$ from $\mathbb{N}$. The following definition explains these formulas also for $\mathsf{Q} = \mathsf{SAT}$, and the following proposition points out that likely these formulas are tautological for some w. Intuitively, these formulas are even harder than $\mathsf{tt}[\mathsf{SAT},n^k], n \in \mathbb{N}$. We shall, however, not need this notation.

Definition 4.5 can be extended to $\mathsf{Q} \in \mathsf{NP}$ as follows. We use standard symbols $(x)_0, (x)_1$ from PV giving the first and second component of the ordered pair coded by x.

Definition 4.7. Let $\mathsf{Q} \subseteq \mathbb{N}$ be defined by $\exists z{<}t(y)\ \varphi(z,y)$ where $t(y)$ is a PV-term and $\varphi(z,y) \in \Sigma^b_0$. For $w(C,x,N) \in \mathsf{PV}$ define $\mathsf{lb}_w[\mathsf{C},\mathsf{Q},n^k]$ as

$$\begin{aligned}&\Big[\!\!\Big[n = |N| \to \Big(w_1 < 1\#N \wedge w_0 < t(w_1) \wedge \\ &\quad \Big(C \text{ is a } \mathsf{C}\text{-circuit of size} \leqslant x^k \to \\ &\qquad \big(C(w_1) = 0 \wedge \varphi(w_0,w_1)\big) \vee \big(C(w_1) = 1 \wedge \neg\varphi(z,w_1)\big)\Big)\Big)\Big]\!\!\Big]\ [2^n-1/N, n/x],\end{aligned} \tag{42}$$

where, for readability, we abbreviated $(w(C,x,N))_0, (w(C,x,N))_1$ by w_0, w_1. The length bound associated to z is $p_t(n)$, that is, the bounding polynomial p_t of t evaluated at the length bound associated to y.

For $\mathsf{Q} \in \mathsf{P}$ the formula $\mathsf{LB}[\mathsf{C},\mathsf{Q}]$ for $s = n^k$ is Σ^b_1, so in case PV_1 proves it, Theorem 2.1 implies that there exists $w \in \mathsf{PV}$ such that $\mathsf{lb}_w[\mathsf{C},\mathsf{Q},n^k]$ is tautological. This reasoning does not apply for $\mathsf{Q} \in \mathsf{NP}$ because then $\mathsf{LB}[\mathsf{C},\mathsf{Q}] \in \Sigma^b_2$. In this case, provability in PV_1 implies by the KPT-theorem [42] that the existential quantifier $\exists y$ is witnessed by a tuple of polynomial time functions $\bar{w}$ determining a constant round Student-Teacher computation. The corresponding translation gives size $n^{O(1)}$ formulas $\mathsf{lb}_{\bar{w}}$ weaker than the formulas lb_w defined above. We omit their definition and discussion here and refer the interested reader to [54]. Instead, we include a proof that, under some plausible hardness assumptions, the stronger witnessing with a single w is possible for $\mathsf{Q} = \mathsf{SAT}$. This improves [52, Proposition 4.3] establishing a one round Student-Teacher computation, and, in fact, is a combination of folklore arguments (e.g. [8, 22, 45] contain similar constructions).

Proposition 4.8. *Assume there exists a* one-way permutation*, that is, a polynomial time computable, length preserving bijection $f : \mathbb{N} \to \mathbb{N}$ such that for all $k, \ell \in \mathbb{N}$ there is $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$ and every size $\leqslant n^k$ circuit C with n variables and n outputs*

$$\Pr_{x<2^n}[C(f(x)) = x] < 1/n^\ell.$$

Assume further that there exists $h : \mathbb{N} \to 2$ computable in time $2^{O(n)}$ with hardness $2^{\Omega(n)}$*, that is, there is $\delta > 0$ such that for all sufficiently large n and all size $2^{\delta n}$ circuits C with n variables and 1 output we have*

$$\Pr_{x<2^n}[C(x) = h(x)] < 1/2 + 1/2^{\delta n}.$$

Then for all $k \in \mathbb{N}$ there are $n_0 \in \mathbb{N}$ and a polynomial time algorithm which given $n \geqslant n_0$ in unary and a circuit C of size $\leqslant n^k$ computes $y < 2^n$ such that C on y does not decide SAT*, i.e. either $y \in \mathsf{SAT}, C(y) = 0$ or $y \notin \mathsf{SAT}, C(y) = 1$. Additionally, the algorithm outputs an assignment z that satisfies y in case $y \in \mathsf{SAT}$.*

In other words, there is $w(C,x,N) \in \mathsf{PV}$ such that $\mathsf{lb}_w[\mathsf{SAT},n^k]$ is tautological for sufficiently large n.

Proof. Given $b \in \mathbb{N}$ we can compute in polynomial time a propositional formula α_b expressing "$f(x) = b$": its variables include $x_0, \ldots, x_{|b|-1}$; it has exactly one satisfying assignment and this assignment assigns $bit(i, f^{-1}(b))$ to x_i; moreover, there is a polynomial time function mapping $f^{-1}(b)$ to a satisfying assignment. For $\bar{\epsilon} \in \{0,1\}^{\leqslant |b|}$ let $\alpha_b[\bar{\epsilon}]$ be the formula obtained from α_b by substituting the i-th bit of $\bar{\epsilon}$ for x_{i-1}.

Let C be a circuit with n variables and size n^k. Choose $n \geqslant m \geqslant n^{\Omega(1)}$ such that the formulas $\alpha_b[\bar{\epsilon}]$ for $b < 2^m$ have size $\leqslant n$ and 'padded versions' $\alpha^n_b[\bar{\epsilon}]$ have size exactly n;

these 'padded versions' are logically equivalent formulas with the same variables and computable in time $n^{O(1)}$.

By the usual self-reducibility argument we find a circuit D which on $b < 2^m$ computes $a := f^{-1}(b)$ if C decides SAT on all formulas $\alpha_b^n[bit(0,a), \ldots, bit(i-1,a), 1], i < m$. As $m \geqslant n^{\Omega(1)}$, the size of D is $\leqslant m^\ell$ for some $\ell \in \mathbb{N}$. Since f is one-way we have, assuming n and hence m is large enough,

$$\Pr_{a<2^m}[D(f(a)) = a] < 1/m.$$

Let D' be a circuit that given $a < 2^m$ checks whether $D(f(a)) = a$. This circuit can be chosen of size $m^{\ell'}$ for some $\ell' \in \mathbb{N}$.

Based on the hard function h as in our assumption we get the Nisan-Wigderson generator [48]. That is, there is a constant $c \in \mathbb{N}$ depending only on ℓ' and a function $G : 2^{c \log m} \to 2^m$ such that (in fact, for all $m^{\ell'}$-size circuits, not only D')

$$\big| \Pr_{a<2^m}[D'(a) = 1] - \Pr_{s<2^{c \log m}}[D'(G(s)) = 1] \big| < 1/m.$$

Moreover, given $s < 2^{c \log m}$ and m (in binary), $G(s)$ can be computed in time $m^{O(1)}$.

It follows that $\Pr_{s<2^{c \log m}}[D'(G(s)) = 1] < 1$, so there exists $s < 2^{c \log m}$ such that $D(f(G(s))) \neq G(s)$. Hence there exists $i < m$ such that C does not decide SAT on the size n formula $\alpha^n_{f(G(s))}[bit(0, G(s)), \ldots, bit(i-1, G(s)), 1]$.

Note these are $\leqslant m^c \cdot m \leqslant n^{c+1}$ many formulas. Our witnessing function w runs C on all of them and outputs the first where C does not decide SAT. This is easy to detect because we know which of our formulas are satisfiable: those with $bit(i, G(s)) = 1$. Clearly, if the formula found is satisfiable, then a satisfying assignment can be computed in polynomial time from $G(s)$. □

4.4 A general upper bound

Given our APC_1 proofs of circuit lower bounds $\mathsf{LB}[\mathsf{C}, \mathsf{Q}]$ we would like to conclude that WF admits short proofs of tautologies $\mathsf{lb}_w[\mathsf{C}, \mathsf{Q}, n^k]$ for some w. Unfortunately, this does not follow directly because a priori the APC_1-proof yields a witnessing w computable not in deterministic but probabilistic polynomial time (see Section 3.5). We deal with this complication by reformulating the simulation in terms of an implication. We observe that for proving a Σ^b_1-formula in APC_1 the truth table of a single hard function can replace $\mathsf{sWPHP}(\mathsf{PV})$ in such a way that, in particular, APC_1-proofs of $\mathsf{LB}[\mathsf{C}, \mathsf{Q}]$ for $s = n^k$ translate to short EF proofs of tautologies stating that a truth-table of a single hard function implies $\mathsf{lb}_w[\mathsf{C}, \mathsf{Q}, n^k]$.

For a tuple $\bar{x} = (x_0, \ldots, x_{k-1})$ of variables we write $|\bar{x}|$ for $\max_{i<k} |x_i|$.

Lemma 4.9. *Suppose* $\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV})$ *proves* $\exists y \varphi(y, \bar{x})$ *for* $\varphi(y, \bar{x}) \in \Sigma^b_1$. *For every rational* $0 < \epsilon < 1$ *there are* $\ell \in \mathbb{N}$ *and* $g \in \mathsf{PV}$ *such that* PV_1 *proves*

$$|N| \geqslant |\bar{x}|^\ell \wedge \mathsf{LB}_{\mathsf{tt}}(f, (g(\bar{x}, f, n, N))_0, |N|^\epsilon, n, N) \ \to\ \varphi((g(\bar{x}, f, n, N))_1, \bar{x}).$$

Proof. It suffices to prove this when $\bar{x}$ is a single variable x. It is well-known (see e.g. [31, Theorem 3.1 (i)]) that $\mathsf{sWPHP}(\mathsf{PV})$ is, over S^1_2, equivalent to the more familiar version with x pigeons and x^2 holes (i.e. replace in (9) the bounds $x|y|$ and $x(|y|+1)$ by x and x^2 respectively). Now, if $\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV})$ proves $\exists y \varphi(y, x)$, then, following Thapen's proof of [65, Theorem 4.2] (based on [64, Section 2]; cf. also [28, Proposition 1.14]), there are $\ell_0 \in \mathbb{N}$ and a unary $h \in \mathsf{PV}$ such that S^1_2 proves

$$\exists y \varphi(y, x) \ \vee\ \forall v{<}2^{8|x|^{\ell_0}} \exists u{<}2^{4|x|^{\ell_0}}\ h(u) = v.$$

By Buss' Witnessing Theorem 2.1 it now suffices to show that for every (standard) positive rational $\epsilon < 1$ there is $\ell \in \mathbb{N}$ such that S^1_2 proves

$$\forall v{<}2^{8|x|^{\ell_0}} \exists u{<}2^{4|x|^{\ell_0}}\ h(u) = v \ \to\ \big(|N| \geqslant |x|^\ell \to \exists C\ \neg\mathsf{LB}_{\mathsf{tt}}(f, C, |N|^\epsilon, n, N)\big).$$

Argue in S^1_2 and set $m := 4|x|^{\ell_0}$. There is $\ell_1 \in \mathbb{N}$ such that h on 2^m, a surjection from 2^m onto 2^{2m}, is computed by a circuit of size m^{ℓ_1}. Following Jeřábek's S^1_2-proof of [28, Proposition 3.5], this implies that every (number) f viewed as a truth table of length $|f|$ is computed by a size $O(m|m| + m^{\ell_1} \cdot |\lceil |f|/m \rceil|)$ circuit with $||f||$ variables. Set $n := ||f||$ and $N := 2^{2^n} - 1$, so that $2^n = |N|$. The size of this circuit is $\leqslant |f|^\epsilon \leqslant |N|^\epsilon$ if $\ell \in \mathbb{N}$ is sufficiently large and if $|N| = 2^{||f||} \geqslant |x|^\ell$ and hence $|f| \geqslant |x|^\ell/2$. □

Recall the formulas (39) from Section 4.2. The following is our main result concerning upper bounds on the lb_w-formulas.

Theorem 4.10. *Let* $\mathsf{Q} \subseteq \mathbb{N}$ *be* Σ^b_0*-defined,* $k, n_0 \in \mathbb{N}$ *and* $0 < \epsilon < 1$. *If* APC_1 *proves*

$$n_0 \leqslant x \to \mathsf{LB}[\mathsf{C}, \mathsf{Q}](C, x^k, x, N),$$

then there are $\ell \in \mathbb{N}$, $w(C, x, N, \tilde{f}, \tilde{x}, \tilde{N}) \in \mathsf{PV}$, $circ(C, x, N, \tilde{f}, \tilde{x}, \tilde{N}) \in \mathsf{PV}$ *and a polynomial time algorithm which given* 2^m *and* n *in unary such that*

$$n \geqslant n_0 \text{ and } m \geqslant (k+1)\ell \log n$$

computes an EF*-proof of*

$$\begin{aligned} &\textstyle\bigvee_{a<2^m} \text{“}circ(C, x, N, \tilde{f}, \cdot)(a) {\neq} \tilde{f}(a)\text{”} \quad [n/x, 2^n - 1/N] \\ &\to \mathsf{lb}_{w(\cdot, \tilde{f}, \tilde{x}, \tilde{N})}[\mathsf{C}, \mathsf{Q}, n^k]\ [m/\tilde{x}, 2^{2^m} - 1/\tilde{N}]; \end{aligned} \tag{43}$$

moreover, PV_1 *proves that* $circ(C, x, N, \tilde{f}, \tilde{x}, \tilde{N})$ *is a circuit of size* $\leqslant |\tilde{N}|^\epsilon$.

Recall from Section 4.2 that the length bounds associated to the variables $\tilde{f}, \tilde{C}, \tilde{x}, \tilde{N}$ are $2^m, 2^m, |m|, 2^m$, and those associated to C, x, N are $n^{k+1}, |n|, n$. Apart from some auxiliary variables, the formula (43) has variables corresponding to C and $\tilde{f}$, both appearing before and after $\to$. Observe that (43) has size $n^{O(1)}$ for $m := \lceil (k+1)\ell \log n \rceil$.

Proof. By the lemma there are $\mathit{circ}, w \in \mathsf{PV}$ and $\ell \in \mathbb{N}$ such that PV_1 proves

$$\begin{aligned}
&|\tilde{N}| \geqslant |N|^{(k+1)\ell} \geqslant n_0^{(k+1)\ell} \\
&\wedge \mathsf{LB}_{\mathsf{tt}}(\tilde{f}, \mathit{circ}(C, x, N, \tilde{f}, \tilde{x}, \tilde{N}), |\tilde{N}|^{\epsilon}, \tilde{x}, \tilde{N}) \\
&\to \mathsf{LB}^0[\mathsf{C}, \mathsf{Q}](C, x^k, x, N, w(C, x, N, \tilde{f}, \tilde{x}, \tilde{N})),
\end{aligned} \tag{44}$$

Here, we used $\max\{|C|, |x|, |N|\} \leqslant |N|^{k+1}$ if $x = |N|$; this holds because then $|C|$ is implicitly bounded in $\mathsf{LB}^0[\mathsf{C}, \mathsf{Q}]$ by x^{k+1}. It is easy to ensure that *circ* satisfies the "moreover" part of the theorem; if necessary modify the function changing every output which is not a size $\leqslant |\tilde{N}|^{\epsilon}$ circuit to some such circuit not computing $\tilde{f}$.

We apply the translation and a substitution to (44). By Cook's Simulation Theorem 4.1, there is a polynomial time algorithm computing EF-proofs of the formulas

$$\begin{aligned}
&\Big([\![|\tilde{N}| \geqslant |N|^{(k+1)\ell} \geqslant n_0^{(k+1)\ell}]\!] \\
&\quad \wedge [\![\mathsf{LB}_{\mathsf{tt}}(\tilde{f}, \mathit{circ}(C, x, N, \tilde{f}, \tilde{x}, \tilde{N}), |\tilde{N}|^{\epsilon}, \tilde{x}, \tilde{N})]\!] \\
&\quad \to [\![\mathsf{LB}^0[\mathsf{C}, \mathsf{Q}](C, x^k, x, N, w(C, x, N, \tilde{f}, \tilde{x}, \tilde{N}))]\!] \Big) [m/\tilde{x}, 2^{2^m} - 1/\tilde{N}, n/x, 2^n - 1/N].
\end{aligned} \tag{45}$$

This is (43) if we can eliminate the first conjunct (45). But since $m \geqslant (k+1)\ell \log n$ and $n \geqslant n_0$ after the substitution (45) is a tautology whose variables are only the auxiliary variables used in the definition of the translation. These do not appear elsewhere in the formula, so substituting them by arbitrary values gives a true propositional formula without variables which is easy to prove. □

4.5 Succinct tautologies via anticheckers

A rather crude way to define succinct formulas expressing circuit lower bounds is to restrict the disjunction $\bigvee_{a<2^n}$ in (40) to a small subdisjunction:

Definition 4.11. Let $\mathsf{Q} \subseteq \mathbb{N}$ be Σ_0^b-defined. An *antichecker* is a sequence $A = (A_n)_{n\in\mathbb{N}}$ of subsets $A_n \subseteq [0, 2^n)$. It is *polynomially bounded* if $|A_n| \leqslant n^{O(1)}$.

Given an antichecker A define

$$\mathsf{lb}_A[\mathsf{C}, \mathsf{Q}, n^k] := \bigvee\nolimits_{a\in A_n} \text{“}C(a) \neq \mathsf{Q}(a)\text{”} \tag{46}$$

The size of this formula is $(|A_n| + n)^{O(1)}$. We do not mention C if it is the class of all circuits, thus writing $\mathsf{lb}_A[\mathsf{Q}, n^k]$.

The following is a result from [43]:

Theorem 4.12 (Lipton, Young 1994)**.** *Let* $\mathsf{Q} \subseteq \mathbb{N}$ *be* Σ^b_0*-defined. For all* $k \in \mathbb{N}$ *there exists* $\ell \in \mathbb{N}$ *such that for all sufficiently large* $n \in \mathbb{N}$ *the following holds: if* Q *restricted to inputs* $y < 2^n$ *cannot be decided by circuits of size* n^ℓ*, then* $\mathsf{lb}_A[\mathsf{Q}, n^k]$ *is tautological for some antichecker* $A = (A_n)_{n\in\mathbb{N}}$ *with* $|A_n| \leqslant n^\ell$*.*

The lb_A-formulas as well as the lb_w-formulas could be hard tautologies for EF or Frege, and the hope is that this might be easier to show than for the tt-formulas. Intuitively, the lb-formulas are even harder than the tt-formulas because they are, for polynomially bounded anticheckers, exponentially shorter but have the same meaning. We give some evidence for this intuition showing that hardness of lb_A-formulas for Frege follows from hardness of tt-formulas for constant depth Frege. Being hard for *constant depth Frege* means being hard for depth d Frege for all $d \in \mathbb{N}$.

Here, we say that a set Γ of propositional formulas *has short proofs* in a given proof system (and it is *hard* otherwise), if there is a polynomial p such that every $F \in \Gamma$ has a proof of size $p(|F|)$ in the system ($|F|$ is the length of the binary string encoding F).

To feed tt-formulas into constant depth Frege we reformulate them as DNFs:

Lemma 4.13. *There is a polynomial time computable function that maps every propositional formula* F *to a DNF formula* $DNF(F)$ *such that*

(a) F *is tautological if and only if so is* $DNF(F)$*;*

(b) the set of formulas of the form $(F \to DNF(F))$ *has short Frege proofs.*

The proof is standard using extension variables for subformulas of F and goes back to Tseitin [68, pp.115f]. We leave it to the reader.

Proposition 4.14. *Let* $\mathsf{Q} \subseteq \mathbb{N}$ *be* Σ^b_0*-defined,* $k \in \mathbb{N}$ *and* $I \subseteq \mathbb{N}$ *infinite. If the formulas*

$$\mathsf{tt}[\mathsf{Q}, n^k]^{DNF} := \bigvee_{a<2^n} DNF(\text{“}C(a){\neq}\mathsf{Q}(a)\text{”})$$

for $n \in I$ *are hard for constant depth Frege, then for all polynomially bounded anticheckers* $A = (A_n)_{n\in\mathbb{N}}$ *the formulas* $\mathsf{lb}_A[\mathsf{Q}, n^k], n \in I$*, are hard for (unbounded depth) Frege.*

Proof. Suppose there is a polynomially bounded antichecker A and an infinite $I \subseteq \mathbb{N}$ such that the formulas $\mathsf{lb}_A[\mathsf{Q}, n^k], n \in I$, have short Frege proofs. By Lemma 4.13 (b) there are short Frege proofs of $\bigvee_{a\in A_n} DNF(\text{“}C(a){\neq}\mathsf{Q}(a)\text{”}), n \in I$. We can assume the conjunctions and disjunctions are written in a balanced form so that the formula has logical depth $O(\log n)$ (i.e. the formula tree has this depth). Then the main result of Filmus et al. [26, Theorem 3.1] (see [47] for a model-theoretic proof) applies and implies that for sufficiently large $d \in \mathbb{N}$ our formula has depth d Frege proofs of size $2^{O(n)}$. Weakening gives size $2^{O(n)}$ depth d Frege proofs of $\mathsf{tt}[\mathsf{Q}, n^k]^{DNF}$. Since $\mathsf{tt}[\mathsf{Q}, n^k]^{DNF}$ has size $\geqslant 2^n$ these proofs are short. □

Note that $\mathsf{lb}_A[\mathsf{C}, \mathsf{Q}, n^k]$ states that the *partial* truth table $\{(a, \mathsf{Q}(a)) \mid a \in A_n\}$ cannot be computed by a size $\leqslant n^k$ circuit in C. In the next section, we aim to prove a non-uniform version of this formula where instead of a fixed problem Q we have a partial truth table f as input. Identify a partial function f on $\{0,1\}^n$ with its graph

$$f = \big\{(x_i, b_i) \in \{0,1\}^n \times \{0,1\} \mid i < \ell\big\}, \tag{47}$$

where $\ell \in \mathbb{N}$ is the *size* of f. Then formula $\mathsf{ptt}[\mathsf{C}, f, s(n), n, \ell]$ has the form

$$\bigvee\nolimits_{i<\ell} \text{“}C(x_i) \neq b_i\text{”} \tag{48}$$

and expresses that there are no size $s(n)$ C-circuits C computing f. Before giving the definition, we informally point out a motivation from learning: given ℓ input/output pairs associated with a function f as above we wish to predict the value $f(x_\ell)$ on a new input $x_\ell \in \{0,1\}^n$. For this to make sense we have to assume that this value is determined by the ℓ given input/output pairs, so $f(x_\ell)$ is computed by any minimal size circuit C computing f on $x_0, \ldots, x_{\ell-1}$. Say, the minimal circuit C has size $s(n)$. Then the task to predict the bit $f(x_\ell)$ can be formulated as the task to prove the lower bound (48) for circuits of size $s(n)$ and with extra disjunct “$C(x_\ell) \neq b$” for the bit $b := 1 - f(x_\ell)$. It has recently been demonstrated that *natural* proofs of circuit lower bounds indeed imply the existence of learning algorithms [15].

To define the formula (48) we give an ad hoc formalization of lower bounds for partial functions in bounded arithmetic and apply the propositional translation. We remind the reader that our choice is immaterial to a large extent, namely EF-provable equivalence.

View f as in (47) as a number $f < 2^{\ell\cdot(n+1)}$ in turn viewed as a binary string consisting of ℓ blocks of length $n+1$, the i-th one being given by $[f]_i^{n,\ell} < 2^{n+1}$ and meant to code the i-th pair (x_i, b_i) in (47); formally, x_i is $\left\lfloor [f]_i^{n,\ell}/2 \right\rfloor < 2^n$ and b_i is $\mathit{bit}(0, [f]_i^{n,\ell}) < 2$. We formalize this using for n, ℓ variables x, z with associated dummy variables N, L. Further, we use $[u]_i^{x,z}$ as a function symbol in PV. Then the following PV_1-formula expresses a size s C-circuit lower bound for the partial truth table $u < 2^{z\cdot(x+1)}$:

$$\begin{aligned}
&\mathsf{LB}_{\mathsf{ptt}}[\mathsf{C}](u, C, s, x, N, z, L) := \\
&\qquad \exists i < |L| \Big(u < L\#(2N) \wedge x = |N| \wedge z = |L| \\
&\qquad\qquad\qquad \wedge\, C \text{ is a } \mathsf{C}\text{-circuit of size} \leqslant s \to C(\lfloor [u]_i^{x,z}/2 \rfloor) \neq \mathit{bit}(0, [u]_i^{x,z}) \Big).
\end{aligned}$$

Note this formula holds trivially if u does not code a partial function (i.e. codes pairs $(a, 0)$ and $(a, 1)$ for some $a \in \{0,1\}^n$).

Definition 4.15. Let $s(x) \in \mathsf{PV}_1$ and recall a circuit of size $\leqslant s(n)$ is coded by a number of length $\leqslant c \cdot s(n) \cdot \log s(n)$ for a suitable constant $c \in \mathbb{N}$. Associate with u, C, x, N, z, L

length bounds $\ell \cdot (n+1), c \cdot s(n) \cdot \log s(n), |n|, n, |\ell|, \ell$ and define

$$\begin{aligned}&\mathsf{ptt}[\mathsf{C}, f, s(n), n, \ell] := \\ &\qquad [\![\mathsf{LB}_{\mathsf{ptt}}[\mathsf{C}](u, C, s(x), x, N, z, L)]\!]\ [f/u, n/x, 2^n - 1/N, \ell/z, 2^\ell - 1/L].\end{aligned}$$

Observe that the quantifier $\exists i < |L|$ translates to a disjunction $\bigvee_{i<\ell}$, so the formula $\mathsf{ptt}[\mathsf{C}, f, s(n), n, \ell]$ is of the form (48) as desired. It has size $(s(n) \cdot \ell \cdot n)^{O(1)}$.

4.6 Propositional naturalization of Smolensky's proof

In this section we formalize a variant of Razborov and Rudich's naturalization of Smolensky's $\mathsf{AC}^0[p]$-lower bound proof, "the most difficult example of naturalization we have encountered" [62, Section 3.2.1]. This will allow us to construct WF proofs of formulas $\mathsf{ptt}[\mathsf{AC}^0[p], f, n\#n, n, \ell]$ for all partial functions f satisfying a technically defined property which is in some sense large, constructive and useful (cf. Theorem 3.27)

To define our succinct natural property we need some notation. Let $f = f(x_1, \dots, x_n)$ be a partial Boolean function on n Boolean variables $x_1, \dots, x_n$, and let ρ be a restriction on these variables leaving n' variables unassigned (see Section 3.2). Then $f{\restriction}\rho := f(\rho(x_1), \dots, \rho(x_n))$ is a partial Boolean function on n' variables with domain $\subseteq \{0,1\}^{n'}$. By abuse of notation we shall denote these n' variables by $x_1, \dots, x_{n'}$. We shall be interested in partial functions which have sufficiently large domain in the sense that $f{\restriction}\rho$ is total for some ρ leaving polylogarithmically many variables unassigned.

Let $p, q \in \mathbb{N}$ be distinct primes, $\omega \neq 1$ a q-th root of unity in $\mathbb{F}_{p^{q-1}}$, and $\mathcal{P} \subseteq \mathbb{F}_{p^{q-1}}[x]$ a set of polynomials in the variables $x = (x_1, \dots, x_{n'})$. We define a $2^{n'} \times |\mathcal{P}|$ matrix $M^{n'}_{p,q}(\mathcal{P})$ over $\mathbb{F}_{p^{q-1}}$: its rows are indexed by tuples $a \in \{\omega, 1\}^{n'}$, its columns by $P(x) \in \mathcal{P}$, and the $(a, P(x))$-th entry is the value $P(a) \in \mathbb{F}_{p^{q-1}}$. Further, for a polynomial $P_0(x)$ we write

$$M^{n'}_{p,q}(P_0) := M^{n'}_{p,q}(P_0 \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'})$$

where $P_0 \cdot \mathcal{L}_{n'} := \{P_0 \cdot Q \mid Q \in \mathcal{L}_{n'}\}$ and $\mathcal{L}_{n'}$ denotes the set of low degree multilinear monomials (we agree that $\prod_{i\in\emptyset} x_i = 1$):

$$\mathcal{L}_{n'} := \Big\{ \textstyle\prod_{i\in T} x_i \mid T \subseteq [n'], |T| \leqslant n'/2 \Big\}. \tag{49}$$

For $g : \{0,1\}^{n'} \to \{0,1\}$, the proof of Theorem 3.12 shows how to explicitly write down a multilinear polynomial $p(x) \in \mathbb{F}_{p^{q-1}}[x]$ coinciding with the function $g(\frac{x_1-1}{\omega-1}, \dots, \frac{x_{n'}-1}{\omega-1})$, which maps $\{\omega, 1\}^{n'}$ into $\{0,1\}$. Then set

$$P[g](x) := (\omega - 1)p(x) + 1.$$

In particular, there is a polynomial time algorithm which given the truth table of g computes $P[g]$ explicitly as a list of coefficients. Roughly speaking, $P[g]$ restricted to $\{1, \omega\}^{n'}$ is the same as g but with $1, \omega$ playing the role of $0, 1$.

Theorem 4.16. *Let $p, q \in \mathbb{N}$ be distinct primes, $d \in \mathbb{N}$ and $0 < \epsilon < 1$ a rational. There are $c, n_0 \in \mathbb{N}$ and $circ(r, u, C, x, N, z, L, \tilde{f}, \tilde{x}, \tilde{N}) \in \mathsf{PV}$ and a polynomial time algorithm which given 2^k in unary and f, ρ such that for some $\ell, n \in \mathbb{N}$ and $m := \lfloor \log^{9d} n \rfloor$*

(i) f is a size ℓ partial Boolean function on n variables and ρ a restriction leaving $m+q$ variables unassigned,

(ii) $f{\restriction}\rho : \{0,1\}^{m+q} \to \{0,1\}$ is total and $M^{m+q}_{p,q}(P[f{\restriction}\rho])$ has rank at least $3/4 \cdot 2^m$,

(iii) $n \geqslant n_0$ and $k \geqslant c \cdot \log(\ell n)$,

computes an EF*-proof of*

$$\begin{aligned} &\textstyle\bigvee_{a<2^k} \text{“}circ(r, u, C, x, N, z, L, \tilde{f}, \cdot)(a) \neq \tilde{f}(a)\text{”}\ [\rho/r, f/u, n/x, 2^n - 1/N, \ell/z, 2^\ell - 1/L] \\ &\to \mathsf{ptt}\big[\mathsf{AC}^0_d[p], f, 2^{|n|^2}, n, \ell\big]; \end{aligned} \tag{50}$$

moreover, PV_1 *proves that $circ(r, u, C, x, N, z, L, \tilde{f}, \tilde{x}, \tilde{N})$ is a circuit of size $\leqslant |\tilde{N}|^\epsilon$.*

Observe that, choosing $k := c\log(\ell n)$ in (iii), the formula (50) has size polynomial in the size of f. Note that (ii) implies that the size of f is at least quasipolynomial in n.

Proof. Jeřábek [29, Theorem 4.3.18] showed that there exists a PV-function which PV_1-provably computes from a given matrix M over $\mathbb{F}_{p^{q-1}}$ a sequence of elementary matrices bringing M in reduced row echelon form. In particular, there exists a PV-symbol which PV_1-provably computes from M a subset (of indices) of rows which form a basis for the row space of M. Given f, ρ with (i) and (ii) one can compute in polynomial time (the list of coefficients of) the multilinear polynomial $P[f{\restriction}\rho]$ and the matrix $M^{m+q}_{p,q}(P[f{\restriction}\rho])$, explicitly as a tuple of elements of $\mathbb{F}_{p^{q-1}}$ (note (ii) implies that f has size $\ell \geqslant 2^{m+q}$). Hence, (i) and (ii) are expressible by Σ^b_0-formulas with variables u, r, x, z for f, ρ, n, ℓ.

We claim that $\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV})$ proves the Σ^b_0-formula

$$\begin{aligned} &\varphi(r, u, C, x, N, z, L) := \\ &\quad x \geqslant n_0 \to \big(u, r, x, z \text{ satisfy (i) and (ii)} \to \mathsf{LB}_{\mathsf{ptt}}[\mathsf{AC}^0_d[p]](u, C, x\#x, x, N, z, L)\big). \end{aligned} \tag{51}$$

We shall argue in S^1_2 that $\neg\varphi$ contradicts $\mathsf{sWPHP}(\mathsf{PV})$. For readability we write again f, ρ, n, ℓ instead of u, r, x, z. Assume the antecedent of φ and that C is a size $\leqslant n\#n$ $\mathsf{AC}^0_d[p]$-circuit computing $f{\restriction}\rho : 2^{m+q} \to 2$ where $m := \lfloor \log^{9d} n \rfloor$. Note that this implies $2^{m+q} \in Log$.

Now follow the proof of Theorem 3.9 and construct an arithmetical circuit P by replacing gates of C by low-degree polynomials: setting the parameters ℓ, ϵ in Theorem 3.9 appropriately, we get $P(x) = (f{\restriction}\rho)(x)$ with probability $1 - 1/2^{4+q}$ over $x < 2^{m+q}$ and that $P(x)$ has syntactic degree $O(|n|^{2d})$. As $2^{m+q} \in Log$, all probabilities can be counted precisely and stated by a Σ^b_0-formula. To define $P(x)$, thus $\mathsf{BB}(\Sigma^b_0)$ is sufficient and this scheme is available in S^1_2.

Applying the inputwise substitution $y = \frac{x-1}{\omega-1}$ to P and replacing its output z by $(\omega-1)z+1$, gives an arithmetical circuit P' of the same syntactic degree such that $P'(x) = P[f{\restriction}\rho](x)$ for many x, namely, for all x from some set $X \subseteq \{\omega, 1\}^{m+q}$ of cardinality $\mathit{Card}(X) \geqslant (1 - 1/2^{4+q}) \cdot 2^{m+q}$.

As mentioned above, we can compute in PV a subset $X' \subseteq \{\omega, 1\}^{m+q}$ of indices of rows forming a basis of the row space of $M^{m+q}_{p,q}(P[f{\restriction}\rho])$. By (ii), $\mathit{Card}(X') \geqslant 3/4 \cdot 2^m$, so $X'' := X \cap X'$ has cardinality $\mathit{Card}(X'') > 2/3 \cdot 2^m$. The rows with index in X'' are the same in the matrices $M^{m+q}_{p,q}(P[f{\restriction}\rho])$ and $M^{m+q}_{p,q}(P')$. The columns of $M^{m+q}_{p,q}(P')$ are indexed by polynomials of degree $\lfloor (m+q)/2 \rfloor + O(|n|^{2d}) < \lfloor \frac{m}{2} \rfloor + m^{1/3}$ assuming n_0 and hence n, m are large enough. Thus, every function $h : X'' \to \mathbb{F}_{p^{q-1}}$ can be written as a polynomial of at most this degree (considering rank, one sees that every such h, viewed as an X''-indexed column vector, is a linear combination of the columns of $M^{m+q}_{p,q}(P')$ restricted to X''). This contradicts the $\mathsf{sWPHP}(\mathsf{PV})$ (see the proof of Theorem 3.12).

This finishes the proof that $\mathsf{S}^1_2 + \mathsf{sWPHP}(\mathsf{PV})$ proves (51).

We now proceed similarly as in the proof of Theorem 4.10. Abbreviating the variables r, u, C, x, N, z, L of φ by $\bar{x}$ for readability, Lemma 4.9 gives a constant $c' \in \mathbb{N}$ and a function $\mathit{circ}(\bar{x}, \tilde{f}, \tilde{x}, \tilde{N}) \in \mathsf{PV}$ such that PV_1 proves

$$|\tilde{N}| \geqslant |\bar{x}|^{c'} \wedge \mathsf{LB}_{\mathsf{tt}}(\tilde{f}, \mathit{circ}(\bar{x}, \tilde{f}, \tilde{x}, \tilde{N}), |\tilde{N}|^{\epsilon}, \tilde{x}, \tilde{N}) \;\to\; \varphi(\bar{x}). \tag{52}$$

As in Theorem 4.10 we find such circ satisfying the "moreover" part of the theorem.

We now describe the polynomial time algorithm. On input $(2^k, f, \rho)$ satisfying (i)-(iii) for certain n, ℓ, it first runs the algorithm from Theorem 4.1 to get an EF-proof of the translation of (52) for the following association of length bounds to the variables. With the variables u, C, x, N, z, L associate $\ell \cdot (n+1), 2^{|n|^3}, |n|, n, |\ell|, \ell$, and with r some length bound $n^{O(1)}$ suitable to hold an encoding of the restriction ρ; note length $2^{|n|^3}$ is enough to code a circuit of size $\leqslant n\#n$. With the variables $\tilde{f}, \tilde{x}, \tilde{N}$ associate $2^k, |k|, 2^k$.

The time needed to construct this EF-proof is polynomial in these length bounds, so polynomial in the length of the input (note $|f| \geqslant \ell \geqslant 2^{\log^{9d} n}$).

Next the algorithm applies the substitution

$$[k/\tilde{x}, 2^{2^k} - 1/\tilde{N}, \rho/r, f/u, n/x, 2^n - 1/N, \ell/z, 2^\ell - 1/L]$$

to the proof. If $c \in \mathbb{N}$ in (iii) is large enough, then $[\![|\tilde{N}| \geqslant |\bar{x}|^{c'}]\!]$ as well as the antecedent of $[\![\varphi]\!]$ become tautologies in auxiliary variables only, so can be eliminated (see the proof of Theorem 4.10). This yields an EF-proof of the formula

$$[\![\mathsf{LB}_{\mathsf{tt}}(\tilde{f}, \mathit{circ}(\bar{x}, \tilde{f}, \tilde{x}, \tilde{N}), |\tilde{N}|^{\epsilon}, \tilde{x}, \tilde{N})]\!] \;\to\; [\![\mathsf{LB}_{\mathsf{ptt}}[\mathsf{AC}^0_d[p]](u, C, |N|\#|N|, x, N, z)]\!]$$

with the above substitution. This is (50). □

Note that we obtained EF proofs in Theorem 4.16 by an application of Cook's Simulation Theorem 4.1 which is enabled by expressing $\mathsf{AC}^0[p]$ lower bounds directly by a Σ_0^b formula (instead of witnessing the Σ_1^b formula $\mathsf{LB}[\mathsf{C},\mathsf{Q}]$ as in Section 4.4). Applying Jeřábek's Simulation Theorem 4.2 we get, as a corollary to the previous proof, unconditional short WF proofs of $\mathsf{AC}^0[p]$ lower bounds:

Corollary 4.17. *Let $p, q \in \mathbb{N}$ be distinct primes and $d \in \mathbb{N}$. There are $n_0 \in \mathbb{N}$ and a polynomial time algorithm which given f, ρ such that for some $\ell, n \in \mathbb{N}$ and $m := \lfloor \log^{9d} n \rfloor$*

(i) f is a size ℓ partial Boolean function on n variables and ρ a restriction leaving $m+q$ variables unset,

(ii) $f{\restriction}\rho : \{0,1\}^{m+q} \to \{0,1\}$ is total and $M_{p,q}^{m+q}(P[f{\restriction}\rho])$ has rank at least $3/4 \cdot 2^m$,

(iii) $n \geqslant n_0$,

computes a WF-proof of $\mathsf{ptt}\left[\mathsf{AC}_d^0[p], f, 2^{|n|^2}, n, \ell\right]$.

Proof. As seen in the previous proof $\mathsf{S}_2^1 + \mathsf{sWPHP}(\mathsf{PV})$ proves the Σ_0^b-formula φ. By Theorem 4.2 we can produce a WF-proof of $[\![\varphi]\!]$ with length bounds as in the previous proof. As there, applying an appropriate substitution allows to eliminate the antecedent, leaving a proof of $\mathsf{ptt}[\mathsf{AC}_d^0[p], f, 2^{|n|^2}, n, \ell]$. □

Remark 4.18. The argument ρ to the algorithms in Theorem 4.16 and Corollary 4.17 can be omitted by slightly increasing the running time: given f one can compute in time $n^{O(m)}$ some ρ such that (i) and (ii) hold, provided there exists one. In particular, fixing $k := \lceil c \cdot \log(\ell n) \rceil$ in Theorem 4.16, we get quasipolynomial time algorithms with single input f.

Corollary 4.19. *Let $p, q \in \mathbb{N}$ be distinct primes and $d \in \mathbb{N}$. There are $n_0 \in \mathbb{N}$ and a quasipolynomial time algorithm that given $n \geqslant n_0$ in unary computes a WF-proof of*

$$\mathsf{ptt}\left[\mathsf{AC}_d^0[p], f, 2^{|n|^2}, n, 2^{m+q}\right],$$

where f is the MOD_q function restricted to $\{0,1\}^{m+q} \times \{0\}^{n-m-q}$ with $m := \lfloor \log^{9d} n \rfloor$.

Proof. Let ρ be the restriction on the variables $x_1, \ldots, x_n$ that leaves $x_1, \ldots, x_{m+q}$ unassigned and maps $x_{m+q+1}, \ldots, x_n$ to 0. Then $f = f{\restriction}\rho$ equals MOD_q on $\{0,1\}^{m+q}$.

For $i < q$ let $\bar{b}_i \in \{\omega, 1\}^q$ be a tuple with $q - i$ many ω's and i many 1's. Since $f{\restriction}\rho = MOD_q$ on $\{0,1\}^{m+q}$, we have for every $a = (a_1, \ldots, a_m) \in \{\omega, 1\}^m$

$$\prod_{i \in [m]} a_i = \sum_{i<q} \omega^i \cdot \frac{P[f{\restriction}\rho](a_1, \ldots, a_m, \bar{b}_i) - 1}{(\omega - 1)}. \tag{53}$$

Observe that $M_{p,q}^{m+q}(P[f{\restriction}\rho])$ and $M_{p,q}^{m+q}(P[f{\restriction}\rho] - 1)$ have the same rank. This is because $(P[f{\restriction}\rho] - 1) \cdot Q$ is a linear combination of $P[f{\restriction}\rho] \cdot Q$ and Q, for $Q \in \mathcal{L}_{m+q}$, and similarly, $P[f{\restriction}\rho]\cdot Q$ is a linear combination of $(P[f{\restriction}\rho]-1)\cdot Q$ and Q. It suffices to show that the rank of $M^{m+q}_{p,q}(P[f{\restriction}\rho]-1)$ is large. Then our claim follows from Corollary 4.17.

By Remark 3.13 the matrix $M^{m+q}_{p,q}(\prod_{i\in[m]} x_i)$ has rank at least 2^m. Choose 2^m many linearly independent columns, say, with indices $I_0\dot\cup I_1$ where $I_0\subseteq(\prod_{i\in[m]} x_i)\cdot\mathcal{L}_m$ and $I_1\subseteq\mathcal{L}_m\setminus((\prod_{i\in[m]} x_i)\cdot\mathcal{L}_m)$. We have $|I_0\dot\cup I_1|=2^m$.

Now, consider the columns of $M^{m+q}_{p,q}(P[f{\restriction}\rho]-1)$ indexed by Q and $(P[f{\restriction}\rho]-1)\cdot Q$ where $Q\in\mathcal{L}_m\subseteq\mathcal{L}_{m+q}$ (see (49)). By (53), for every $a\in\{\omega,1\}^m$, there is a linear combination of rows of $M^{m+q}_{p,q}(P[f{\restriction}\rho]-1)$ resulting in a row with values $(\prod_{i\in[m]} a_i)\cdot Q(a)$ on the positions indexed by $(P[f{\restriction}\rho]-1)\cdot Q$, and with values

$$\sum_{i<q}\frac{\omega^i}{\omega-1}\cdot Q(a)=\frac{\omega^q-1}{(\omega-1)^2}\cdot Q(a)=0 \tag{54}$$

on the positions indexed by Q. Adding these 2^m many rows to $M^{m+q}_{p,q}(P[f{\restriction}\rho]-1)$ does not increase the rank. Hence it suffices to show that the resulting matrix M has rank at least 2^m.

The I_1-columns of M prolong the I_1-columns of $M^{m+q}_{p,q}(\prod_{i\in[m]} x_i)$. By (54) these columns have 0 in the newly added 2^m many rows. The I_0-columns of M read like those of $M^{m+q}_{p,q}(\prod_{i\in[m]} x_i)$ in these newly added rows. It follows that the $(I_0\dot\cup I_1)$-columns of M are linearly independent. □

Recalling the motivation from learning, we finally observe for $q=2$ that there are many partial functions satisfying (ii) in Theorem 4.16.

Proposition 4.20. *Let $p>2$ be prime and $n'\in\mathbb{N}$. Then the matrix $M^{n'}_{p,2}(P[g])$ over $\mathbb{F}_p$ has rank at least $3/4\cdot 2^{n'}$ for at least half of all functions $g:\{0,1\}^{n'}\to\{0,1\}$.*

Proof. Let us call a polynomial over $\mathbb{F}_p$ with variables $x=(x_1,\dots,x_{n'})$ *representing* if it maps $\{-1,1\}^{n'}$ into $\{-1,1\}$. Obviously, representing polynomials are closed under multiplication. We claim that for every representing $P=P(x)$ at least one of the matrices $M^{n'}_{p,2}(P)$ or $M^{n'}_{p,2}(P\cdot\prod_{i\in[n']} x_i)$ has rank $\geqslant 3/4\cdot 2^{n'}$.

We check that the claim implies the proposition. Map $g:\{-1,1\}^{n'}\to\{-1,1\}$ to a pair (b,g') with $b\in\{-1,1\}$ and $g':\{-1,1\}^{n'}\to\{-1,1\}$ defined as follows: if $M^{n'}_{p,2}(P[g])$ has rank $\geqslant 3/4\cdot 2^{n'}$, then $b:=0$ and $g':=g$; otherwise $b:=1$ and g' is such that $P[g']$ equals $P[g]\cdot\prod_{i\in[n']} x_i$ on $\{-1,1\}^{n'}$. This map is injective, and, by the claim, every value (b,g') is such that $M^{n'}_{p,2}(P[g'])$ has rank $\geqslant 3/4\cdot 2^{n'}$

We are left to prove the claim. For a set $\mathcal{P}$ of representing polynomials, let $V(\mathcal{P})$ denote the vector space spanned by the columns of $M^{n'}_{p,2}(\mathcal{P})$. Observe that for a representing polynomial $P(x)\in\mathbb{F}_p[x]$ we have

$$\dim\big(V(\mathcal{P})\big)=\dim\big(V(P\cdot\mathcal{P})\big). \tag{55}$$

Indeed, $M^{n'}_{p,2}(P \cdot \mathcal{P})$ is obtained from $M^{n'}_{p,2}(\mathcal{P})$ by multiplying every row with a non-zero scalar, namely $P(a) \in \{-1, 1\}$ for the row with index $a \in \{-1, 1\}^{n'}$, and this preserves the rank.

For the set of monomials $\mathcal{M}_{n'} := \left\{ \prod_{i \in T} x_i \mid T \subseteq [n'] \right\}$ we have

$$\dim V(\mathcal{M}_{n'}) = 2^{n'}. \tag{56}$$

because every function from $\{-1, 1\}^{n'}$ to $\{-1, 1\}$ is computed by a multilinear representing polynomial. Hence, the $2^{n'}$ columns of $M^{n'}_{p,2}(\mathcal{M}_{n'})$ are linearly independent, so we have

$$\dim V(\mathcal{L}_{n'}) \geqslant |\mathcal{L}_{n'}| \geqslant 2^{n'}/2. \tag{57}$$

Further, the monomials in $\mathcal{M}_{n'}$ and the polynomials in $(\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}$ compute the same functions from $\{-1, 1\}^{n'}$ to $\{-1, 1\}$. Hence,

$$V(\mathcal{M}_{n'}) = V((\textstyle\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}). \tag{58}$$

We aim to show that the dimension of $V(P \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'})$ or $V((P \cdot \prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'})$ is $\geqslant 3/4 \cdot 2^{n'}$. Using (55) and noting $P^2 = 1$ we get

$$\begin{aligned}
&\dim V((P \cdot \textstyle\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}) - \dim V(\mathcal{L}_{n'}) \\
&= \dim V((\textstyle\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup P \cdot \mathcal{L}_{n'}) - \dim V(P \cdot \mathcal{L}_{n'}) \\
&= \dim \Big(V((\textstyle\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup P \cdot \mathcal{L}_{n'}) \;/\; V(P \cdot \mathcal{L}_{n'}) \Big) \\
&\geqslant \dim \Big(V((\textstyle\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup P \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}) \;/\; V(P \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}) \Big) \\
&= \dim V(\mathcal{M}_{n'}) - \dim V(P \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}),
\end{aligned}$$

where the last equality uses (58). Rearranging gives

$$\begin{aligned}
&\dim V((P \cdot \textstyle\prod_{i \in [n']} x_i) \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}) + \dim V(P \cdot \mathcal{L}_{n'} \cup \mathcal{L}_{n'}) \\
&\quad \geqslant \dim V(\mathcal{M}_{n'}) + \dim V(\mathcal{L}_{n'}).
\end{aligned}$$

By (56) and (57) the r.h.s. is at least $3/2 \cdot 2^{n'}$. This implies our claim. □

5 Questions

In the Introduction we said that a large part of contemporary complexity theory can be formalized in PV_1 or slight extensions of it. Table 1 lists some such results.

As announced in the Introduction we believe the given proofs of Theorems 1.1, 1.2 and 1.3 show that $\mathsf{sWPHP}(\mathsf{PV})$ allows for a *natural* formalization of these circuit lower bounds. Remarks 3.8, 3.15 and 3.20 detail the role of $\mathsf{sWPHP}(\mathsf{PV})$.

Theory	Theorem	Reference
PV_1	Cook-Levin Theorem	folklore
	PCP Theorem	[53]
	Hardness amplification	[29]
APC_1	AC^0 lower bounds	Section 3.2
	$\mathsf{AC}^0[p]$ lower bounds (with $2^{\log^{O(1)} n} \in Log$)	Section 3.3
	Monotone circuit lower bounds	Section 3.4
HARD^A	Nisan-Wigderson's derandomization	[28]
	Impagliazzo-Wigderson's derandomization	[29]
	Goldreich-Levin theorem	[24]
APC_1^+	Natural proof barrier	Section 3.6
APC_2	Graph isomorphism in coAM	[32]
$\mathsf{APC}_2^{\oplus_p P}$	Toda's theorem	[12]

Table 1: A list of formalizations.

It is natural to ask whether sWPHP can be avoided, that is, whether Theorems 1.1, 1.2 and 1.3 hold for PV_1 instead of APC_1. A positive answer for Theorem 1.2 could be interesting as this seems to require some new insights and a new proof. For Theorem 1.3 one might suspect a positive answer with a similar proof, vaguely because the circuits witnessing the approximate counting are particularly simple and transparent. We have, however, not been able to give such a proof.

On the other hand, proving independence from PV_1 is presumably very difficult. An already challenging open problem is to show that the theory V^0 corresponding to AC^0-reasoning [23] does not prove $\mathsf{LB}(\mathsf{AC}^0_d, \mathsf{PARITY})$ for $s = n^k$, or, more precisely, a suitable second-order formulation of this formula (see e.g. [52]).

A weaker task than finding PV_1-proofs is to derandomize the witnessing functions derived from particular APC_1-proofs of circuit lower bounds. For instance and more precisely: is there a deterministic polynomial time Turing machine satisfying Corollary 3.22?

Concerning Theorem 1.2 we also leave open the question whether polynomial lower bounds can be proved assuming only $n \in Log$, that is: does APC_1 prove $\mathsf{LB}[\mathsf{AC}^0_d[p], \mathsf{MOD}_q]$ for $s = n^k$ and large enough $n \in Log$?

On the propositional side the obvious question is whether our conditional upper bounds can be made unconditional. For instance and more precisely: are there short EF-proofs of $\mathsf{lb}_w[\mathsf{AC}^0_d, \mathsf{PARITY}, n^k]$ for some w? It would already be interesting to find quasipolynomial size WF-proofs. An interesting route to achieve this would be to witness $\mathsf{LB}(\mathsf{AC}^0_d, \mathsf{PARITY})$ for $s = n^k$ by a deterministic $w \in \mathsf{PV}$ provably in APC_1. This in turn could be achievable by derandomizing the Switching Lemma formally in APC_1 (cf. [67]). A positive answer would be interesting not just for the lb_w-formulation but any succinct formulation of AC^0_d-lower bounds, for example, the ptt-formulation. Corollary 4.19 achieves WF-proofs of

$\mathsf{AC}^0_d[p]$ lower bounds for MOD_q by formalizing the naturalization of this lower bound.

It is possible to approach similarly the naturalization of the AC^0 lower bounds based on the Switching lemma (see [62, Section 3.1]). Following the proof of Theorem 1.1, one can show how to generate a set of polynomially many restrictions such that every AC^0-circuit is collapsed by some of them. The set is generated by a probabilistic algorithm or, alternatively, using a Nisan-Wigderson generator based on a hard function. A candidate succinct natural property of partial functions f would thus require f to be non-constant after any of the generated restrictions. However, it is not clear to us if this property is large in some sense. Moreover, WF-proofs of $\mathsf{ptt}[\mathsf{AC}^0, f, n^k, n, \ell]$ (say, for f a partial PARITY) do not seem to follow since the property depends on the hard function of the Nisan-Wigderson generator.

Acknowledgements

We thank the anonymous referee for several excellent reviews. The reviews pointed out several mistakes and many places where we could clarify or otherwise improve this long and quite technical text. We thank Igor Carboni Oliveira for helpful discussions. Most of the work was done while both authors were supported by the Austrian Science Fund (FWF) under project number P 28699. Some initial parts have been undertaken while the second author was supported by the NSERC. The revisions following the mentioned reviews have been done with the first author currently supported by the European Research Council (ERC) under the European Unions Horizon 2020 research programme (grant agreement ERC-2014-CoG 648276 AUTAR) and with the second author currently supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2014)/ERC Grant Agreement no. 615075.

References

[1] Ajtai, M.; Σ^1_1 *formulae on finite structures*, Annals of Pure and Applied Logic, 24 (1): 1-48, 1983.

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

[3] Arora S., Barak B.; *Computational Complexity: A Modern Approach*, Cambridge University Press, 2009.

[4] Atserias A., Thapen N.; *The ordering principle in a fragment of approximate counting*, ACM Transactions on Computational Logic, 15:4, article 29, 2014.

[5] Boppana R.B., Sipser M.; *The complexity of finite functions*, van Leeuwen J. (ed.), Handbook of theoretical computer science (vol. A), pp. 758-804, Elsevier, 1990.

[6] Buss S.R.; *Bounded Arithmetic*, Bibliopolis, Naples, 1986.

[7] Buss S.R.; *Bounded arithmetic and propositional proof complexity*, in H. Schwichtenberg (ed.), Logic of Computation, Springer, pp. 67-122, 1997.

[8] Buss S.R.; *Bounded arithmetic, cryptography and complexity.* Theoria 63: 147-167, 1997.

[9] Buss S.R.; *First-order theory of arithmetic*, in Buss S.R. (ed), Handbook Proof Theory, Elsevier, pp. 79-147, 1998.

[10] Buss S.R. et al.; *Weak formal systems and connections to computational complexity.* Student written lecture notes for topics course Math 271 in Berkeley, Winter/Spring 1988. Available at `http://www.math.ucsd.edu/~sbuss/ResearchWeb/weakformaltopics/WeakFormalTopics_OCR.pdf`.

[11] Buss S.R., Kołodziejczyk L.A., Thapen N.; *Fragments of approximate counting*, Journal of Symbolic Logic, 79(2): 496-525, 2014.

[12] Buss S.R., Kołodziejczyk L.A., Zdanowski K.; *Collapsing modular counting in bounded arithmetic and constant depth propositional proofs*, Transactions of the AMS, 367: 7517-7563, 2015.

[13] Bydzovsky J., Müller M.; *Polynomial time ultrapowers and the consistency of circuit lower bounds*, Archive for Mathematical Logic, to appear.

[14] Bydzovsky J., Krajíček J., Oliveira C.I.; *Consistency of circuit lower bounds with bounded theories*, preprint arXiv:1905.12935 [cs.CC], 2019.

[15] Carmosino M., Impagliazzo R., Kabanets V., Kolokolova A.; *Learning algorithms from natural proofs*, Proceedings of the 2016 Conference on Computational Complexity (CCC), 10:1-10:24, 2016.

[16] Chapman B., Williams R.; *The Circuit-input game, natural proofs, and testing circuits with data*, Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS), pp. 263-270, 2015.

[17] Chow T.Y.; *Almost-natural proofs*, Journal of Computer and System Sciences, 77(4): 728-737, 2011.

[18] Chow T.Y.; *What is a natural proof?*, Notices of the AMS, 58(11): 1586–1587, 2011.

[19] 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.

[20] Cook S.A.; *Feasibly constructive proofs and the propositional calculus*, Proceedings of the 7th Annual ACM Symposium on Theory of Computing (STOC), ACM Press, pp. 83-97, 1975.

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

[22] Cook S.A. and Mitchell D.; *Finding hard instances of the satisfiability problem: a survey*, DIMACS Series in Discrete Mathematics and Theoretical Computer Science 35: 1-17, 1997.

[23] Cook S.A., Nguyen P.; *Logical Foundations of Proof Complexity*, Cambridge University Press, 2010.

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

[25] Furst M., Saxe J. B., Sipser M.; *Parity, circuits, and the polynnomial-time hierarchy*; Mathematical systems Theory, 17: 13-27, 1984.

[26] Filmus Y., Pitassi T., Santhanam R.; *Exponential lower bounds for AC^0-Frege imply superpolynomial Frege lower bounds*, Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP), pp. 618-629, 2011.

[27] Håstad J.; *Computational limitations for small depth circuits*, PhD thesis, M.I.T. press, 1986.

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

[29] Jeřábek E.; *Weak pigeonhole principle, and randomized computation*, Ph.D. thesis, Faculty of Mathematics and Physics, Charles University, Prague, 2005.

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

[31] Jeřábek E.; *On independence of variants of the weak pigeonhole principle*, Journal of Logic and Computation, 17: 587-604, 2007.

[32] Jeřábek E.; *Approximate counting by hashing in bounded arithmetic*, Journal of Symbolic Logic, 74:829-860, 2009.

[33] Kołodziejczyk L., Thapen N.; *The polynomial and linear hierarchies in models where the weak pigeonhole principle fails*, Journal of Symbolic Logic 73:2, pp. 578-592, 2008.

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

[35] Krajíček J.; *Bounded Interpolation theorems, lower bounds for proof systems, and independence results for bounded arithmetic*, Journal of Symbolic Logic, 62(2): 457-486, 1997.

[36] Krajíček J.; *Extensions of models of PV*, in: Logic Colloquium'95, Eds. J.A.Makowsky and E.V.Ravve, ASL/Springer Series, Lecture Notes in Logic 11, pp. 104-114, 1998.

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

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

[39] Krajíček J.; *Forcing with random variables and proof complexity*, London Mathematical Society Lecture Note Series, No.382, Cambridge University Press, 2011.

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

[41] Krajíček J., Pudlák P.; *Propositional provability and models of weak arithmetic*, in: Computer Science Logic (CSL), Börger E., Kleine Büning H., Richter M. M. (eds.), LNCS 440, Springer, pp. 193-210, 1989.

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

[43] Lipton R.J., Young N.E.; *Simple strategies for large zero-sum games with applications to complexity theory*, Proceedings of the 26th Annual ACM Symposium on Theory of Computing (STOC), pp. 734-740, 1994.

[44] Maciel A., Pitassi T., and Woods A.; *A new proof of the weak pigeonhole principle*, Journal of Computer Systems Sciences, 64: 843-872, 2002.

[45] Maly J., Müller M.; *A remark on pseudo proof systems and hard instances of the satisfiability problem.* Mathematical Logic Quarterly, 64 (6): 418-428, 2018.

[46] Müller M., Pich J.; *Provability of weak circuit lower bounds*, unpublished, available from the second author's homepage.

[47] Müller S.; *Polylogarithmic cuts in models of V0*, Logical Methods in Computer Science, 9(1), 2013.

[48] Nisan N., Wigderson A.; *Hardness vs. randomness*, Journal of Computer and System Sciences, 49 (2): 149-167, 1994.

[49] Oliveira C.I., Santhanam R.; *Hardness magnification for natural problems*, Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 65-76, 2018.

[50] Oliveira C.I., Pich J., Santhanam R.; *Hardness magnification near state-of-the-art lower bounds*, Proceedings of the 34th Computational Complexity Conference (CCC), to appear, 2019.

[51] Parikh R.; *Existence and feasibility in arithmetic*, Journal of Symbolic Logic, 36: 494-508, 1971.

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

[53] 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.

[54] Pich J.; *Complexity Theory in Feasible Mathematics*, Ph.D. thesis, Faculty of Mathematics and Physics, Charles University, Prague, 2014.

[55] Raz R.; *Resolution lower bounds for the weak pigeonhole principle*, Journal of the ACM, 51 (2): 115-138, 2004.

[56] Razborov A.A.; *Lower bounds on the monotone complexity of some Boolean functions*, Doklady Akademii Nauk SSSR, 281 (4), pp. 798-801 (in Russian), 1985.

[57] Razborov A.A.; *Lower bounds on the size of bounded depth networks over a complete basis with logical addition* (in Russian), Matematicheskie Zametki, 41(4): 598-607, 1987.

[58] Razborov A.A.; *On provably disjoint NP-pairs*, Basic Research in Computer Science BRICS RS-94-36, 1994.

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

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

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

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

[63] Smolensky R.; *Algebraic methods in the theory of lower bounds for Boolean circuit complexity*, Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC), pp. 77-82, 1987.

[64] Thapen N.; *A model-theoretic characterization of the weak pigeonhole principle*, Annals of Pure and Applied Logic, 118 (1-2): 75-195, 2002.

[65] Thapen N.; *Structures interpretable in models of bounded arithmetic*, Annals of Pure and Applied Logic, 136 (3): 247-266, 2005.

[66] Thiele R.; *Hilbert's twenty-fourth problem.* American Mathematical Monthly, January 2003.

[67] Trevisan L., Xue T.; *A derandomized switching lemma and an improved derandomization of* AC^0, Proceedings of the 28th Conference of Computational Complexity (CCC), pp. 242-247, 2013.

[68] Tseitin G.S.; *On the complexity of derivation in propositional calculus*, in Slisenko A.O. (ed.), Studies in Constructive Mathematics and Mathematical Logic, Part II, pp. 115-125, 1970.