Some subsystems of constant-depth Frege with parity

Michal Garlík* and Leszek Aleksander Kołodziejczyk†

Institute of Mathematics
University of Warsaw
Banacha 2
02-097 Warszawa, Poland
email: mgarlik,lak@mimuw.edu.pl

March 22, 2017

Abstract

We consider three relatively strong families of subsystems of $\mathrm{AC}^0[2]$-Frege for which exponential lower bounds on proof size are known. In order of increasing strength, these subsystems are: AC^0-Frege with parity axioms and the treelike and daglike versions of systems introduced by Krajíček which we call $\mathrm{PK}^c_d(\oplus)$. In a $\mathrm{PK}^c_d(\oplus)$-proof, lines are cedents in which all formulas have depth at most d, parity connectives can only appear as the outermost connectives in formulas, and all but c formulas contain no parity connective at all.

We give simple arguments proving that $\mathrm{AC}^0[2]$-Frege is exponentially stronger than daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, which is exponentially stronger than treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$. On the other hand, we point out that the best available technique for comparing the performance of such systems on De Morgan formulas, due to Impagliazzo and Segerlind, only leads to superpolynomial separations. In fact, we prove that treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$

*Partially supported by Polish National Science Centre grant no. 2013/09/B/ST1/04390 and by ERC Advanced Grant 339691 (FEALORA). Most of this work was carried out during the tenure of an ERCIM 'Alain Bensoussan' Fellowship Programme.

†Partially supported by Polish National Science Centre grant no. 2013/09/B/ST1/04390.

is quasipolynomially but not polynomially equivalent to AC^0-Frege with parity axioms. This leads us to ask the question whether any of our systems is quasipolynomially equivalent to $AC^0[2]$-Frege on De Morgan formulas; a positive answer would imply an exponential lower bound for $AC^0[2]$-Frege.

We also study Itsykson and Sokolov's system Res-Lin. We prove that an extension of treelike Res-Lin is polynomially simulated by a system related to daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, and obtain an exponential lower bound for this system.

1 Introduction

The work presented in this paper is inspired by the following long-standing open problem in propositional proof complexity:

Prove superpolynomial or better lower bounds
on proof size for $AC^0[2]$-Frege.

Here $AC^0[2]$-Frege systems are proof systems in which lines are constant-depth formulas in the language of $\neg$, unbounded fan-in $\wedge$ and $\vee$, and an unbounded fan-in parity connective $\oplus$. The survey paper [6] considers this to be one of the two main challenges currently facing *Cook's programme* of approaching the NP = coNP problem via lower bound proofs for increasingly strong proof systems.

Our point of departure is the observation that there are relatively strong subsystems of $AC^0[2]$-Frege, combining the full power of AC^0-Frege with some ability to reason about parity, for which good lower bounds are known. The most familiar of these systems is AC^0-Frege with parity axioms, which requires exponential size to prove the counting mod 3 principle Count_3 [9] and the pigeonhole principle PHP [4]. Two others are the treelike and daglike versions of a system studied by Krajíček [11] which we call $\mathrm{PK}^c_d(\oplus)$. The idea of $\mathrm{PK}^c_d(\oplus)$ is that all formulas are required to have depth at most d, and even though parity connectives are allowed, the parts of the syntactic trees that are between the root and a parity must have constant size, as specified by c. Treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ requires exponential size to prove PHP [11], and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ has no short proofs of Count_3 ([11] combined with [7]).

How plausible is it that lower bounds for systems such as those "already imply" lower bounds for $AC^0[2]$-Frege?

For instance, how plausible is the following scenario? Take a suitable model $(M, \oplus^M)$ of the (relativized) bounded arithmetic theory corresponding to $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ such that M violates some combinatorial principle, for instance, Count_3. By [20], M satisfies the weak pigeonhole principle for bounded formulas in the language with the usual quantifiers $\exists, \forall$ but without the parity quantifier $\oplus$. Prove that M has some model-theoretic property that makes either $(M, \oplus^M)$ or M expanded by a different interpretation of $\oplus$ satisfy WPHP in the language with $\oplus$. By doing this, you will have proved that the principle violated in M is independent of bounded arithmetic with the $\oplus$ quantifier, as that theory is provably equivalent to a subtheory of the theory for $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ extended by certain instances of WPHP for functions whose definitions involve $\oplus$ [10]. By doing this in a way that is sufficiently uniform in M, you will have actually obtained the elusive lower bound for $\mathrm{AC}^0[2]$-Frege.

In the authors' opinion, the scenario is not all that likely, and has a "pigs can fly" feel. However, what concrete arguments can we give against it? The ideal argument would be to exhibit a uniform family of tautologies in the language of $\neg, \wedge, \vee$ separating $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ from $\mathrm{AC}^0[2]$-Frege: if M violates the separating principle, it cannot be expanded so as to satisfy WPHP in the language with $\oplus$. (On the other hand, if no such family of tautologies exists, then $\mathrm{AC}^0[2]$-Frege has no short proofs of Count_3 or PHP.)

At first glance, it might seem that we have the separation needed to rule out the scenario above. In [15], Impagliazzo and Segerlind proved a separation of AC^0-Frege with parity axioms from $\mathrm{AC}^0[2]$-Frege. Moreover, we show that the technique they use can be adapted to prove that the strength of $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, as a system for refuting CNFs, is strictly intermediate between AC^0-Frege with parity axioms and $\mathrm{AC}^0[2]$-Frege.

There is a catch, however. The method of [15] can only prove barely superpolynomial separations, and that is not quite enough. In the context of constant-depth systems, the right notion of simulation seems to be at least *quasipolynomial* rather than polynomial: the translation from bounded arithmetic to constant-depth proofs is quasipolynomial, and in general the area abounds in results on quasipolynomial-size provability and/or simulation that are either open or false in the polynomial-size case, e.g. [5, 18, 1, 10].

The main technical result of our paper is one more argument in favour of demanding superquasipolynomial rather than merely superpolynomial separations: inspired by another paper by Impagliazzo and Segerlind, we prove that AC^0-Frege with parity axioms and treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ are quasipolynomially equivalent, even though they are *not* polynomially equivalent.

While the possibility that this quasipolynomial equivalence extends all the way to $\mathrm{AC}^0[2]$-Frege seems farfetched, it is nevertheless not disproved. Thus, the main "ideological message" of our paper is that the following question mentioned in [15] deserves more attention than it has received: can the superpolynomial separation of AC^0-Frege with parity axioms from $\mathrm{AC}^0[2]$-Frege be improved—to superquasipolynomial, at the very least? The follow-up question, of course, would be: can such an improved separation be extended to systems such as $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$?

The paper is organized as follows. We introduce the necessary definitions and background in Section 2. In Section 3, we verify that $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ is complete and that the treelike version polynomially simulates AC^0-Frege with parity axioms. The proof of the quasipolynomial simulation of treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ by AC^0-Frege with parity axioms takes up Sections 4–7. The structure of that proof is described at the beginning of Section 4.

In Section 8, we provide exponential separations between treelike and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ and $\mathrm{AC}^0[2]$-Frege as refutation systems for families of formulas with $\oplus$; we sketch a proof of the superpolynomial separations without $\oplus$ in Section 9.

In Section 10, we consider the loosely related topic of a refutation system introduced by Itsykson and Sokolov [17] in which proof lines are disjunctions of parities of literals. [17] proves a lower bound on the treelike version of this system. We point out that even a generalization of the treelike system in which the parities can have arbitrary constant-depth formulas as inputs is quasipolynomially simulated by daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, and therefore unable to give short proofs of Count_3.

2 Preliminaries

Propositional formulas are built up from Boolean variables x and their negations $\overline{x}$ using the unbounded fan-in connectives $\bigvee, \bigwedge, \oplus^0, \oplus^1$. The input to an unbounded fan-in connective is a sequence of formulas. We allow the input to be the empty sequence, $\emptyset$, in which case we will write $\bot$ and $\top$ for $\bigvee \emptyset$ and $\bigwedge \emptyset$, respectively. The parity connective $\oplus^b$, for $b \in \{0,1\}$, is interpreted to be true if the number of its true inputs is congruent to b modulo 2. The negation operator is extended to all formulas by defining $\overline{\overline{x}}$ to be x for a variable x, and inductively defining $\overline{\bigvee_{i\in I} \varphi_i}$, $\overline{\bigwedge_{i\in I} \varphi_i}$ and $\overline{\oplus^b_{i\in I}\varphi_i}$ to be $\bigwedge_{i\in I} \overline{\varphi_i}$, $\bigvee_{i\in I} \overline{\varphi_i}$ and $\oplus^{1-b}_{i\in I}\varphi_i$, respectively. The *depth*, $\mathrm{dp}(\varphi)$, of a formula φ, is the maximum number of alternating blocks of connectives along any

branch in φ (viewed as a tree), except that we consider $\bot$ and $\top$ to have depth 0 and extend that to formulas containing $\bot$ and $\top$.

The propositional proof systems we consider are Tait-style systems, i.e., the lines in a proof are *cedents*. Our convention is that a cedent is a sequence of formulas. We use capital Greek letters $\Gamma, \Delta, \ldots$ as names for both cedents and inputs to the unbounded fan-in connectives. The intended meaning of a cedent Γ is $\bigvee\Gamma$.

The logical axioms are:

$$\oplus^0\emptyset \quad \text{and} \quad x, \overline{x}$$

for a propositional variable x.

The inference rules are:

$$\frac{\Gamma}{\Gamma, \Delta}\ \text{Weakening} \qquad \frac{\Gamma, \Delta, \Lambda, \Psi}{\Gamma, \Lambda, \Delta, \Psi}\ \text{Exchange}$$

$$\frac{\Gamma, \Delta}{\Gamma, \bigvee\Delta}\ \text{OR} \qquad \frac{\Gamma, \varphi_i \quad \text{for all } i \in I}{\Gamma, \bigwedge_{i\in I}\varphi_i}\ \text{AND}$$

$$\frac{\Gamma, \varphi, \varphi}{\Gamma, \varphi}\ \text{Contraction} \qquad \frac{\Gamma, \varphi \qquad \Gamma, \overline{\varphi}}{\Gamma}\ \text{Cut}$$

$$\frac{\Gamma, \oplus^a\Phi \qquad \Gamma, \oplus^b\Psi}{\Gamma, \oplus^{a+b}(\Phi, \Psi)}\ \text{Add} \qquad \frac{\Gamma, \oplus^a(\Phi, \Psi) \qquad \Gamma, \oplus^b\Psi}{\Gamma, \oplus^{a-b}\Phi}\ \text{Subtract}$$

$$\frac{\Gamma, \overline{\varphi}, \oplus^{b-1}\Phi \qquad \Gamma, \varphi, \oplus^b\Phi}{\Gamma, \oplus^b(\Phi, \varphi)}\ \text{MOD}$$

for each $a, b \in \{0, 1\}$.

The unorthodox form of the exchange rule is intended to make the height of proofs (see below) a more useful measure. The form of some of the rules, for instance Subtract, is inspired by [19].

Let $\mathcal{A}$ be a set of *non-logical axioms*, that is, $\mathcal{A}$ is a set of cedents. The intended meaning of $\mathcal{A}$ is $\bigwedge\{\bigvee\Xi : \Xi \in \mathcal{A}\}$. A PK($\oplus$)*-derivation of* Γ *from* $\mathcal{A}$ is a finite sequence $\Theta_1, \ldots, \Theta_k$ of cedents such that the last cedent Θ_k is Γ and every Θ_i either is a logical or non-logical axiom or is inferred from

some of the earlier cedents Θ_j $(j < i)$ using one of the inference rules. If Γ is the empty cedent, then the derivation is called a PK($\oplus$)-*refutation of* $\mathcal{A}$. We mostly think of PK($\oplus$) and its subsystems as refutation systems, i.e. we view a refutation of an unsatisfiable set of non-logical axioms $\mathcal{A}$ as a proof of the tautology $\bigvee\{\overline{\bigvee\Xi} : \Xi \in \mathcal{A}\}$.

A PK($\oplus$)-derivation $\Theta_1, \ldots, \Theta_k$ is called *treelike* if every Θ_i is a premise of at most one inference in the derivation.

The complexity of derivations will be measured in three ways. We define the *size* of a derivation P, denoted by $\mathbf{s}(P)$, to be the number of symbols in P. By the *cedent-size* of P, denoted by $\mathbf{cs}(P)$, we mean the number of occurrences of cedents, i.e. the "number of steps", in P. The *height* of P is the maximum number h such that there is a sequence $\Phi_0, \ldots, \Phi_h$ of cedents in P in which Φ_i is a premise of the inference yielding Φ_{i+1}, for each $i < h$.

We are interested in subsystems of PK($\oplus$) which we call $\mathrm{PK}^c_d(\oplus)$, where c, d are natural numbers. The systems $\mathrm{PK}^c_d(\oplus)$ are a variant of the family of proof systems introduced in [11][1]. A PK($\oplus$)-derivation is called a $\mathrm{PK}^c_d(\oplus)$-*derivation* if each formula in the derivation is of depth at most d, and further, each cedent in the derivation contains at most c formulas of the form $\oplus^b\Gamma$, where $b \in \{0, 1\}$ and the formulas in Γ do not contain any parity connective, while all the remaining formulas in the cedent do not contain any parity connective.

We can generalize the systems $\mathrm{PK}^c_d(\oplus)$ to $\mathrm{PK}^f_d(\oplus)$ for various functions $f\colon \mathbb{N} \to \mathbb{N}$. In a $\mathrm{PK}^f_d(\oplus)$ derivation P, each cedent may contain up to $f(\mathbf{s}(P))$ parities; the other syntactic conditions are as for $\mathrm{PK}^c_d(\oplus)$. For instance, if $\mathrm{id}\colon \mathbb{N} \to \mathbb{N}$ is the identity function, the system $\mathrm{PK}^{\mathrm{id}}_d(\oplus)$ may reason with arbitrary disjunctions (cedents) in which each disjunct is either a depth d formula in the De Morgan language or a parity of depth $d-1$ formulas.

Subsystems of PK($\oplus$) in which there is an $O(1)$ bound on the depth of formulas, but there is no additional bound on the nesting of $\bigwedge, \bigvee$, and $\oplus$ connectives, are collectively known as *constant-depth Frege with parity* or $\mathrm{AC}^0[2]$-*Frege*. The systems $\mathrm{PK}^0_{O(1)}(\oplus)$ are collectively known as *constant-depth Frege* or AC^0-*Frege*. An important family of systems intermediate

[1] The systems $\mathrm{F}^c_d(\mathrm{MOD}_p)$ of [11] are Hilbert-style rather than Tait-style, and the restriction on the shape of lines determined by c is somewhat more liberal than in our setting. Besides, [11] does not explicitly consider the systems as refutation systems. Despite these differences, our main results about $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$—Proposition 3, Theorems 11, 12, and 13—hold for $\mathrm{F}^{O(1)}_{O(1)}(\mathrm{MOD}_p)$ modulo the translation of cedents in one kind of system into lines in the other.

between AC^0- and $\mathrm{AC}^0[2]$-Frege is AC^0-*Frege with parity axioms*, in which the syntactic conditions on cedents are as in AC^0-Frege, but a derivation from the set of non-logical axioms $\mathcal{A}$ is allowed to use additional *parity axioms*, which are cedents of the form

$$\left(\bigwedge_{e \ni i} \neg \varphi_e : 1 \leq i \leq n \right), \ (\varphi_e \wedge \varphi_f : e \perp f) .$$

whenever n is some odd number, the φ_e's are indexed by two-element subsets of $[n] = \{1, \ldots, n\}$, and the entire cedent satisfies the appropriate condition on depth. Here $e \perp f$ stands for $\emptyset \subsetneq e \cap f \subsetneq e$. Note that the axiom says that the φ_e's do not define a partition of the odd-sized set $[n]$ into two-element subsets.

We will need to refer to some results on the algebraic proof system known as *Polynomial Calculus* over $\mathbb{F}_2$, first introduced under a different name in [13]. This is a system for refuting unsatisfiable families of polynomial equations over $\mathbb{F}_2$. Lines in a derivation are multivariate polynomials; each such polynomial p is understood to represent the equation $p = 0$. In addition to a given set of non-logical axioms $\mathcal{A}$, a Polynomial Calculus derivation may use axioms of the form $x^2 - x$ for any variable x. The rules are: from p derive xp where x is a variable; and from p, q derive $p+q$. The *degree* of a derivation is the highest degree of a line in it. A refutation is a derivation whose last line is the constant polynomial 1.

3 Basic properties of $\mathrm{PK}^c_d(\oplus)$

The purpose of this section is verify that the $\mathrm{PK}^c_d(\oplus)$ systems have two basic desirable properties. The first is that each $\mathrm{PK}^c_d(\oplus)$ is a complete proof system, in a reasonably strong sense of the term.

Proposition 1. *For all $c, d \geq 0$, the system $\mathrm{PK}^c_d(\oplus)$ is implicationally complete, in the sense that if $\mathcal{A}$ is a set of $\mathrm{PK}^c_d(\oplus)$ cedents, Γ is a $\mathrm{PK}^c_d(\oplus)$ cedent, and $\mathcal{A} \models \Gamma$, then there is a $\mathrm{PK}^c_d(\oplus)$ proof of Γ from $\mathcal{A}$.*

Proof. Let a set of $\mathrm{PK}^c_d(\oplus)$ cedents $\mathcal{A}$ semantically imply the $\mathrm{PK}^c_d(\oplus)$ cedent Γ.

We first prove the following *Claim*: every $\mathrm{PK}^c_d(\oplus)$ cedent Γ can be derived in $\mathrm{PK}^c_d(\oplus)$ from a set of non-logical axioms each of which: (i) either is the logical axiom $\oplus^0\emptyset$ or contains no formulas with $\oplus$, (ii) semantically follows from Γ, (iii) is falsified by at most one assignment to the variables of Γ.

We build the claimed $\mathrm{PK}^c_d(\oplus)$ derivation backwards from the endcedent Γ. First, a series of cuts on literals derives Γ from cedents containing Γ and, for each variable in Γ, either the variable itself or its negation. These new cedents satisfy the properties (ii) and (iii) from the claim. Then, using a series of MOD inferences, each of these cedents is derived from a set of $\mathrm{PK}^c_d(\oplus)$ cedents containing no formulas with $\oplus$ other than $\oplus^0\emptyset$ or $\oplus^1\emptyset$. Since the conclusion of a MOD inference implies each of its premises, properties (ii) and (iii) are preserved. Finally, each cedent containing $\oplus^0\emptyset$ is derived by weakening from a logical axiom, and each remaining cedent containing $\oplus^1\emptyset$ is derived by weakening from a semantically equivalent cedent without $\oplus^1\emptyset$. This gives us also property (i), thus proving the claim.

Now assume $\mathcal{A}$ semantically implies Γ and that Γ has the properties from the claim. Since Γ has at most one falsifying assignment, it is semantically implied by a single cedent $\Delta \in \mathcal{A}$. Let Δ be $\delta_1, \ldots, \delta_k$. By a series of cuts on the δ_j's, we can derive Γ from the cedent Γ, Δ and cedents of the form $\Gamma, \delta_1, \ldots, \delta_j, \neg\delta_{j+1}$. This is a $\mathrm{PK}^c_d(\oplus)$ derivation because Γ contains no $\oplus$'s. Now, the cedent Γ, Δ follows from Δ by weakening. On the other hand, the claim implies that each of the tautological cedents $\Gamma, \delta_1, \ldots, \delta_j, \neg\delta_{j+1}$ can be derived in $\mathrm{PK}^c_d(\oplus)$ from a set of tautological cedents without parities, which are derivable in $\mathrm{PK}^0_d(\oplus)$ by the completeness of the latter system. □

We now check that the treelike version of $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ polynomially simulates AC^0-Frege with parity axioms. This follows immediately from Proposition 3 below.

Lemma 2. *There is a polytime procedure which given a depth-d formula $\oplus^b(\Gamma, \varphi, \psi, \Delta)$ and the index of φ among the inputs to $\oplus^b$ produces a treelike $\mathrm{PK}^3_d(\oplus)$ derivation of this formula from $\oplus^b(\Gamma, \psi, \varphi, \Delta)$.*

Proof. Left to the reader as an (not entirely trivial) exercise. The reader who eschews such low-brow diversions is free to add the corresponding rule to the system. □

Proposition 3. *For each $c \geq 3, d \geq 1$, there is a polynomial-time procedure which, given a $\mathrm{PK}^c_d(\oplus)$ cedent Γ which is a parity axiom, outputs a treelike $\mathrm{PK}^1_{d+1}(\oplus)$ derivation $P(\Gamma)$ of Γ.*

Proof. Let Γ be

$$\left(\bigwedge_{e \ni i} \neg\varphi_e : 1 \leq i \leq n\right), (\varphi_e \wedge \varphi_f : e \perp f)\,.$$

For each $j = 0, \ldots, n$, we will refer to the sequence of formulas

$$\bigvee_{e \ni 1} \varphi_e, \ldots, \bigvee_{e \ni j} \varphi_e$$

as Θ_j, and to

$$\bigvee_{\substack{e \ni 1 \\ e \cap [j] \neq \emptyset}} \varphi_e, \ldots, \bigvee_{\substack{e \ni n \\ e \cap [j] \neq \emptyset}} \varphi_e$$

as Ξ_j. Note that Θ_n is the same as Ξ_n.

The final inference of $P(\Gamma)$ derives Γ by a cut on the formula

$$\oplus^0 \Theta_n.$$

The cedent $\Gamma, \oplus^1 \Theta_n$ is obtained by successively deriving $\Gamma, \oplus^{j \bmod 2} \Theta_j$ for $j = 0, \ldots, n$. The cedent $\Gamma, \oplus^0 \Theta_0$ follows by weakening from the logical axiom $\oplus^0 \Theta_0$, and $\Gamma, \oplus^b \Theta_{j+1}$ is easy to derive from (a single occurrence of) $\Gamma, \oplus^{1-b} \Theta_j$ by a MOD inference involving the formula $\bigvee_{e \ni j+1} \varphi_e$.

On the other hand, $\Gamma, \oplus^0 \Theta_n$, which coincides with $\Gamma, \oplus^0 \Xi_n$, is obtained by successively deriving $\Gamma, \oplus^0 \Xi_j$ for $j = 0, \ldots, n$. The cedent $\Gamma, \oplus^0 \Xi_0$ follows by weakening from $\oplus^0 \Xi_0$, which can be derived by a balanced tree of additions from multiple copies of the constant-size tautology $\oplus^0 \bot$. The cedent $\Gamma, \oplus^0 \Xi_{j+1}$ is derived by the subtraction rule from $\Gamma, \oplus^0 \Xi_j$ and $\Gamma, \oplus^0 (\Xi_j, \Xi_{j+1})$. Thus, we need to give a polysize treelike derivation of $\Gamma, \oplus^0 (\Xi_j, \Xi_{j+1})$.

It is easy to give a small treelike $\mathrm{PK}^0_d(\oplus)$ derivation of Γ from the cedents $\Gamma, \overline{\varphi_{\{j+1,\ell\}}}$ for $\ell = 1, \ldots, j, j+2, \ldots, n$. Using Lemma 2 and a balanced tree of addition inferences, $\Gamma, \overline{\varphi_{\{j+1,\ell\}}}, \oplus^0 (\Xi_j, \Xi_{j+1})$ can be derived from the cedents

$$\Gamma, \overline{\varphi_{\{j+1,\ell\}}}, \oplus^0 \left(\bigvee_{\substack{e \ni i \\ e \cap [j] \neq \emptyset}} \varphi_e, \bigvee_{\substack{e \ni i \\ e \cap [j+1] \neq \emptyset}} \varphi_e \right), \tag{1}$$

for i different from $j+1, \ell$, and the cedent

$$\Gamma, \overline{\varphi_{\{j+1,\ell\}}}, \oplus^0 \left(\bigvee_{\substack{e \ni j+1 \\ e \cap [j] \neq \emptyset}} \varphi_e, \bigvee_{\substack{e \ni \ell \\ e \cap [j] \neq \emptyset}} \varphi_e, \bigvee_{\substack{e \ni j+1 \\ e \cap [j+1] \neq \emptyset}} \varphi_e, \bigvee_{\substack{e \ni \ell \\ e \cap [j+1] \neq \emptyset}} \varphi_e \right). \tag{2}$$

Each cedent (1) is obtained by a small treelike derivation formalizing the argument "if Γ fails and $j+1$ is matched to ℓ, then the two formulas under

$\oplus^0$ are equivalent". The derivation of (2) depends on whether $\ell < j+1$ or $j+1 < \ell$. In the first case, we formalize the argument "if Γ fails and $j+1$ is matched to ℓ, then all four formulas under $\oplus^0$ are true"; in the second case, it is "...the last two formulas under $\oplus^0$ are true and the others false". $\square$

4 Balancing

Over the next four sections, we prove that for each c, d the treelike version of the system $\mathrm{PK}^c_d(\oplus)$ is quasipolynomially simulated by AC^0-Frege with parity axioms (a polynomial simulation is ruled out by [22], see Theorem 13 part (b)). The simulation is inspired by the "counting axioms simulate Nullstellensatz" argument of [16], but technically more complicated. The high-level overview of the argument is as follows. In Section 4, we show by a standard argument that a treelike $\mathrm{PK}^c_d(\oplus)$ refutation can be transformed into a balanced treelike refutation at the cost of allowing logarithmically many $\oplus$'s per cedent rather than a constant number. In Section 5 we transform a balanced treelike $\mathrm{PK}^{O(\log)}_{O(1)}(\oplus)$ refutation into a balanced treelike refutation in a system that has only one $\oplus$ per line, and some new rules. In Section 6, we show how to eliminate the subtraction rule, at the cost of a quasipolynomial increase in size and replacing the inputs Ψ to $\oplus$ by Ψ, Θ, Θ for various Θ. In 7, we show how to simulate the single-$\oplus$ system without subtraction in AC^0-Frege with parity axioms.

Lemma 4. *Suppose that $c \geq 1$ and P is a treelike $\mathrm{PK}^c_d(\oplus)$-derivation of Γ from $\mathcal{A}$. Suppose further that P has size σ and cedent-size τ. Then Γ can be derived from $\mathcal{A} \cup \{\varphi, \overline{\varphi} : \varphi$ is an element of a cedent of $P\}$ by a treelike $\mathrm{PK}^{2c+\log\tau}_{d+1}(\oplus)$-derivation of height $O(c \log \tau)$ and size $\sigma^{\log O(c)}$.*

Proof. First we prove that P can be transformed into a treelike $\mathrm{PK}^{2c+\log\tau}_{d+1}(\oplus)$-derivation P' which has height $\leq (2c+6)\log\tau$, cedent-size $\tau^{\log(6c+6)}$, and uses as non-logical axioms $\mathcal{A} \cup \mathsf{sa}(P)$, where $\mathsf{sa}(P)$ is the set of cedents of the form $\oplus^a\Phi, \overline{\oplus}^a\Phi$ such that $\oplus^a\Phi$ is an element of some cedent of P, or of the form $\bigwedge_i \overline{\gamma_i}, \Phi$ such that Φ is a cedent in P and $\gamma_1, \ldots, \gamma_\ell$ is the subsequence of Φ consisting of all formulas that do not contain any parity connective. (In this proof we write $\log\tau$ to mean $\max\{1, \log_2\tau\}$.) We shall prove the statement by induction on τ in a Brent-Spira divide-and-conquer fashion. To simplify the inductive argument, we actually show that P' can be a treelike $\mathrm{PK}^{2c+\log(\tau-1)}_{d+1}(\oplus)$-derivation of height $\leq (2c+6)\log(\tau-1)$.

This is obviously true for $\tau \leq (2c+6)\log(\tau-1)$. In the inductive step, we find an inference I in P with lower cedent Ψ and with the following

properties:

(a) for each premise Π of I, the subderivation H_Π of P with endcedent Π has cedent-size $\leq \lceil \tau/2 \rceil$,

(b) if we remove from P the subderivation ending with I, leaving Ψ as an initial cedent, then the resulting derivation (call it D) has cedent-size $\leq \tau/2$.

By the induction hypothesis, for each premise Π of the inference I there is a treelike $\mathrm{PK}^{2c+\log(\tau-1)-1}_{d+1}(\oplus)$-derivation of Π from $\mathcal{A} \cup \mathsf{sa}(H_\Pi)$ which has height $\leq (2c+6)(\log(\tau-1)-1)$. Put these derivations together with the inference I, and let Q denote the resulting derivation. Q is a treelike $\mathrm{PK}^{2c+\log(\tau-1)-1}_{d+1}(\oplus)$-derivation of Ψ from $\mathcal{A}$ with height $\leq (2c+6)(\log(\tau-1)-1)+1 = (2c+6)\log(\tau-1) - 2c - 5$.

Now we move on to the other part of the derivation P. As defined in (b), D is a treelike derivation of Γ from the hypotheses $\mathcal{A} \cup \{\Psi\}$, and $1 \leq \mathbf{cs}(D) \leq \tau/2$. In the case where $\mathbf{cs}(D) = 1$, we have $\Psi = \Gamma$, and we define the desired derivation P' to be just Q. Assume $\mathbf{cs}(D) > 1$. By the induction hypothesis, there is a treelike $\mathrm{PK}^{2c+\log(\tau-1)-1}_{d+1}(\oplus)$-derivation R of Γ from $\mathcal{A} \cup \{\Psi\} \cup \mathsf{sa}(D)$ of height $\leq (2c+6)(\log(\tau-1)-1)$. The idea is to transform R into several derivations from (only) the hypotheses $\mathcal{A} \cup \mathsf{sa}(D)$ and combine them with Q by a repeated use of the cut rule to derive Γ.

Let Δ be the subsequence $\delta_1, \ldots, \delta_m$ of Ψ consisting of all formulas that do not contain any parity connective. Denote by Δ' the sequence $\overline{\delta_1}, \ldots, \overline{\delta_m}$. Let $\oplus^{a_1}\Phi_1, \ldots, \oplus^{a_k}\Phi_k$ be the subsequence of Ψ consisting of all formulas with parities. We have $k \leq c$. For $i \in \{0, \ldots, k\}$ denote the sequence $\oplus^{a_1}\Phi_1, \ldots, \oplus^{a_i}\Phi_i$ by Θ_i. (Hence Θ_0 is the empty sequence.) We will modify R to construct derivations R' and R_i, $i = 1, \ldots, k$, of $\Gamma, \Theta_k, \bigwedge \Delta'$ and $\Gamma, \Theta_{i-1}, \overline{\oplus}^{a_i}\Phi_i$, respectively.

Construction of R': First, wherever Ψ occurs as an initial cedent in R, replace it with the cedent $\bigwedge \Delta', \Psi$. Second, replace every other initial cedent Ξ in R with

$$\dfrac{\dfrac{\Xi}{\Xi, \bigwedge \Delta'}}{\bigwedge \Delta', \Xi}$$

Third, add the formula $\bigwedge \Delta'$ to the left of every other cedent in R. Finally, apply to the endcedent $\bigwedge \Delta', \Gamma$ of the resulting derivation a weakening with Θ_k and one exchange to derive $\Gamma, \Theta_k, \bigwedge \Delta'$.

The construction of R_i, $i = 1, \ldots, k$, is similar. Every occurrence of Ψ as an initial cedent in R is now replaced with the derivation

$$\dfrac{\dfrac{\oplus^{a_i}\Phi_i, \overline{\oplus}^{a_i}\Phi_i}{\text{one weakening,}}}{\dfrac{\text{one exchange}}{\overline{\oplus}^{a_i}\Phi_i, \Psi}}$$

and every other initial cedent Ξ of R is replaced with

$$\dfrac{\dfrac{\Xi}{\Xi, \overline{\oplus}^{a_i}\Phi_i}}{\overline{\oplus}^{a_i}\Phi_i, \Xi}$$

To the left of every other cedent in R we add the formula $\overline{\oplus}^{a_i}\Phi_i$, hence obtaining a derivation with endcedent $\overline{\oplus}^{a_i}\Phi_i, \Gamma$. By one additional weakening and one exchange we get the derivation R_i of $\Gamma, \Theta_{i-1}, \overline{\oplus}^{a_i}\Phi_i$.

Finally, we combine the derivations Q, R' and R_i, $i = 1, \ldots, k$ with $k+1$ cut inferences to get the desired derivation P', as shown in Figure 1. Notice that just below the endcedent of the subderivation Q we inserted some necessary exchanges, a weakening, and a $\bigvee$-inference to make the cedent ready for cutting.

By their construction, each of the derivations R' and R_i, $i = 1, \ldots, k$, has at most $2c + \log(\tau - 1)$ parities per cedent and contains formulas of depth at most $d + 1$. We checked that Q is a treelike $\mathrm{PK}_{d+1}^{2c+\log(\tau-1)-1}(\oplus)$-derivation. Therefore, by construction, P' is a treelike $\mathrm{PK}_{d+1}^{2c+\log(\tau-1)}(\oplus)$-derivation.

The heights of R' and R_i, $i = 1, \ldots, k$, are at most four greater than the height of R, hence they are $\leq (2c+6)(\log(\tau-1) - 2c - 2$. From the height bound for Q and the construction it follows that P' has height at most $(2c+6)\log(\tau-1)$ as desired. The same is, of course, true in the case $\mathbf{cs}(D) = 1$ (which has $P' = Q$).

Next, we bound the cedent-size of P'. Denote $t := \log(6c+6)$. By the induction hypothesis, we have

$$\mathbf{cs}(Q) \leq 1 + \sum_{\Pi} \mathbf{cs}(H_\Pi)^t \leq 1 + (\sum_{\Pi} \mathbf{cs}(H_\Pi))^t \leq 1 + (\tau - \mathbf{cs}(D))^t,$$

where Π in the sums ranges over the premises of the inference I. Each of R', R_i, $i = 1, \ldots, k$, has cedent-size $\leq 3\mathbf{cs}(R)$. Hence, by construction,

$$\begin{aligned} \mathbf{cs}(P') &\leq \mathbf{cs}(Q) + (c+1)3\mathbf{cs}(R) + 2c + 4 \\ &< (\tau - \mathbf{cs}(D))^t + (3c+3)\mathbf{cs}(D)^t + 2c + 6. \end{aligned} \tag{3}$$

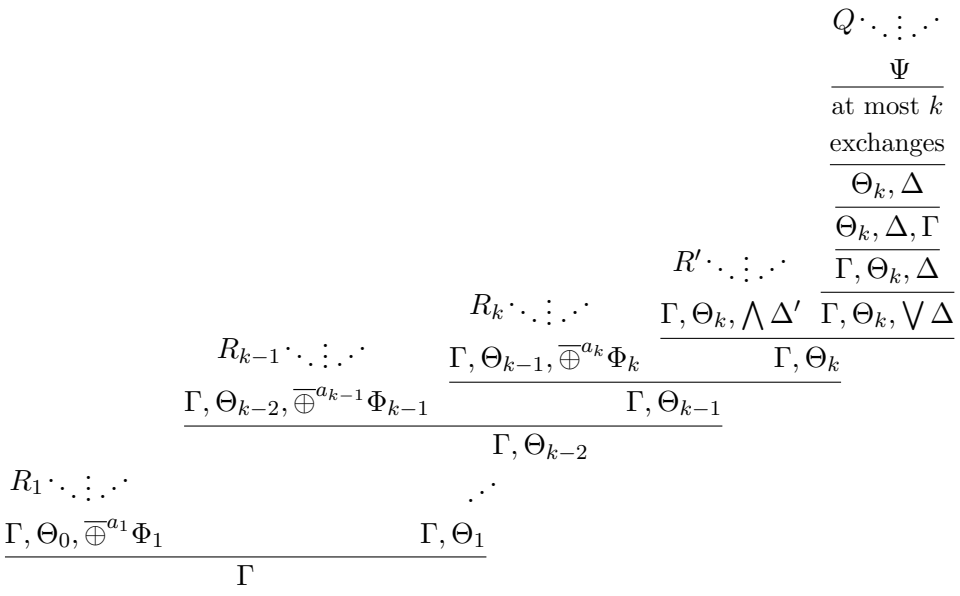

Figure 1: The derivation P'.

This is true in the case $\mathbf{cs}(D) = 1$ as well. The definition of t and some elementary calclulus implies that, as a function of $\mathbf{cs}(D)$, the last expression in (3) is decreasing on the interval $[1, \ldots, \tau/3]$. So, assume $\mathbf{cs}(D) = 1$. We have

$$\begin{aligned}\tau^t - (\tau-1)^t &\geq \tau^t - \tau^t + t\tau^{t-1} - \binom{t}{2}\tau^{t-2} = \tau^{t-2} t(\tau - (t-1)/2) \\ &= \tau^{\log(\frac{3c+3}{2})} \log(6c+6)\,(\tau - \log(3c+3)/2) \geq 5c+9.\end{aligned}$$

Here the last inequality holds because we assumed τ to be large enough ($\tau \geq (2c+6)\log\tau$). Thus $\mathbf{cs}(P') \leq \tau^t$ for $1 \leq \mathbf{cs}(D) \leq \tau/3$. Now assume $\tau/3 \leq \mathbf{cs}(D) \leq \tau/2$. Then

$$\begin{aligned}(\tau - \mathbf{cs}(D))^t + (3c+3)\mathbf{cs}(D)^t &\leq (\tau - \tau/3)^t + (3c+3)(\tau/2)^t \\ &= \tau^{\log(6c+6)} \left((2/3)^{\log(6c+6)} + 1/2 \right) \leq \tau^{\log(6c+6)} - (2c+6).\end{aligned}$$

Again, the last inequality follows from our assumption on τ, and it gives $\mathbf{cs}(P') \leq \tau^t$ for $\tau/3 \leq \mathbf{cs}(D) \leq \tau/2$. Thus, we have verified $\mathbf{cs}(P') \leq \tau^t$ for $1 \leq \mathbf{cs}(D) \leq \tau/2$, as was required.

It remains to transform P', which is a derivation from the axioms $\mathcal{A} \cup \mathsf{sa}(P)$, into a derivation from $\mathcal{A} \cup \{\varphi, \overline{\varphi} : \varphi \text{ is an element of some cedent of } P\}$. The only axioms of P' that are not as required are those of the form $\bigwedge_i \overline{\gamma_i}, \Phi$ such that Φ is a cedent in P and $\gamma_1, \ldots, \gamma_\ell$ is the subsequence of Φ consisting of all formulas that do not contain parity connectives. We construct a derivation P'' by attaching to P' the following derivation of each such cedent:

$$\dfrac{\dfrac{\dfrac{\gamma_i, \overline{\gamma_i}}{\text{one weakening,}}}{\text{two exchanges}}}{\dfrac{\Phi, \overline{\gamma_i}}{\dfrac{\Phi, \bigwedge \overline{\gamma_i}}{\bigwedge \overline{\gamma_i}, \Phi}}} \; , i = 1, \ldots, \ell$$

Let λ denote the maximum length of any cedent in P. Then

$$\mathbf{cs}(P'') \leq \mathbf{cs}(P') + (4\lambda + 1)\mathbf{cs}(P') \leq (4\lambda + 2)\tau^{\log(6c+6)}.$$

It follows from the construction that every cedent in P', and hence in P'', has size $\leq 2\sigma$. So, P'' has size $\leq 2\sigma(4\lambda + 2)\tau^{\log(6c+6)} = \sigma^{\log O(c)}$.

By construction, P'' is a treelike $\mathrm{PK}^{2c+\log\tau}_{d+1}(\oplus)$-derivation of Γ from the required set of axioms and has height $\leq 5 + (2c+6)\log\tau = O(c\log\tau)$.

□

Lemma 5. *Let Φ be a cedent of length ℓ of formulas which do not contain parity connectives and have depth d. Let φ be a formula in Φ. Let S and S_φ be the sizes of Φ and φ, respectively. There exist:*

(a) a treelike $\mathrm{PK}^0_d(\oplus)$-derivation of $\varphi, \overline{\varphi}$ of height $O(d)$ and size $O(S^2_\varphi)$,

(b) a treelike $\mathrm{PK}^2_d(\oplus)$-derivation of $\oplus^0\varphi, \oplus^1\varphi$ of height $O(d)$ and size $O(S^2_\varphi)$,

(c) a treelike $\mathrm{PK}^3_d(\oplus)$-derivation of $\oplus^0\Phi, \oplus^1\Phi$ of height $O(d + \log\ell)$ and size $O(\ell^{\log 3}S^2)$.

Proof. Using the first four rules of inference it is easy to construct (a). There is a height $O(1)$ size $O(S_\varphi)$ treelike $\mathrm{PK}^2_d(\oplus)$-derivation of $\oplus^0\varphi, \oplus^1\varphi$ from $\varphi, \overline{\varphi}$. Together with (a) this gives (b). Let Φ be Φ_1, Φ_2 where the length of Φ_1 is $\lfloor \ell/2 \rfloor$. There is a height $O(1)$ size $O(S)$ treelike $\mathrm{PK}^3_d(\oplus)$-derivation of $\oplus^0\Phi, \oplus^1\Phi$ which has six initial cedents: three $\oplus^0\Phi_1, \oplus^1\Phi_1$ and three $\oplus^0\Phi_2, \oplus^1\Phi_2$. Iterating this $\lceil \log\ell \rceil$ times gives a height $O(\log\ell)$

size $O(S\ell^{\log 3})$ treelike $\mathrm{PK}_d^3(\oplus)$-derivation of $\oplus^0\Phi, \oplus^1\Phi$ which has as initial cedents $O(\ell^{\log 3})$ copies of $\oplus^0\psi, \oplus^1\psi$ for each occurrence of ψ in Φ. Together with (b) this gives (c). □

Theorem 6. *Suppose that $c \geq 1$ and P is a treelike $\mathrm{PK}_d^c(\oplus)$-derivation of Γ from $\mathcal{A}$. Suppose further that P has size σ and cedent-size τ. Then there is a treelike $\mathrm{PK}_{d+1}^{2c+\log\tau}(\oplus)$-derivation of Γ from $\mathcal{A}$ which has height $O(d+c\log\sigma)$ and size $\sigma^{\log O(c)}$.*

Proof. By Lemma 4, there is a treelike $\mathrm{PK}_{d+1}^{2c+\log\tau}(\oplus)$-derivation P_1 of Γ from $\mathcal{A}\cup\{\varphi, \overline{\varphi} : \varphi \text{ is an element of a cedent of } P\}$ which has height $O(c\log\tau)$ and size $\sigma^{\log O(c)}$. Using derivations (a) and (c) of Lemma 5 to derive the non-logical axioms in P_1 that are not in $\mathcal{A}$ we obtain the desired derivation. □

5 Translation into a one-parity system

We want to transform a $\mathrm{PK}_d^c(\oplus)$-derivation into a derivation in a similar proof system which allows only one parity connective per line.

Definition. Let φ be a formula. Define $\mathsf{cnj}(\varphi)$ to be Θ if φ is of the form $\bigwedge\Theta$, and (φ) otherwise. Let Φ and Ψ be $(\varphi_i)_{i\in I}$ and $(\psi_i)_{i\in J}$, respectively. $\Phi \times \Psi$ denotes the sequence $\left(\bigwedge(\mathsf{cnj}(\varphi_i), \mathsf{cnj}(\psi_j))\right)_{i\in I, j\in J}$. Further, define recursively

$$\prod_{i=1}^{0}\Phi_i = \left(\bigwedge\emptyset\right) = (\top) \quad \text{and} \quad \prod_{i=1}^{k}\Phi_i = \left(\prod_{i=1}^{k-1}\Phi_i\right) \times \Phi_k$$

for sequences $\Phi_1, \ldots, \Phi_k$ of formulas.

The *depth d one-parity system*, denoted by $\mathrm{PK}_d^{\mathrm{one}\oplus}$, has lines of the form

$$\Gamma, \oplus\Phi$$

where Γ and Φ are sequences of formulas that do not contain parity connectives and have depth at most d, and moreover the formulas in Φ have $\bigwedge$ as their topmost connective. The intended meaning is $\bigvee(\Gamma, \oplus^0\Phi)$.

The logical axioms are:

$$\emptyset, \oplus\emptyset \quad \text{and} \quad x, \overline{x}, \oplus(\top)$$

for a propositional variable x.

The inference rules are:

$$\frac{\Gamma, \oplus\Phi}{\Gamma, \Delta, \oplus\Phi}\ \text{Weakening} \qquad \frac{\Gamma, \Delta, \Lambda, \Psi, \oplus\Phi}{\Gamma, \Lambda, \Delta, \Psi, \oplus\Phi}\ \text{Exchange}$$

$$\frac{\Gamma, \Delta, \oplus\Phi}{\Gamma, \bigvee\Delta, \oplus\Phi}\ \text{OR} \qquad \frac{\Gamma, \varphi_i, \oplus\Phi \qquad \text{for all } i \in I}{\Gamma, \bigwedge_{i\in I} \varphi_i, \oplus\Phi}\ \text{AND}$$

$$\frac{\Gamma, \varphi, \varphi, \oplus\Phi}{\Gamma, \varphi, \oplus\Phi}\ \text{Contraction} \qquad \frac{\Gamma, \varphi, \oplus\Phi \qquad \Gamma, \overline{\varphi}, \oplus\Phi}{\Gamma, \oplus\Phi}\ \text{Cut}$$

$$\frac{\Gamma, \oplus\Phi \qquad \Gamma, \oplus\Psi}{\Gamma, \oplus(\Phi, \Psi)}\ \text{Add} \qquad \frac{\Gamma, \oplus(\Phi, \Psi) \qquad \Gamma, \oplus\Psi}{\Gamma, \oplus\Phi}\ \text{Subtract}$$

$$\frac{\Gamma, \overline{\varphi}, \oplus\Phi \qquad \Gamma, \varphi, \oplus(\Phi, \top)}{\Gamma, \oplus(\varphi, \Phi, \top)}\ \text{MOD} \qquad \frac{\Gamma, \oplus\Phi}{\Gamma, \oplus(\Phi \times \Psi)}\ \text{Multiply}$$

$$\frac{\Gamma, \oplus\Phi}{\Gamma, \oplus\pi(\Phi)}\ \text{Permute}$$

where $\pi(\Phi)$ is a permutation of the formulas in Φ.

Lemma 7. *Each of the following cedents of the system* $\mathrm{PK}_d^{\mathrm{one}\oplus}$ *has a treelike* $\mathrm{PK}_d^{\mathrm{one}\oplus}$*-derivation of height* $O(d)$ *and size* $O(S^2)$*, where* S *is the size of the cedent.*

(a) $\varphi, \overline{\varphi}, \oplus(\top)$

(b) $\emptyset, \oplus(\bigwedge(\Gamma, \Phi), \bigwedge(\Gamma, \Phi, \Phi))$

(c) $\emptyset, \oplus(\bigwedge(\Gamma, \Phi, \Psi), \bigwedge(\Gamma, \Psi, \Phi))$

(d) $\overline{\bigwedge\Phi}, \oplus(\bigwedge(\Psi, \Phi), \bigwedge\Psi)$

(e) $\bigwedge\Phi, \oplus\bigwedge(\Psi, \Phi)$

Proof. (a) is proved in the same way as Lemma 5 (a). Each of (b) - (e) reduces by a height $O(1)$ size $O(S)$ treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation to constantly many cedents of the form $\emptyset, \oplus(\top, \top)$ or of the form $(\overline{\varphi_i})_{i\in I}, \bigwedge_{i\in I} \varphi_i, \oplus(\top)$; the latter are proved in the same way as (a). □

Definition. We map each cedent Γ of the system $\mathrm{PK}_d^c(\oplus)$ to a cedent $(\Gamma)^{\mathrm{one}\oplus}$ of the system $\mathrm{PK}_{d+1}^{\mathrm{one}\oplus}$ in the following way. Let Γ^{M} be the subse-

quence of Γ consisting of all formulas that do not contain any parity connective (M stands for de Morgan), and let $\oplus^{a_1}\Gamma_1, \ldots, \oplus^{a_k}\Gamma_k$ be the subsequence of Γ consisting of all formulas with parities. $(\Gamma)^{\mathrm{one}\oplus}$ is the cedent

$$\Gamma^{\mathsf{M}}, \oplus \prod_{i=1}^{k} \Gamma_i^{\mathsf{P}}$$

where for $i = 1, \ldots, k$ we put Γ_i^{P} to be Γ_i if $a_i = 0$ and $\Gamma_i, \top$ if $a_i = 1$.

Lemma 8. *Let P be a treelike $\mathrm{PK}_d^c(\oplus)$-derivation of Ω from $\mathcal{A}$. Suppose that P has size σ and height h. Then there is a treelike $\mathrm{PK}_{d+1}^{\mathrm{one}\oplus}$-derivation of $(\Omega)^{\mathrm{one}\oplus}$ from $\{(\Xi)^{\mathrm{one}\oplus} : \Xi \in \mathcal{A}\}$ which has size $O(\sigma^{2c})$ and height $O(h + d + c \log \sigma)$.*

Proof. Obtain P' by replacing each cedent Γ of P by its translation $(\Gamma)^{\mathrm{one}\oplus}$. Note that the logical axioms in P translate into logical axioms of $\mathrm{PK}_{d+1}^{\mathrm{one}\oplus}$. To make P' a $\mathrm{PK}_{d+1}^{\mathrm{one}\oplus}$-derivation we need to derive for every inference rule of $\mathrm{PK}_d^c(\oplus)$ the translation of its conclusion from the translations of its premises.

The Contraction rule either translates into Contraction or, if a parity is being contracted, becomes

$$\frac{\Gamma^{\mathsf{M}}, \oplus \left(\left(\prod_{i=1}^{k-2} \Gamma_i^{\mathsf{P}} \right) \times \Gamma_{k-1}^{\mathsf{P}} \times \Gamma_k^{\mathsf{P}} \right)}{\Gamma^{\mathsf{M}}, \oplus \left(\left(\prod_{i=1}^{k-2} \Gamma_i^{\mathsf{P}} \right) \times \Gamma_k^{\mathsf{P}} \right)} \tag{4}$$

where $\Gamma_{k-1}^{\mathsf{P}} = \Gamma_k^{\mathsf{P}}$. Denote $\prod_{i=1}^{k-2} \Gamma_i^{\mathsf{P}}$ by Θ and let Γ_k^{P} be $(\varphi_j)_{j \in I}$. By Lemma 7 (b), for each γ in Θ and $j \in I$, there is a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation of

$$\emptyset, \oplus \left(\bigwedge \left(\mathsf{cnj}(\gamma), \mathsf{cnj}(\varphi_j) \right), \bigwedge \left(\mathsf{cnj}(\gamma), \mathsf{cnj}(\varphi_j), \mathsf{cnj}(\varphi_j) \right) \right).$$

Attach these derivations to the leaves of a balanced tree of Adds whose root is (after one Permute) the cedent

$$\emptyset, \oplus \left(\Theta \times \Gamma_k^{\mathsf{P}}, \Theta \times \left(\mathsf{cnj}(\varphi_j) \wedge \mathsf{cnj}(\varphi_j) \right)_{j \in I} \right).$$

Call this derivation Q_1. Similarly, for each γ in Θ and $i < j \in I$, there is, by Lemma 7 (c), a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation of

$$\emptyset, \oplus \left(\bigwedge \left(\mathsf{cnj}(\gamma), \mathsf{cnj}(\varphi_i), \mathsf{cnj}(\varphi_j) \right), \bigwedge \left(\mathsf{cnj}(\gamma), \mathsf{cnj}(\varphi_j), \mathsf{cnj}(\varphi_i) \right) \right).$$

Put these derivations together with a balanced tree of Adds and one Permute to derive

$$\emptyset, \oplus\left(\Theta \times (\mathsf{cnj}(\varphi_i) \wedge \mathsf{cnj}(\varphi_j))_{i<j\in I}, \Theta \times (\mathsf{cnj}(\varphi_j) \wedge \mathsf{cnj}(\varphi_i))_{i<j\in I}\right).$$

Denote this derivation by Q_2. We derive (4) by:

$$\dfrac{\dfrac{\dfrac{\dfrac{Q_1 \cdot\cdot\cdot \quad Q_2 \cdot\cdot\cdot}{\emptyset, \oplus\left(\Theta \times \Gamma_k^{\mathsf{P}}, \Theta \times \Gamma_k^{\mathsf{P}} \times \Gamma_k^{\mathsf{P}}\right)}\,\text{Add \& Permute}}{\Gamma^{\mathsf{M}}, \oplus\left(\Theta \times \Gamma_k^{\mathsf{P}}, \Theta \times \Gamma_k^{\mathsf{P}} \times \Gamma_k^{\mathsf{P}}\right)}\,\text{Weakening} \qquad \Gamma^{\mathsf{M}}, \oplus\left(\Theta \times \Gamma_k^{\mathsf{P}} \times \Gamma_k^{\mathsf{P}}\right)}{\Gamma^{\mathsf{M}}, \oplus\left(\Theta \times \Gamma_k^{\mathsf{P}}\right)}\,\text{Subtract}$$

One of the two MOD rules translates into

$$\dfrac{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus\left(\left(\prod_{i=1}^k \Gamma_i^{\mathsf{P}}\right) \times \Phi\right) \qquad \Gamma^{\mathsf{M}}, \varphi, \oplus\left(\left(\prod_{i=1}^k \Gamma_i^{\mathsf{P}}\right) \times (\Phi, \top)\right)}{\Gamma^{\mathsf{M}}, \oplus\left(\left(\prod_{i=1}^k \Gamma_i^{\mathsf{P}}\right) \times (\varphi, \Phi, \top)\right)} \tag{5}$$

The other MOD rule translates into

$$\dfrac{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus\left(\left(\prod_{i=1}^k \Gamma_i^{\mathsf{P}}\right) \times (\Phi, \top)\right) \qquad \Gamma^{\mathsf{M}}, \varphi, \oplus\left(\left(\prod_{i=1}^k \Gamma_i^{\mathsf{P}}\right) \times \Phi\right)}{\Gamma^{\mathsf{M}}, \oplus\left(\left(\prod_{i=1}^k \Gamma_i^{\mathsf{P}}\right) \times (\varphi, \Phi)\right)} \tag{6}$$

Denote $\prod_{i=1}^k \Gamma_i^{\mathsf{P}}$ by Θ. Let us derive (5) first. We use Lemma 7 (d) and (e) to get for each γ in Θ a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation of

$$\overline{\varphi}, \oplus\left(\bigwedge(\mathsf{cnj}(\gamma), \mathsf{cnj}(\varphi)), \bigwedge \mathsf{cnj}(\gamma)\right) \quad \text{and} \quad \varphi, \oplus \bigwedge(\mathsf{cnj}(\gamma), \mathsf{cnj}(\varphi)),$$

respectively, and we attach each of these two sets of derivations to a balanced tree of Adds to obtain derivations Q_3 and Q_4 of

$$\overline{\varphi}, \oplus(\Theta \times (\varphi, \top)) \quad \text{and} \quad \varphi, \oplus(\Theta \times (\varphi)),$$

respectively. Denote the left premise of (5) by (L) and the right premise by (R). We can derive (5) by:

$$\dfrac{\dfrac{\dfrac{\dfrac{Q_3}{\overline{\varphi}, \oplus(\Theta \times (\varphi, \top))}}{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus(\Theta \times (\varphi, \top))}\,\text{Weak.} \quad \dfrac{\dfrac{Q_4}{\varphi, \oplus(\Theta \times (\varphi))}}{\Gamma^{\mathsf{M}}, \varphi, \oplus(\Theta \times (\varphi))}\,\text{Weak.}}{\dfrac{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus(\Theta \times (\varphi, \top)), \Theta \times \Phi)}{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus(\Theta \times (\varphi, \Phi, \top))}\,\text{Perm.}}\,\text{Add} \quad \dfrac{\Gamma^{\mathsf{M}}, \varphi, \oplus(\Theta \times (\varphi), \Theta \times (\Phi, \top))}{\Gamma^{\mathsf{M}}, \varphi, \oplus(\Theta \times (\varphi, \Phi, \top))}\,\text{Perm.}}{\Gamma^{\mathsf{M}}, \oplus(\Theta \times (\varphi, \Phi, \top))}\,\text{Cut}}\,\text{Add}$$

To derive (6), we can proceed (up to some permutations and weakenings) as follows: first add $\oplus(\Theta \times (\top, \top))$ to its right premise, then use the derivation of (5) (with Φ replaced by $\Phi, \top$), and then subtract $\oplus(\Theta \times (\top, \top))$.

Deriving the translations of the remaining rules is easier. The translation of the Exchange rule is derived with the help of Lemma 7 (c) together with a balanced tree of Adds to deal with the parity part of the premise and using one Exchange for the part outside parity. The translation of the Weakening rule is derived by one Weakening and one Multiply. The Cut rule either becomes Cut or, if the cut formula contains a parity, is derived by one Permute and one Subtract. The translations of the Add rules and Subtract rules are derived using the corresponding rule and Permute, possibly also requiring an addition and subtraction of $\oplus(\Theta \times (\top, \top))$.

Fill P' with the described derivations to obtain a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+1}$-derivation P''. Note that for any cedent Γ in P of size S the size of $(\Gamma)^{\mathrm{one}\oplus}$ is $O(S^c)$. Consider the inference J in P of which Γ is the conclusion. Any balanced tree of Adds that we had to attach to the translation of J has height $O(\log(S^c))$ and size $O(S^c \log(S^c))$. To the leaves of this tree we appended derivations of height $O(d)$ and of total size $O(S^{2c})$. Hence, what we added to the translation of J has height $O(d + c \log S)$ and size $O(S^{2c})$. Thus, P'' has the required properties. □

6 Delay of subtractions

Our strategy for obtaining the simulation of treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ by AC^0-Frege with parity axioms is to show that, given a balanced treelike small $\mathrm{PK}^{\mathrm{one}\oplus}_{O(1)}$ derivation from a set of (translations of) $\oplus$-free axioms $\mathcal{A}$, there is a family of small constant-depth formulas which describe matchings witnessing that if the cedents in $\mathcal{A}$ are satisfied, then each cedent $\Gamma, \oplus\Phi$ in the derivation is also satisfied—i.e. either Γ holds or there is a perfect matching on the satisfied elements of Φ. For this, the translation to $\mathrm{PK}^{\mathrm{one}\oplus}_{O(1)}$ was needed because a constant-depth definable matching for a disjunction

$\oplus\Phi_1, \ldots, \oplus\Phi_c$ of parities should try to witness the entire disjunction rather than try to choose one of the disjuncts (e.g., $(\oplus^0\Phi, \oplus^1\Phi)^{\mathrm{one}\oplus}$ is easy to witness, but trying to witness either $\oplus^0\Phi$ or $\oplus^1\Phi$ would in general require a small constant-depth formula for parity).

A further problem with this strategy is caused by the Subtract rule

$$\frac{\Gamma, \oplus(\Phi, \Psi) \qquad \Gamma, \oplus\Psi}{\Gamma, \oplus\Phi}$$

which is the one-parity system's way of rendering both subtractions and cuts on parities. The issue is that some $\varphi \in \Phi$ may be matched in the left premise to $\psi \in \Psi$, which cannot be maintained in the conclusion since ψ does not appear in it. We could try to match φ to the $\varphi' \in \Phi$ such that φ' is matched to $\psi' \in \Psi$ in the left premise and ψ' is matched to ψ in the right premise. However, keeping track of this by means of constant-depth formulas leads to exponential blowup. The following lemma helps us deal with the problematic rule.

Lemma 9. *Let P be a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation of $\Gamma, \oplus\Phi$ from $\mathcal{A}$. Suppose that P has size σ and height h. Then for some sequence Θ there is a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation of $\Gamma, \oplus(\Theta, \Theta, \Phi)$ from $\mathcal{A}$ which does not use the Subtract rule and which has size $O(\sigma^h)$ and height $O(h)$.*

Proof. The lemma is clear for $h = 1$, and we will proceed by induction on height. Let $h > 1$ and let J be the last inference of P. First, we apply the induction hypothesis to each of the subderivations P_i of the premises $\Gamma_i, \oplus\Phi_i$ of J, for $i = 1, \ldots, k$, to obtain derivations P_i' of $\Gamma_i, \oplus(\Theta_i, \Theta_i, \Phi_i)$. Then we form the desired derivation P' by putting the derivations P_i' together with a simple derivation Q which depends on J.

If J is Weakening, Exchange, OR, Contraction, or Permute, then we apply the same rule to $\Gamma_1, \oplus(\Theta_1, \Theta_1, \Phi_1)$ to derive $\Gamma, \oplus(\Theta_1, \Theta_1, \Phi)$ and we let Θ be Θ_1.

If J is Add, apply Add to the premises $\Gamma, \oplus(\Theta_i, \Theta_i, \Phi_i)$, $i = 1, 2$, and use one Permute to derive $\Gamma, \oplus(\Theta, \Theta, \Phi)$, where Θ is Θ_1, Θ_2.

If J is Subtract with premises $\Gamma, \oplus(\Phi, \Psi)$ and $\Gamma, \oplus\Psi$, then Q is

$$\dfrac{\dfrac{\Gamma, \oplus(\Theta_1, \Theta_1, \Phi, \Psi) \qquad \Gamma, \oplus(\Theta_2, \Theta_2, \Psi)}{\Gamma, \oplus(\Theta_1, \Theta_1, \Phi, \Psi, \Theta_2, \Theta_2, \Psi)}\,\mathrm{Add}}{\Gamma, \oplus(\Theta_1, \Theta_2, \Psi, \Theta_1, \Theta_2, \Psi, \Phi)}\,\mathrm{Permute}$$

So Θ is Θ_1, Θ_2, Ψ.

If J is Multiply with premise $\Gamma, \oplus\Phi_1$ and conclusion $\Gamma, \oplus(\Phi_1 \times \Psi)$, then Q is

$$\frac{\Gamma, \oplus(\Theta_1, \Theta_1, \Phi_1)}{\Gamma, \oplus(\Theta_1 \times \Psi, \Theta_1 \times \Psi, \Phi_1 \times \Psi)}\ \text{Multiply}$$

So Θ is $\Theta_1 \times \Psi$.

For the remaining rules, AND, Cut and MOD, we obtain $\Gamma, \oplus(\Theta, \Theta, \Phi)$ by the same rule (plus one Permute in the case of MOD), but in order to use this rule we must first equalize the overheads Θ_i. Let us consider only the AND rule with premises $\Gamma_i, \oplus\Phi$, $i = 1, \ldots, k$; Cut and MOD are easier. Let $\widetilde{\Theta}_i$ denote the sequence $\Theta_1, \ldots, \Theta_{i-1}, \Theta_{i+1}, \ldots, \Theta_k$ and let Θ be $\Theta_1, \ldots, \Theta_k$. Define Q to be the following derivation:

$$\dfrac{\dfrac{\dfrac{}{\Gamma_i, \oplus(\Theta_i, \Theta_i, \Phi)}\quad \dfrac{R}{\Gamma_i, \oplus\left(\widetilde{\Theta}_i, \widetilde{\Theta}_i\right)}}{\Gamma_i \oplus (\Theta, \Theta, \Phi)}\ \text{Add \& Permute} \qquad i = 1, \ldots, k}{\Gamma, \oplus(\Theta, \Theta, \Phi)}\ \text{AND}$$

Here R is a short derivation which first derives $\oplus(\top, \top)$ and then applies Multiply and Weakening to it.

The size and height bounds on P' follow easily from the construction. $\square$

7 Simulation by parity axioms

We are now ready to complete our simulation of treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ by AC^0-Frege with parity axioms. For a sequence Ψ of formulas we denote its length by $\mathrm{lh}(\Psi)$, and we will often list the members of Ψ as $\psi_1, \ldots, \psi_{\mathrm{lh}(\Psi)}$.

Definition. Let Ψ be a sequence of formulas and $\mathrm{lh}(\Psi) = k$. For each $e \in [k]^2$ let μ_e be a formula. We say that the μ_e's *define a matching on the satisfied elements* of Ψ if the following formula holds:

$$\bigwedge_{\{u,v\} \in [k]^2} \left(\mu_{\{u,v\}} \to (\psi_u \wedge \psi_v)\right) \wedge \left(\bigwedge_{\substack{e,f \in [k]^2 \\ e \perp f}} \overline{\mu}_e \vee \overline{\mu}_f \right) \wedge \bigwedge_{u \in [k]} \left(\psi_u \to \bigvee_{\substack{e \in [k]^2 \\ u \in e}} \mu_e \right).$$

Lemma 10. *Let $\mathcal{A}$ be a set of cedents consisting of formulas which do not contain parity connectives. Let P be a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_d$-derivation of $\Delta, \oplus(\Theta, \Theta, \top)$ from $\{(\Upsilon)^{\mathrm{one}\oplus} : \Upsilon \in \mathcal{A}\}$ which does not use the Subtract rule. Suppose that P has size σ. Then there is a AC^0-Frege with parity axioms derivation of Δ from $\mathcal{A}$ with size polynomial in σ.*

Proof. For each cedent $\Omega, \oplus\Xi$ in P, we will construct constant-depth formulas which attempt to define a matching on the satisfied elements of Ξ. Using $\mathcal{A}$ as a hypothesis, we prove in AC^0-Frege that if Ω is false, then the formulas do in fact define a matching.

In the particular case of the final cedent $\Delta, \oplus(\Theta, \Theta, \top)$, we will be able to conclude that if Δ is false, then there is a matching on the satisfied elements of $(\Theta, \Theta, \top)$. However, there is an obvious matching on the satisfied elements of (Θ, Θ) defined by formulas $\nu_e = \theta_i$ for $e = \{i, i + \mathrm{lh}(\Theta)\}$, $i \in [\mathrm{lh}(\Theta)]$, and $\nu_e = \bot$ for all other $e \in [2 \cdot \mathrm{lh}(\Theta)]^2$. It is not difficult to rule out the coexistence of these two contradictory matchings by a polysize proof in AC^0-Frege with parity axioms (see [16]), and this is actually the only place where we use the parity axioms.

We now describe how to construct the formulas defining the matchings. It will be clear from the construction that we need only a polynomial number of polysize constant-depth formulas, and that the proofs of their properties also need only size $\mathrm{poly}(\sigma)$.

If C denotes the line $\Omega, \oplus\Xi$ of P, call any $u \in [\mathrm{lh}(\Xi)]$ a *parity input index* of C, and let $\xi_{C,u}$ denote the formula in Ξ with this index.

First, we define an obvious notion of a predecessor of a parity input index u, intended to be the index of a direct ancestor of ξ_u in a premise of the rule used to derive C. Let J be an inference rule of $\mathrm{PK}_d^{\mathrm{one}\oplus}$, denote by C the conclusion of J and let C' be some premise of J. Let u, u' be some parity input indices of C, C', respectively. The binary relation $(C, u)\, R\, (C', u')$, *the parity input index u' of C' is a predecessor of the parity input index u of C*, is defined as follows (referring to the rules as stated in Section 5):

If J is one of the rules Weakening, Exchange, OR, AND, Contraction, Cut, then $(C, u)\, R\, (C', u')$ iff $u = u'$.
If J is Add, then $(C, u)\, R\, (C', u')$ iff either $u = u'$ and C' is the left premise, or $u = u' + \mathrm{lh}(\Phi)$ and C' is the right premise.
If J is MOD, then $(C, u)\, R\, (C', u')$ iff $u = u' + 1$.
If J is Multiply, then $(C, u)\, R\, (C', u')$ iff $\lceil u/\mathrm{lh}(\Psi) \rceil = u'$.
If J is Permute, then $(C, u)\, R\, (C', u')$ iff $u = \pi(u')$.

Observe that, with the exception of the Multiply rule, the formulas $\xi_{C,u}$ and $\xi_{C',u'}$ are identical.

Next, let $0 < m \in \mathbb{N}$, and suppose that for each $i \in [m]$, $C_i = \Omega_i, \oplus\Xi_i$ is a line in P obtained by a rule J_i, which has C_{i+1} as a premise in case $i < m$. Let $e_i = \{u_i, v_i\} \in [\mathrm{lh}(\Xi_i)]^2$ for $i \in [m]$. We say that the sequence

$$B = (C_1, e_1), (C_2, e_2), \ldots, (C_m, e_m)$$

is a *matching branch of* (C_1, e_1) if and only if J_m is MOD, $e_m = \{1, \mathrm{lh}(\Xi_m)\}$, and for each $i \in [m-1]$,

$$(C_i, u_i)\, R\, (C_{i+1}, u_{i+1}) \wedge (C_i, v_i)\, R\, (C_{i+1}, v_{i+1});$$

and additionally, if J_i is Multiply with $\Psi := \Psi_i$, then $u_i \equiv v_i \pmod{\mathrm{lh}(\Psi_i)}$.

So, a matching branch leads to a line where some ancestors of u_1, v_1 appeared together for the first time. Moreover, we insist on a specific form of the line and of u_m, v_m: ξ_{C_m,u_m} should be the formula entering inside $\oplus$ in the conclusion of a MOD inference, and ξ_{C_m,v_m} should be the final $\top$ inside that $\oplus$. We want B to give rise to an edge between u_1, v_1 only if the following condition β^B holds:

$$\xi_{C_m,1} \wedge \bigwedge_{i \in [m-1]} \alpha_{J_i,C_{i+1}},$$

where $\alpha_{J_i,C_{i+1}}$ is a formula justifying the choice of premise C_{i+1} made by B if J_i was one of the rules Cut, MOD, AND, and it is defined (only for these rules) as follows:

$$\alpha_{\mathrm{Cut},C'} = \begin{cases} \varphi & \text{if } C' \text{ is the right premise} \\ \overline{\varphi} & \text{if } C' \text{ is the left premise} \end{cases}$$

$$\alpha_{\mathrm{MOD},C'} = \begin{cases} \varphi & \text{if } C' \text{ is the left premise} \\ \overline{\varphi} & \text{if } C' \text{ is the right premise} \end{cases}$$

$$\alpha_{\mathrm{AND},C'} = \overline{\varphi}_\ell \wedge \bigwedge_{j=1}^{\ell-1} \varphi_j \quad \text{if } C' \text{ is the } \ell\text{th premise.}$$

Here φ is, respectively, the cut formula in case of Cut and the formula entering inside $\oplus$ in case of MOD; and φ_j is the jth auxiliary formula of the AND rule.

Now, for a line $C = \Omega, \oplus\Xi$ in P with $k = \mathrm{lh}(\Xi) \geq 2$ and for $e = \{u, v\} \in [k]^2$, let $\mu_{C,e}$ be the formula

$$\xi_{C,u} \wedge \xi_{C,v} \wedge \bigvee_{B \in \mathcal{B}} \beta^B,$$

where $\mathcal{B}$ is the set of all matching branches of (C, e). We prove by induction that if $\bigvee \Omega$ is false, then the formulas $\mu_{C,e}$ define a matching on the satisfied formulas of Ξ. This means that we have to give constant-depth Frege

derivations from $\mathcal{A}$ of the following cedents:

$$\Omega, \overline{\mu}_{C,e}, \xi_{C,u} \quad \text{for each } u \in e \in [k]^2, \tag{7}$$

$$\Omega, \overline{\xi}_{C,u}, (\mu_{C,e})_{e \ni u} \quad \text{for each } u \in [k], \tag{8}$$

$$\Omega, \overline{\mu}_{C,e}, \overline{\mu}_{C,f} \quad \text{for each } e, f \in [k] \text{ such that } e \perp f. \tag{9}$$

Using Lemma 5 (a) we derive $\overline{\xi}_{C,u}, \xi_{C,u}$ from which (7) follows easily. So we now concentrate on (8) and (9). If C is an axiom, both are obvious.

Assume that C is obtained from $C' = \Omega', \oplus\Xi'$ by one of the rules Weakening, Exchange, OR, Contraction. For $u \in [k]$ we have $\mu_{C,e} = \mu_{C',e}$ for each $e \in [k]^2$ with $u \in e$, so (8) and (9) are easily obtained from the induction hypothesis.

Assume that C is obtained by AND. If the rule has no premises, (8) is obvious and (9) as well, since Ω is $\top$. So assume that the AND inference has $n \geq 1$ premises and $k \geq 2$ (the case $k \leq 1$ is straightforward), and let $u \in e \in [k]^2$. Notice that for each premise C_ℓ, to each matching branch B' of (C_ℓ, e) corresponds a unique matching branch B of (C, e), such that β^B is equivalent to $\alpha_{\mathrm{AND},C_\ell} \wedge \beta^{B'}$. Denote by $\mathcal{B}_\ell^e$ the set of matching branches of (C_ℓ, e). To obtain (8), one may use, for each $\ell \in [n]$, the induction hypothesis

$$\Gamma, \varphi_\ell, \overline{\xi}_{C,u}, \left(\xi_{C,u} \wedge \xi_{C,v} \wedge \bigvee_{B \in \mathcal{B}_\ell^{\{u,v\}}} \beta^B \right)_{v \in \{u,v\} \in [k]^2}$$

to derive

$$\Gamma, \overline{\alpha}_{\mathrm{AND},C_\ell}, \overline{\xi}_{C,u}, \left(\xi_{C,u} \wedge \xi_{C,v} \wedge \alpha_{\mathrm{AND},C_\ell} \wedge \bigvee_{B \in \mathcal{B}_\ell^{\{u,v\}}} \beta^B \right)_{v \in \{u,v\} \in [k]^2}.$$

Cutting these n cedents against the tautology $\alpha_{\mathrm{AND},C_1}, \ldots, \alpha_{\mathrm{AND},C_n}, \bigwedge_{\ell \in [n]} \varphi_\ell$, which has a small proof, we obtain

$$\Gamma, \left(\bigwedge_{\ell \in [n]} \varphi_\ell \right), \overline{\xi}_{C,u}, \left(\xi_{C,u} \wedge \xi_{C,v} \wedge \alpha_{\mathrm{AND},C_\ell} \wedge \bigvee_{B \in \mathcal{B}_\ell^{\{u,v\}}} \beta^B \right)_{\substack{v \in \{u,v\} \in [k]^2, \\ \ell \in [n]}}$$

from which the desired

$$\Gamma, \left(\bigwedge_{\ell \in [n]} \varphi_\ell \right), \overline{\xi}_{C,u}, \left(\xi_{C,u} \wedge \xi_{C,v} \wedge \bigvee_{\ell \in [n]} \bigvee_{B \in \mathcal{B}_\ell^{\{u,v\}}} \alpha_{\mathrm{AND},C_\ell} \wedge \beta^B \right)_{v \in \{u,v\} \in [k]^2}$$

follows using OR rules and some easily derivable regrouping and distributivity properties. To obtain (9), let $e = \{u, v\} \perp f = \{u, w\} \in [k]^2$ and use the induction hypothesis

$$\Gamma, \varphi_\ell, \overline{\xi_{C,u} \wedge \xi_{C,v} \wedge \bigvee_{B \in \mathcal{B}^e_\ell} \beta^B}, \overline{\xi_{C,u} \wedge \xi_{C,w} \wedge \bigvee_{B \in \mathcal{B}^f_\ell} \beta^B}.$$

for every $\ell \in [n]$, to derive

$$\Gamma, \overline{\xi}_{C,u}, \overline{\xi}_{C,v}, \overline{\bigvee_{B \in \mathcal{B}^e_\ell} \alpha_{\mathrm{AND},C_\ell} \wedge \beta^B}, \overline{\xi}_{C,u}, \overline{\xi}_{C,w}, \overline{\bigvee_{B \in \mathcal{B}^f_\ell} \alpha_{\mathrm{AND},C_\ell} \wedge \beta^B}. \tag{10}$$

For $\ell \neq \ell'$, the tautology $\overline{\alpha_{\mathrm{AND},C_\ell}}, \overline{\alpha_{\mathrm{AND},C_{\ell'}}}$ has a small proof, and we can use it to derive

$$\Gamma, \overline{\xi}_{C,u}, \overline{\xi}_{C,v}, \overline{\bigvee_{B \in \mathcal{B}^e_\ell} \alpha_{\mathrm{AND},C_\ell} \wedge \beta^B}, \overline{\xi}_{C,u}, \overline{\xi}_{C,w}, \overline{\bigvee_{B \in \mathcal{B}^f_{\ell'}} \alpha_{\mathrm{AND},C_{\ell'}} \wedge \beta^B}. \tag{11}$$

Applying the AND rule $n + 1$ many times to the sets of cedents (10), (11) and the OR rule twice, we obtain

$$\Gamma, \overline{\xi_{C,u} \wedge \xi_{C,v} \wedge \bigvee_{\ell \in [n]} \bigvee_{B \in \mathcal{B}^e_\ell} \alpha_{\mathrm{AND},C_\ell} \wedge \beta^B}, \overline{\xi_{C,u} \wedge \xi_{C,w} \wedge \bigvee_{\ell \in [n]} \bigvee_{B \in \mathcal{B}^f_\ell} \alpha_{\mathrm{AND},C_\ell} \wedge \beta^B},$$

from which (9) follows by a weakening with $\bigwedge_{\ell \in [n]} \varphi_\ell$.

If C is one of the remaining rules, Cut, Add, MOD, Multiply, Permute, the verifications of (8) and (9) are easier than for the AND rule and use similar ideas, so we will omit them. □

Theorem 11. *For each $c, d \in \mathbb{N}$, there is a quasipolynomial-time procedure which, given a treelike $\mathrm{PK}^c_d(\oplus)$ refutation of a set of $\oplus$-free cedents $\mathcal{A}$, produces a refutation of $\mathcal{A}$ in AC^0-Frege with parity axioms.*

Proof. Assuming that $\mathcal{A}$ has a treelike $\mathrm{PK}^c_d(\oplus)$ refutation of size σ:

(i) apply Theorem 4 to obtain a treelike $\mathrm{PK}^{\log \sigma + O(1)}_{d+1}(\oplus)$ refutation of $\mathcal{A}$ of size $\sigma^{O(1)}$ and height $O(\log \sigma)$,

(ii) apply Lemma 8 to obtain a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+2}$ refutation of $\{(\Xi)^{\mathrm{one}\oplus} : \Xi \in \mathcal{A}\}$ of size $\sigma^{O(\log \sigma)}$ and height $O(\log^2 \sigma)$,

(iii) apply Lemma 9 to obtain a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+2}$ derivation of $\oplus(\Theta,\Theta)$ from $\{(\Xi)^{\mathrm{one}\oplus} : \Xi \in \mathcal{A}\}$ which has size $\sigma^{O(\log^3 \sigma)}$, height $O(\log^2 \sigma)$ and does not use the Subtract rule,

(iv) apply Lemma 10 to obtain a refutation of $\mathcal{A}$ in AC^0-Frege with parity axioms which has size polynomial in $\sigma^{O(\log^3 \sigma)}$.

All the results used in the simulation are proved by explicit constructions which can be carried out in time polynomial in $\sigma^{O(\log^3 \sigma)}$. □

Remark. By part (b) of Theorem 13, the simulation in Theorem 11 cannot be improved to polynomial.

8 Exponential separations for formulas with $\bigoplus$

In this section we show that it is easy to separate treelike and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ from each other and from $\mathrm{AC}^0[2]$-Frege if we treat them as refutation systems, and, crucially, if the refuted sets of cedents are allowed to contain $\oplus$ connectives.

Theorem 12. *There exist families $\{\mathcal{A}_n\}_{n\in\omega}$ and $\{\mathcal{B}_n\}_{n\in\omega}$ of unsatisfiable sets of $\mathrm{PK}^1_2(\oplus)$ cedents such that:*

(a) each $\mathcal{A}_n$ has a $\mathrm{poly}(n)$-size refutation in $\mathrm{PK}^{\mathrm{id}}_2(\oplus)$, but requires $2^{n^{\Omega(1)}}$-size refutations in $\mathrm{PK}^c_d(\oplus)$ for any constants c, d,

(b) each $\mathcal{B}_n$ has a $\mathrm{poly}(n)$-size refutation in $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, but requires $2^{n^{\Omega(1)}}$-size refutations in treelike $\mathrm{PK}^c_d(\oplus)$ for any constants c, d.

Proof. We prove (a). The idea is to take a family of narrow CNF's which have small constant-depth Frege refutations but no low-degree Polynomial Calculus refutations, and to replace each variable by a parity of fresh variables. For concreteness, consider the ordering principle restricted to an expander graph as formulated in [14] (weak PHP would do just as well except that it is only known to have a quasipolynomial-size constant-depth proof). Given a degree 9 expander $G = (V, E)$ on n vertices, the CNF $\mathrm{GOP}(G)$ consists of the following clauses in the variables x_{ij}, $i<j\in[n]$:

$$\begin{aligned} \overline{x_{ij}}, \overline{x_{jk}}, x_{ik}, &\qquad i<j<k\in[n],\\ x_{ij}, x_{jk}, \overline{x_{ik}} &\qquad i<j<k\in[n],\\ \left(\overline{x_{ji}} \colon (i,j)\in E, j<i\right), \left(x_{ij} \colon (i,j)\in E, j>i\right) &\qquad i\in[n] \end{aligned}$$

Intuitively, if we think of the variables as describing a linear ordering on $[n]$, where x_{ij} means that i is below j in the ordering, $\mathrm{GOP}(G)$ says that each element of $[n]$ is smaller than one of its neighbours in G. The algebraic reformulation of $\mathrm{GOP}(G)$ is the following set of polynomials over $\mathbb{F}_2$:

$$\begin{aligned} x_{ij}x_{jk}(1+x_{ik}), &\qquad i<j<k\in[n],\\ (1+x_{ij})(1+x_{jk})x_{ik}, &\qquad i<j<k\in[n],\\ \prod_{\substack{(i,j)\in E\\ j<i}} x_{ji}\cdot \prod_{\substack{(i,j)\in E\\ j>i}} (1+x_{ij}) &\qquad i\in[n] \end{aligned}$$

We obtain $\mathcal{A}_n$ by replacing each x_{ij} with $\sum_\ell x_{ij\ell}$ for distinct fresh variables $x_{ij\ell}$, $\ell\in[n]$, and rewriting the resulting polynomials as $\oplus$'s of conjunctions. So, $\mathcal{A}_n$ consists of the formulas:

$$\begin{aligned} &\oplus^0(\{x_{ij\ell_1}\wedge x_{jk\ell_2} : \ell_1,\ell_2\in[n]\},\\ &\qquad \{x_{ij\ell_1}\wedge x_{jk\ell_2}\wedge x_{ik\ell_3} : \ell_1,\ell_2,\ell_3\in[n]\}),\\ &i<j<k\in[n],\\ &\oplus^0(\{x_{ik\ell_3} : \ell_3\in[n]\},\\ &\qquad \{x_{ij\ell_1}\wedge x_{ik\ell_3} : \ell_1,\ell_3\in[n]\},\\ &\qquad \{x_{jk\ell_2}\wedge x_{ik\ell_3} : \ell_2,\ell_3\in[n]\},\\ &\qquad \{x_{ij\ell_1}\wedge x_{jk\ell_2}\wedge x_{ik\ell_3} : \ell_1,\ell_2,\ell_3\in[n]\}),\\ &i<j<k\in[n], \end{aligned}$$

and the somewhat messy formulas corresponding to the width-9 clauses of $\mathrm{GOP}(G)$.

A polysize refutation of $\mathcal{A}_n$ in $\mathrm{PK}_2^{\mathrm{id}}(\oplus)$ can be obtained by first deriving the clauses of the CNF statement of $\mathrm{GOP}(G)$ with $\oplus^1_\ell x_{ij\ell}$'s substituted for the x_{ij}'s, and then performing the same substitution in a polysize resolution refutation of $\mathrm{GOP}(G)$ [23].

On the other hand, any $\mathrm{PK}^c_d(\oplus)$ refutation of $\mathcal{A}_n$ requires size $2^{n^{\Omega(1/d)}}$. To see this, let P be a $\mathrm{PK}^c_d(\oplus)$ refutation of size $S = 2^{n^{a\cdot(1/d)}}$ where the constant a is small enough, as determined by the argument below. Let $n_0 = \binom{n}{2}n$ and $n_{i+1} = n_i/(O(\log S))$. Apply a series of random restrictions $\rho_1\dots\rho_{d+1}$ as in [2, Sections 2 and 6.1], with ρ_i leaving n_i out of n_{i-1} variables unassigned. Assuming S is small enough, w.h.p. $\rho=\rho_1\dots\rho_{d+1}$ switches all $\oplus$-free subformulas of formulas appearing in P, as well as all the formulas $\bigvee\Gamma$ for $\Gamma,\oplus\Psi_1,\dots,\oplus\Psi_c$ a cedent in P, into canonically defined decision trees of height $\log S = n^{a\cdot(1/d)}$. Also w.h.p. assuming S is small enough,

$\rho_1 \dots \rho_{d+1}$ leaves at least one $x_{ij\ell}$ unassigned for each $i<j\in[n]$. Apply an additional restriction τ which for each i and j sets all $x_{ij\ell}$'s except one, for instance to 0. Simplify the decision trees accordingly.

As a result of the above procedure, each cedent $\Gamma, \oplus\Psi_1, \dots, \oplus\Psi_c$ in P gets mapped to a product of canonically defined polynomials $p_\Gamma p_{\Psi_1} \dots p_{\Psi_c}$ obtained from the decision trees for $\bigvee\Gamma$ and for the elements of $\Psi_1, \dots, \Psi_c$: p_Γ is 0 exactly if $\bigvee\Gamma$ holds, and p_{Ψ_m} is 0 exactly if $\oplus\Psi_m$ holds for $m \in [c]$. Each polynomial $p_\Gamma p_{\Psi_1} \dots p_{\Psi_c}$ has degree at most $h = (c+1)n^{a\cdot(1/d)}$. Moreover, a tedious but straighforward verification reveals that for every inference in P, the polynomial representing the conclusion can be derived from the polynomials representing the premises in degree-$O(h)$ Polynomial Calculus (in fact, these derivations are treelike and their number of lines is polynomial in $\max(\text{number of premises}, 2^h)$). This gives a degree-$O(h)$ PC refutation of $\mathcal{A}_n\restriction_{\rho\tau}$. However, $\mathcal{A}_n\restriction_{\rho\tau}$ is essentially GOP(G), which by [14, Theorem 2] has no Polynomial Calculus refutation of degree less than $n/108$—a contradiction if a was chosen small enough.

Part (b) is proved analogously, except that to define $\mathcal{B}_n$ we need tautologies that have low degree proofs in Polynomial Calculus but not in the Nullstellensatz proof system of [3]—for instance, the housesitting tautologies of [13, 8]. □

9 The Impagliazzo-Segerlind argument

We now verify that some superpolynomial separations between the systems we are studying can be witnessed by formulas in the De Morgan language. In particular, this is the case for AC^0-Frege with parity axioms and treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, which are quasipolynomally equivalent by Theorem 11 and Proposition 3.

Theorem 13. *There exist families $\{\mathcal{A}_n\}_{n\in\omega}$ and $\{\mathcal{B}_n\}_{n\in\omega}$ of unsatisfiable CNF's such that:*

(a) each $\mathcal{A}_n$ has a poly(n)*-size refutation in* $\mathrm{PK}_2(\oplus)$*, but requires $n^{\omega(1)}$-size refutations in* $\mathrm{PK}^c_d(\oplus)$ *for any constants c, d,*

(b) each $\mathcal{B}_n$ has a poly(n)*-size refutation in treelike* $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$*, but requires $n^{\omega(1)}$-size refutations in* AC^0*-Frege with parity axioms.*

Proof. Both parts of the theorem are proved using a technique of Impagliazzo and Segerlind presented in [15] and described in full detail in [22, Chapter VI]. We expect the reader to have those two texts at hand.

Part (b) is actually witnessed by $\mathcal{B}_n = \mathrm{IS}(\mathcal{U})$, where $\mathcal{U}$ is what [15] would call an (n,n)-universe and the $\mathrm{IS}(\mathcal{U})$'s are CNF's used in to separate constant-depth Frege with parity axioms from constant-depth Frege with parity gates. For a fixed n, the formula $\mathcal{B}_n$ has variables $x_{i\ell}$ for $i,\ell \in [n]$ as well as some auxiliary variables, and claims in an obfuscated way that all of

$$\oplus^0(x_{1\ell} : \ell \in [n]), \tag{12}$$

$$\oplus^0(x_{i\ell} : \ell \in [n]) \Rightarrow \oplus^0(x_{(i+1)\ell} : \ell \in [n]), \qquad i \in [n], \tag{13}$$

$$\oplus^1(x_{n\ell} : \ell \in [n]), \tag{14}$$

are satisfied. Each $\mathcal{B}_n$ has size $\mathrm{poly}(n)$ and has a polysize Polynomial Calculus refutation, but requires superpolynomial-size refutations in constant-depth with parity axioms. However, it is straightforward to give a polysize *treelike* $\mathrm{PK}_2^3(\oplus)$ refutation of $\mathcal{B}_n$: first, derive the formulas (12)-(14) from the clauses of $\mathcal{B}_n$, and then arrange a refutation of (12)-(14) into a balanced tree of cuts.

We sketch a proof of (a). The idea is essentially the same as that of [15] and [22], except that instead of considering (and suitably modifying) formulas that are easy for Polynomial Calculus but somewhat hard for Nullstellensatz, we work with formulas that are easy for constant-depth systems but hard for Polynomial Calculus. Since the technical part of the lower bound proof is quite involved[2] but conceptually almost identical to the one described by Impagliazzo and Segerlind, we focus on describing the construction of $\mathcal{A}_n$ and explain the lower bound argument only briefly, outsourcing the details to [22, Chapter VI].

As in the case of Theorem 12 part (a), our starting point is a family of tautologies that are hard for Polynomial Calculus but not for constant-depth systems. Once again, we want to replace individual variables by parities, but now the resulting formulas are not allowed to contain $\oplus$ connectives, so every statement of the form "an even number of $\varphi_1, \ldots, \varphi_k$ are true" has to be reexpressed using auxiliary variables that give a perfect matching on the satisfied elements of $\{\varphi_1, \ldots, \varphi_k\}$. Moreover, in doing this we have to avoid making too many parity statements implicitly definable by constant-depth formulas in the new variables; otherwise, $\mathcal{A}_n$ will be easy for AC^0-Frege.

As our underlying family of tautologies, this time we choose the weak

[2] Chapter VI of [22] is almost 60 pages.

pigeonhole principle PHP_m^{2m}. As a set of polynomials, this consists of:

$$\begin{aligned} 1 + \sum_{j \in [m]} x_{ij}, &\qquad && i \in [2m], \\ x_{i_1 j} \cdot x_{i_2 j}, &\qquad && i_1 < i_2 \in [2m], j \in [m] \end{aligned}$$

The reason for preferring weak PHP to the GOP formulas used in Theorem 12 part (a) is that it leads to a definition of $\mathcal{A}_n$ that is more perspicuous and easier to work with, whereas the fact that GOP has slightly smaller constant-depth refutations is no longer helpful because we are only proving a superpolynomial separation.

Now, given n, where w.l.o.g. n is even, choose m quasipolynomially smaller than n such that there exist constant-depth refutations of PHP_m^{2m} of size n. Replacing each x_{ij} by a sum of n variables x_{ijk}, $k \in [n]$, and rewriting the polynomials as $\oplus$'s of conjunctions, leads to the set of formulas:

$$\oplus^1 \left(\{ x_{ijk} : j \in [m], k \in [n] \} \right), \qquad i \in [2m], \tag{15}$$

$$\oplus^0 \left(\{ x_{i_1 jk} \wedge x_{i_2 j\ell} : k, \ell \in [n] \} \right), \qquad i_1 < i_2 \in [2m], j \in [m] \tag{16}$$

To obtain $\mathcal{A}_n$, we introduce an additional set of $nm + 1$ "type-1 extra points" for each i, and a set of n^2 "type-2 extra points" for each triple (i_1, i_2, j); note that $nm + 1$ is an odd number and n^2 is even. We then reexpress (15) for a given i by saying that there is a perfect matching on the union of the set of type-1 extra points and the set of x_{ijk}'s with value 1. We reexpress (16) for (i_1, i_2, j) by saying that there is a perfect matching on the union of the set of type-2 extra points and the set of pairs (k, ℓ) such that both $x_{i_1 jk}$ and $x_{i_2 j\ell}$ evaluate to 1. In both cases, it helps to simplify things if all edges in the matchings are required to contain at least one extra point.

In more detail, $\mathcal{A}_n$ is a CNF in the variables:

$$\begin{aligned} x_{ijk}, &\qquad && i \in [2m], j \in [m], k \in [n], \\ y_{ijkp}, &\qquad && i \in [2m], j \in [m], k \in [n], p \in [mn + 1], \\ v_{ie}, &\qquad && i \in [2m], e \in [[mn + 1]]^2, \\ z_{i_1 i_2 jk\ell q}, &\qquad && i_1 < i_2 \in [2m], j \in [m], k, \ell \in [n], q \in [n^2], \\ w_{i_1 i_2 jf}, &\qquad && i_1 < i_2 \in [2m], j \in [m], f \in [[n^2]]^2. \end{aligned}$$

Intuitively, y_{ijkp} says that x_{ijk} is matched to the type-1 extra point p; v_{ie} says that the two extra points in e are matched for the given i; $z_{i_1 i_2 jk\ell q}$ says that the pair of $x_{i_1 jk}$ and $x_{i_2 j\ell}$ is matched to the type-2 extra point q; and

$w_{i_1 i_2 j f}$ says that the two extra points in f are matched for (i_1, i_2, j). The clauses of $\mathcal{A}_n$ are:

$$\overline{x_{ijk}} \vee \bigvee_{p\in[nm+1]} y_{ijkp} \qquad i\in[2m], j\in[m], k\in[n],$$

$$\overline{y_{ijkp}} \vee x_{ijk} \qquad i\in[2m], j\in[m], k\in[n], p\in[mn+1],$$

$$\bigvee_{j\in[m],k\in[n]} y_{ijkp} \vee \bigvee_{e\ni p} v_{ie} \qquad i\in[2m], p\in[mn+1],$$

$$\overline{y_{ijkp}} \vee \overline{y_{ij'k'p}} \qquad i\in[2m], (j,k)\neq(j',k'), p\in[mn+1],$$

$$\overline{y_{ijkp}} \vee \overline{y_{ijkp'}} \qquad i\in[2m], j\in[m], k\in[n], p\neq p',$$

$$\overline{y_{ijkp}} \vee \overline{v_{ie}} \qquad i\in[2m], j\in[m], k\in[n], p\in e,$$

$$\overline{v_{ie}} \vee \overline{v_{ie'}} \qquad i\in[2m], e\perp e',$$

$$\overline{x_{i_1jk}} \vee \overline{x_{i_2j\ell}} \vee \bigvee_{q\in[n^2]} z_{i_1i_2jk\ell q} \qquad i_1<i_2\in[2m], j\in[m], k,\ell\in[n],$$

$$\overline{z_{i_1i_2jk\ell q}} \vee x_{i_1jk} \qquad i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\in[n^2],$$

$$\overline{z_{i_1i_2jk\ell q}} \vee x_{i_2j\ell} \qquad i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\in[n^2],$$

$$\bigvee_{k,\ell\in[n]} z_{i_1i_2jk\ell q} \vee \bigvee_{f\ni q} w_{i_1i_2jf}, \qquad i_1<i_2\in[2m], j\in[m], q\in[n^2],$$

$$\overline{z_{i_1i_2jk\ell q}} \vee \overline{z_{i_1i_2jk'\ell' q}} \qquad i_1<i_2\in[2m], j\in[m], (k,\ell)\neq(k',\ell'), q\in[n^2],$$

$$\overline{z_{i_1i_2jk\ell q}} \vee \overline{z_{i_1i_2jk\ell q'}} \qquad i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\neq q',$$

$$\overline{z_{i_1i_2jk\ell q}} \vee \overline{w_{i_1i_2jf}} \qquad i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\in f,$$

$$\overline{w_{i_1i_2jf}} \vee \overline{w_{i_1i_2jf'}} \qquad i_1<i_2\in[2m], j\in[m], f\perp f'.$$

A poly(n)-size $\mathrm{AC}^0[2]$-Frege refutation of $\mathcal{A}_n$ can be obtained by first deriving the clauses of PHP^{2m}_m with $\oplus^1(\{x_{ijk} : k\in[n]\})$ substituted for x_{ij}, and then making the same substitution in the size-n AC^0-Frege refutation of PHP^{2m}_m.

To prove that $\mathcal{A}_n$ cannot have polynomial-size $\mathrm{PK}^c_d(\oplus)$ refutations P for any c, d we employ a switching lemma argument. As in [15, 22], in order to make the switching work, we have to consider not just random restrictions but random simplifications, which allow some variables to be replaced by other variables rather than by $0, 1$ values. More specifically, we consider *random* (n, n^ϵ)*-simplifications*, where ϵ is sufficiently small and it is assumed w.l.o.g. that $n - n^\epsilon$ is divisible by 4. Performing such a simplification splits into the following steps:

(i) for each i, j:

(a) set $\frac{n-n^\epsilon}{2}$ randomly chosen x_{ijk}'s to 1 and $\frac{n-n^\epsilon}{2}$ randomly chosen x_{ijk}'s to 0; set all y, z variables involving the latter x_{ijk}'s to 0; leave the n^ϵ remaining x_{ijk}'s unset;

(b) choose at random a perfect matching on the set of x_{ijk}'s that have been set to 1;

(ii) for each i:

(a) choose at random an injection which assigns a type-1 extra point to each pair (j, k) such that x_{ijk} has been set to 1; set all corresponding y variables to 1 and all conflicting y and v variables to 0;

(b) choose at random a set of $m\frac{n-n^\epsilon}{2}$ hitherto unused type-1 extra points and a perfect matching on it; set all corresponding v variables to 1 and all conflicting y and v variables to 0; thus, $mn^\epsilon+1$ type-1 extra points remain unmatched;

(iii) for each i_1, i_2, j:

(a) choose at random an injection which assigns a type-2 extra point to each pair (k, ℓ) such that both x_{i_1jk} and $x_{i_2j\ell}$ have been set to 1; set all corresponding z variables to 1 and all conflicting z and w variables to 0;

(b) choose at random an injection which assigns a type-2 extra point to each pair (k, ℓ) such that one of x_{i_1jk} and $x_{i_2j\ell}$ has been set to 1 and the other is unset; set all conflicting z variables to 0 (w variables are dealt with in steps (c), (d) below);

(c) for each k such that x_{i_1jk} remains unset: for each ℓ such that $x_{i_2j\ell}$ has been set to 1, substitute x_{i_1jk} for $z_{i_1jk\ell p}$, where p is the extra point assigned to (k, ℓ) in step (b); substitute $\neg x_{i_1jk}$ for the $w_{i_1i_2jf}$ such that f consists of the extra points assigned to (k, ℓ) and (k, ℓ') where $x_{i_2j\ell'}$ has been set to 1 and $x_{i_2j\ell}, x_{i_2j\ell'}$ are matched in step (i)(b); set all $w_{i_1i_2jf'}$ variables for $f \perp f'$ to 0;

(d) perform a step analogous to (c) with the roles of i_1 and i_2 interchanged;

(e) choose at random a set of $n^2 - \left(\frac{(n-n^\epsilon)^2}{4} + n^\epsilon(n-n^\epsilon) + n^{2\epsilon}\right)$ hitherto unused extra points and a perfect matching on it; set all corresponding w variables to 1 and all conflicting z and w variables to 0; thus, $n^{2\epsilon}$ extra points remain unmatched through steps (a)-(e).

A decision tree for $\mathcal{A}_n$ is allowed to ask one of the following: whether an x variable is true—and get a yes/no answer; for a given i, what an x_{ijk} is matched to—and get as an answer either "x_{ijk} is 0" or a y_{ijkp} together with "x_{ijk} is 1"; for a given i, what a type-1 extra point p is matched to—and get as an answer either a v_{ie} or a y_{ijkp} together with "x_{ijk} is 1"; for given i_1, i_2, j, what $(x_{i_1jk}, x_{i_2j\ell})$ is matched to—and get as an answer either a list of values for $x_{i_1jk}, x_{i_2j\ell}$ which are not both 1 or a z variable and the information that both x's are 1; finally, for given i_1, i_2, j, what a type-2 extra point q is matched to—and get as an answer either a w variable or a z variable together with the information that the appopriate two x's are both 1. A decision tree *strongly represents* a DNF φ if each branch of the tree either contains a term of φ or is inconsistent with each term of φ (in the natural loose sense of inconsistency where a y or w variable can be inconsistent with a negated x variable, two y's can be inconsistent with each other etc.).

Since P has only polynomial size, we can apply a switching lemma analogous to [15, Theorem 7.3.1]/[22, Theorem 83] d+1 times[3]. Let ρ stand for the simplification (*simplifying restriction* in the terminology of [15]) built during the iteration of the lemma. This ρ induces a mapping from all $\oplus$-free

[3]We use this opportunity to point out two apparent issues with [22], at least one of which is not addressed in [15]. Firstly, the definition of presimplification used in [22] is missing a step analogous to our (i)(b). To use our setting and notation, this is as if the mapping between $x_{i_2j\ell}$'s set to 1 induced in (iii)(c) by replacing some z's by an x_{i_1jk}, some w's by $\neg x_{i_1jk}$ and some other w's by 0's was defined separately for each x_{i_1jk} that remains unset; and similarly for (iii)(d) (cf. [22, Definition VI.G.2, part 7.]). The effect is that the first paragraph of the proof of [22, Lemma 89]—specifically, the claim "the number of L-presimplifications which are extended by a given $(L-e)$-presimplification depends only on L and e"—fails, and the statement of that lemma is actually false. The fix we know is to add a step analogous to (i)(b). This leads to a factor of $\frac{1}{2}$ appearing in the statement of the corrected version of [22, Lemma 89], but that has no bearing on the eventual statement of [22, Theorem 83].

The other issue is to some extent fixed in [15], but some of the details are not present in that extended abstract. We explain the problem in the language of [22]. The proof of [22, Lemma 91], specifically the argument justifying the claim "there is a literal of T_{i+1} not set by B^ρ", does not seem to work under the current definitions of independent set and of s-encoding [22, Definitions VI.G.10, VI.G.11]. A solution is to change those definitions by requiring the terms in a "B-independent set for F with respect to ρ" to be ρ-consistent with B and not just with each other (this change is already made in [15, Definition 7.3.5]) and requiring an "s-encoding for ρ with respect to F" to satisfy not just the s terms of an independent set but also the associated set B (this is not discussed in [15]). However, the size of B—w.l.o.g., at most $O(r^2s^2)$—must then be taken into account in the proof and statement of [22, Lemma 90]. Eventually, this results in much worse bounds in [22, Lemma 84]: the term L^{Cr} has to be replaced by L^{Cr^2s}. Once again, though, no changes to the statement of [22, Theorem 83] are needed.

subformulas of formulas in P, as well as all formulas $\bigvee\Gamma$ for Γ the $\oplus$-free part of a cedent in P, to constant-height decision trees strongly representing them. Given a bound $h\in\mathbb{N}$ on the height of the decision trees, the mapping is an h-evaluation in a sense analogous to [22, Definition VI.H.1] or of [4]. Each line of $P\restriction_\rho$ can be viewed as a degree-h polynomial. Moreover, using the properties of h-evaluations, one verifies that for each inference in P, the polynomial representing the conclusion restricted by ρ can be derived in constant-degree Polynomial Calculus from the polynomials representing the premises restricted by ρ together with $\mathcal{A}_n\restriction_\rho$. Apply a further substitution of variables by constants, τ, so that $\mathcal{A}_n\restriction_{\rho\tau}$ becomes identical to PHP^{2m}_m. Now $P\restriction_{\rho\tau}$ is a constant-degree Polynomial Calculus refutation of PHP^{2m}_m, which does not exist by [21]. □

Remark. In all likelihood, applying the techniques of [15] to a suitable family of formulas hard for Nullstellensatz/treelike Polynomial Calculus but not for daglike Polynomial Calculus would give a family of DNFs with polysize refutations in daglike but not treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$. However, we have not investigated this in detail.

10 The Itsykson-Sokolov system

In [17], Itsykson and Sokolov study a refutation system they call Res-Lin in which lines are cedents of $\oplus$'s of literals. They obtain strong lower bounds on refutation size for the treelike version of their system, but leave lower bounds for the daglike version as an open problem. Krajíček [12] has recently developed a randomized version of the feasible interpolation method aimed at attacking the problem.

In this section, we prove a simple result that illustrates a difference between the treelike and daglike versions of systems like Res-Lin in which lines express disjunctions of parities, and suggests that proving a lower bound for daglike Res-Lin might require special-purpose techniques. Namely, we show that treelike $\mathrm{PK}^{\mathrm{id}}_{O(1)}(\oplus)$, which strengthens treelike Res-Lin by allowing arbitrary constant-depth formulas rather than just literals as inputs to the $\oplus$ connectives, is simulated by (daglike) $\mathrm{PK}^{\log}_{O(1)}(\oplus)$, which yields lower bounds proved in more or less the same way as for $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$. As we explain in a remark below, such lower bounds are unlikely to be easily provable for daglike $\mathrm{PK}^{\mathrm{id}}_{O(1)}(\oplus)$.

Theorem 14. *Suppose that $\mathcal{A}$ is a set of $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$ cedents each of which contains at most p formulas with an $\oplus$ connective. If $\mathcal{A}$ has a treelike $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$-refutation of size σ and cedent-size τ, then it also has a $\mathrm{PK}_d^{\log\tau+p+3}(\oplus)$-refutation of size $\mathrm{poly}(\sigma)$.*

Proof. Let us refer to the number of formulas with $\oplus$ occurring in the cedent as the $\oplus$-*width* of the cedent. For a set of cedents $\mathcal{A}$ and a derivation P we denote by $\mathcal{A}^P$ the set of cedents of the form Θ, Ω or of the form $\Theta, \varphi, \overline{\varphi}$, where $\Omega \in \mathcal{A}$ and φ and each element of Θ occurs (possibly negated) as an element of a line in P.

By induction on τ we prove the following: if $\mathcal{A}$ is a set of cedents of $\oplus$-width p and $\Xi = \xi_1, \ldots, \xi_k$ has a treelike $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$-derivation P from $\mathcal{A}$ of size σ and cedent-size τ, then there is a $\mathrm{PK}_d^{\log\tau+p+3}(\oplus)$-refutation P' of $\mathcal{A}^P \cup \{(\overline{\xi}_i) : i \in [k]\}$ of cedent-size $5\sigma\tau$, such that for each occurrence of a formula in P there are at most two occurrences of the same formula (possibly negated) in each line of P', and such that each element of a line in P' occurs (possibly negated) as an element of a line in P.

If $\tau = 1$ and $\Xi \in \mathcal{A}$, we obtain a refutation of $\oplus$-width at most p by cutting Ξ against each $(\overline{\xi}_i)$. The case where Ξ is a logical axiom is obvious.

For the inductive step, if Ξ is derived by Weakening, Exchange, or Contraction, then the induction hypothesis applied to the premise of the rule already gives the refutation P' we need.

In considering the other rules, we write Γ to stand for the side formulas of the conclusion Ξ and assume that $\Gamma = \gamma_1, \ldots, \gamma_k$.

If $\Xi = \Gamma, \bigvee \Delta$ is derived by the OR rule, denote by Q the subderivation of P with endcedent Γ, Δ. Let $\Delta = \delta_1, \ldots, \delta_\ell$. By the induction hypothesis, we have a $\mathrm{PK}_d^{\log(\tau-1)+p+3}(\oplus)$-refutation Q' of $\mathcal{A}^Q \cup \{(\overline{\gamma}_i) : i \in [k]\} \cup \{(\overline{\delta}_i) : i \in [\ell]\}$. Obtain P' by attaching 5 cedents above each axiom $(\overline{\delta}_i)$ of Q' to form its derivation from $(\bigwedge_{i\in[\ell]} \overline{\delta}_i)$ and $(\overline{\delta}_i, \delta_i)$. The refutation P' has the required properties, in particular $\mathbf{cs}(P') \leq \mathbf{cs}(Q') + 5\ell \leq 5\mathbf{s}(Q)\mathbf{cs}(Q) + 5\ell \leq 5(\mathbf{s}(\Xi) + \mathbf{s}(Q))(1 + \mathbf{cs}(Q)) = 5\sigma\tau$.

If $\Xi = \Gamma, \bigwedge_{j\in[m]} \varphi_j$ is derived by the AND rule, denote by Q_j the subderivation of P with endcedent Γ, φ_j. By the induction hypothesis, we have for each $j \in [m]$ a $\mathrm{PK}_d^{\log(\tau-1)+p+3}(\oplus)$-refutation Q'_j of $\mathcal{A}^{Q_j} \cup \{(\overline{\gamma}_i) : i \in [k]\} \cup \{(\overline{\varphi}_j)\}$. Obtain Q''_j by adding $(\overline{\varphi}_1, \ldots, \overline{\varphi}_{j-1})$ in front of every line in Q'_j, and arrange these derivations so that for $j \in [m-1]$ the axiom $(\overline{\varphi}_1, \ldots, \overline{\varphi}_j)$ of Q''_j is the endcedent of Q''_{j+1}. Add at most km cedents so that each axiom of Q''_j, $2 \leq j \leq m$, containing some $\overline{\gamma}_i$ is derived from $(\overline{\gamma}_i)$ by a weakening and exchange. Attach $4m + 2$ cedents to the axiom $(\overline{\varphi}_1, \ldots, \overline{\varphi}_m)$ of Q''_m to form its derivation from $(\bigvee_{j\in[m]} \overline{\varphi}_j)$

and $\{(\overline{\varphi}_j, \varphi_j) : j \in [m]\}$. The resulting refutation P' satisfies $\mathbf{cs}(P') \leq \sum_{j\in[m]} \mathbf{cs}(Q'_j) + km + 4m + 2 \leq 5\sum_{j\in[m]} \mathbf{s}(Q_j)\mathbf{cs}(Q_j) + (k+4)m + 2 \leq 5(\mathbf{s}(\Xi) + \sum_{j\in[m]} \mathbf{s}(Q_j))(1 + \sum_{j\in[m]} \mathbf{cs}(Q_j)) = 5\sigma\tau$ as well as the remaining requirements.

If $\Xi = \Gamma, \oplus^b(\Phi, \varphi)$ is derived by the MOD rule, let Q_1 and Q_2 denote, respectively, the subderivation of P with endcedent $(\Gamma, \overline{\varphi}, \oplus^{b-1}\Phi)$ and $(\Gamma, \varphi, \oplus^b\Phi)$. Suppose that $\mathbf{cs}(Q_1) \leq \mathbf{cs}(Q_2)$ (the opposite case is treated analogously). So $\mathbf{cs}(Q_1) < \tau/2$ and $\mathbf{cs}(Q_2) < \tau - 1$. By the induction hypothesis, there is a $\mathrm{PK}_d^{\log\tau+p+2}(\oplus)$-refutation Q'_1 of $\mathcal{A}^{Q_1} \cup \{(\overline{\gamma}_i) : i \in [k]\} \cup \{(\varphi), (\oplus^b\Phi)\}$, and there is a $\mathrm{PK}_d^{\log(\tau-1)+p+3}(\oplus)$-refutation Q'_2 of $\mathcal{A}^{Q_2} \cup \{(\overline{\gamma}_i) : i \in [k]\} \cup \{(\overline{\varphi}), (\oplus^{b-1}\Phi)\}$. Obtain a derivation Q''_1 of $\overline{\varphi}$ by adding $\overline{\varphi}$ in front of every line in Q'_1, and similarly obtain a derivation Q'''_1 of $\oplus^{b-1}\Phi$ by adding $\oplus^{b-1}\Phi$ in front of each line in Q'_1. Attach Q'_2 to Q''_1 and Q'''_1. Add at most $3k$ cedents so that each axiom of Q''_1 and Q'''_1 containing some $\overline{\gamma}_i$ is derived from $(\overline{\gamma}_i)$ by a weakening and exchange. Attach 9 cedents to the axiom $(\overline{\varphi}, \oplus^b\Phi)$ of Q''_1 to form its derivation from cedents $(\oplus^{b-1}(\Phi, \varphi))$, $(\oplus^b\Phi, \oplus^{b-1}\Phi)$ and $(\overline{\varphi}, \varphi)$. Similarly, attach 8 cedents to the axiom $(\oplus^{b-1}\Phi, \varphi)$ of Q'''_1 to form its derivation from cedents $(\oplus^{b-1}(\Phi, \varphi))$, $(\varphi, \overline{\varphi})$, and $(\oplus^{b-1}\Phi, \oplus^b\Phi)$. Call the resulting refutation P'. Because of the assumption $\mathbf{cs}(Q_1) \leq \mathbf{cs}(Q_2)$, we have $\mathbf{cs}(P') \leq \mathbf{cs}(Q'_2) + 2\mathbf{cs}(Q'_1) + 3k + 17 \leq 5\mathbf{s}(Q_2)\mathbf{cs}(Q_2) + 10\mathbf{s}(Q_1)\mathbf{cs}(Q_1) + 3k + 17 \leq 5(\mathbf{s}(Q_2) + \mathbf{s}(Q_1) + \mathbf{s}(\Xi))(1 + \mathbf{cs}(Q_2) + \mathbf{cs}(Q_1)) = 5\sigma\tau$. The $\oplus$-width of Q'''_1 is one greater than the $\oplus$-width of Q'_1, hence it is at most $\log\tau + p + 3$. This number also bounds the $\oplus$-width of the additional derivations of axioms of Q''_1 and Q'''_1. The remaining requirements on P' are easy to check.

In the remaining cases, when Ξ is derived by Cut, Add, or Subtract, we can utilize the ideas used in the case of MOD, and the derivations are in fact simpler, so we will omit them.

Taking for P the refutation from the statement of the theorem, we thus obtain a $\mathrm{PK}_d^{\log\tau+p+3}(\oplus)$-refutation P' of $\mathcal{A}^P$ of cedent-size $5\sigma\tau$, such that each line has size at most $2\sigma^2$. Attach to each axiom of P' at most two cedents forming its derivation from $\mathcal{A} \cup \{(\varphi, \overline{\varphi}) : \varphi \text{ is an element of a cedent in } P\}$. Finally, derive each axiom of the form $(\varphi, \overline{\varphi})$ using Lemma 5. The resulting $\mathrm{PK}_d^{\log\tau+p+3}(\oplus)$-refutation has size $\leq 10\sigma\tau \cdot \sigma^2 + 5\sigma\tau \cdot O(\sigma^4) = O(\sigma^6)$. □

Theorem 14 implies:

Corollary 15. *For every d there is some $\epsilon > 0$ such that formulas expressing the counting principle* Count_3^n *require 2^{n^ϵ}-size refutations in treelike* $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$.

Proof. For an appropriate δ, a 2^{n^δ} lower bound on $\mathrm{PK}_d^{\log}(\oplus)$ refutations of Count_3^n can be obtained by combining an argument analogous to that of [11] with the degree lower bounds of [7]. □

Remark. It would be possible to prove Corollary 15 via a quasipolynomial simulation of treelike $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$ by daglike $\mathrm{PK}_{O(1)}^3(\oplus)$. To obtain the simulation, one combines the simulation of Theorem 14 with the translation $(\cdot)^{\mathrm{one}\oplus}$ of Section 5. The resulting sequence of $\mathrm{PK}_{O(1)}^1(\oplus)$ cedents can be made into a $\mathrm{PK}_{O(1)}^3(\oplus)$-derivation by adding a polysize derivation of the $(\cdot)^{\mathrm{one}\oplus}$-translation of the conclusion of each inference from the $(\cdot)^{\mathrm{one}\oplus}$-translations of the premises.

Remark. We do not expect that a result analogous to Corollary 15 can be easily obtained for daglike $\mathrm{PK}_{O(1)}^{\mathrm{id}}(\oplus)$. In fact, already bounds for a subsystem of $\mathrm{PK}_2^{\mathrm{id}}(\oplus)$ in which the inputs to $\oplus$ are log-sized conjunctions will probably be hard to prove. That system corresponds to the apparently strong bounded arithmetic theory $T_2^{2,\oplus\mathrm{P}}(\alpha)$ [10], which has not been separated from full bounded arithmetic with parity quantifiers. In particular, the system has quasipolynomial-size refutations of the surjective weak pigeonhole principle for functions defined in terms of $\oplus$'s of log-sized conjunctions—a principle which seems to be a major source of the strength of $\mathrm{AC}^0[2]$-Frege.

Acknowledgement. The authors are grateful to Nathan Segerlind for some clarifications concerning the status of [15] and [22].

References

[1] Albert Atserias, Moritz Müller, and Sergi Oliva. Lower bounds for DNF-refutations of a relativized weak pigeonhole principle. *Journal of Symbolic Logic*, 80(2):450–476, 2015.

[2] Paul Beame. A switching lemma primer. Unpublished, available at homes.cs.washington.edu/∼beame/papers, 1994.

[3] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, and Pavel Pudlák. Lower bounds on Hilbert's Nullstellensatz and propositional proofs. *Proceedings of the London Mathematical Society*, 73(3):1–26, 1996.

[4] Paul Beame and Søren Riis. More on the relative strength of counting principles. In P. Beame and S. Buss, editors, *Proof Complexity and Feasible Arithmetics*, pages 13–36. American Mathematical Society, 1997.

[5] Arnold Beckmann and Jan Johannsen. An unexpected separation result in linearly bounded arithmetic. *MLQ. Mathematical Logic Quarterly*, 51(2):191–200, 2005.

[6] S. R. Buss. Towards NP-P via proof complexity and search. *Annals of Pure and Applied Logic*, 163(7):906–917, 2012.

[7] Sam Buss, Dima Grigoriev, Russell Impagliazzo, and Toniann Pitassi. Linear gaps between degrees for the polynomial calculus modulo distinct primes. *Journal of Computer and System Sciences*, 62(2):267–289, 2001.

[8] Samuel R. Buss. Lower bounds on Nullstellensatz proofs via designs. In *Proof complexity and feasible arithmetics (Rutgers, NJ, 1996)*, volume 39 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 59–71. Amer. Math. Soc., Providence, RI, 1998.

[9] Samuel R. Buss, Russell Impagliazzo, Jan Krajíček, Pavel Pudlák, Alexander A. Razborov, and Jiří Sgall. Proof complexity in algebraic systems and bounded depth Frege systems with modular counting. *Computational Complexity*, 6:256–298, 1996/1997.

[10] Samuel R. Buss, Leszek Aleksander Kołodziejczyk, and Konrad Zdanowski. Collapsing modular counting in bounded arithmetic and constant depth propositional proofs. *Transactions of the American Mathematical Society*, 367(11):7517–7563, 2015.

[11] Jan Krajíček. Lower bounds for a proof system with an expontential speed-up over constant-depth Frege systems and over polynomial calculus. In *Proceedings of MFCS '97*, volume 1295 of *Lecture Notes in Computer Science*, pages 85–90. Springer, 1997.

[12] Jan Krajíček. Randomized feasible interpolation and monotone circuits with a local oracle, 2016. Preprint, `https://arxiv.org/abs/1611.08680`.

[13] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Groebner basis algorithm to find proofs of unsatisfiability. In *Proceedings of STOC 1996*, pages 174–183. ACM, 1996.

[14] Nicola Galesi and Massimo Lauria. Optimality of size-degree tradeoffs for polynomial calculus. *ACM Transactions on Computational Logic*, 12(1):Article 4, 2010.

[15] Russell Impagliazzo and Nathan Segerlind. Counting axioms do not polynomially simulate counting gates. In *Proceedings of FOCS 2001*, pages 200–209. IEEE, 2001.

[16] Russell Impagliazzo and Nathan Segerlind. Constant-depth Frege systems with counting axioms polynomially simulate Nullstellensatz refutations. *ACM Transactions on Computational Logic*, 7(2):199–218, 2006.

[17] Dmitry Itsykson and Dmitry Sokolov. Lower bounds for splittings by linear combinations. In E. Csuhaj-Varjú, M. Dietzfelbinger, and Z. Ésik, editors, *Proceedings of MFCS 2014, Part II*, volume 8635 of *Lecture Notes in Computer Science*, pages 372–383. Springer, 2014.

[18] A. Maciel, T. Pitassi, and A. R. Woods. A new proof of the weak pigeonhole principle. *Journal of Computer and System Sciences*, 64:843–872, 2002.

[19] Alexis Maciel, Phuong Nguyen, and Toniann Pitassi. Lifting lower bounds for tree-like proofs. *Computational Complexity*, 23(4):585–636, 2014.

[20] J. B. Paris, A. J. Wilkie, and A. R. Woods. Provability of the pigeonhole principle and the existence of infinitely many primes. *Journal of Symbolic Logic*, 53:1235–1244, 1988.

[21] A. A. Razborov. Lower bounds for the polynomial calculus. *Computational Complexity*, 7:291–324, 1998.

[22] Nathan Segerlind. *New Separations in Propositional Proof Complexity.* PhD thesis, University of California, San Diego, 2003.

[23] Gunnar Stålmarck. Short resolution proofs for a sequence of tricky formulas. *Acta Informatica*, 33(3):277–280, 1996.