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 7, 2018

Abstract

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

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. On the other hand, we prove that treelike $\mathrm{PK}_{O(1)}^{O(1)}(\oplus)$ is quasipolynomially but

[*] 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.

not polynomially equivalent to constant-depth systems with parity axioms. This leads us to ask the question whether any of our systems is quasipolynomially equivalent to $\mathrm{AC}^0[2]$-Frege on De Morgan formulas.

We also study Itsykson and Sokolov's proof 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 $\mathrm{AC}^0[2]$-Frege.

Here $\mathrm{AC}^0[2]$-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 [7] 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 families of subsystems of $\mathrm{AC}^0[2]$-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 is AC^0-Frege with parity axioms, which requires exponential size to prove the counting mod 3 principle Count_3 [10] and the pigeonhole principle PHP [5]. Two other families consist of the treelike and daglike versions of systems studied by Krajíček [20] which we call $\mathrm{PK}^c_d(\oplus)$, where $c, d \in \mathbb{N}$ are constants. The intuitive idea behind $\mathrm{PK}^c_d(\oplus)$ is that lines of the proof are $\oplus$'s of constant-depth De Morgan formulas, but insisting on that particular restricted form of lines would require inference rules that are hard to understand. In [20], lines were obtained by substituting $\oplus$'s of De Morgan formulas of constant *depth* (as measured by d) into De Morgan formulas of constant *size* (as specified by c). We use a somewhat modified definition in which lines are disjunctions (cedents) of constant-depth De Morgan formulas and a constant *number* of $\oplus$'s of constant-depth De Morgan formulas (see Section 2 for details). Both in Krajíček's version and in ours, a line can be translated into an equivalent single $\oplus$ at the cost of a polynomial increase in size (the depth of inputs to $\oplus$ will increase by a constant).

Already the treelike version of $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ polynomially simulates AC^0-Frege with parity axioms (see Section 3 below). On the other hand, by [20], treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ needs exponential size to prove PHP. Moreover, daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ needs exponential size to prove Count_3 ([20] combined with [8]).

It is natural to ask whether lower bounds for systems such as these could "already imply" lower bounds for $\mathrm{AC}^0[2]$-Frege. For instance, one could envision a scenario in which the already known result on the unprovability of Count_3 (or some other principle not mentioning $\oplus$) in the bounded arithmetic theory corresponding to $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ is extended to a strengthening of the theory by a suitable variant of the weak pigeonhole principle for functions defined by formulas with no nested occurrences of $\oplus$. By [11], this would give an unprovability result for the theory corresponding to $\mathrm{AC}^0[2]$-Frege. If the unprovability result is obtained in a sufficiently general way (i.e. holds for a sufficiently large class of models), it will actually yield the elusive lower bound for $\mathrm{AC}^0[2]$-Frege.

In the authors' opinion, scenarios like the one above are probably wishful thinking, since $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ is likely to be much weaker than $\mathrm{AC}^0[2]$-Frege, even as a system for proving DNF's/refuting CNF's. However, what concrete arguments can one give to justify this opinion? The ideal argument would be to exhibit a family of CNF's (or at least formulas in the language of $\neg, \wedge, \vee$) giving a strong separation of $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ from $\mathrm{AC}^0[2]$-Frege. Note that if no such family of tautologies exists, then $\mathrm{AC}^0[2]$-Frege has no short proofs of Count_3.

So far, there is only one method known to be useful for separating $\mathrm{AC}^0[2]$-Frege and AC^0-Frege with parity axioms—namely, the switching lemma-based technique used by Impagliazzo and Segerlind [16] to exhibit a family of CNF's with polysize refutations in $\mathrm{AC}^0[2]$-Frege but not in AC^0-Frege with parity axioms. Unfortunately, this technique has a serious limitation, in that it can only prove barely superpolynomial separations.

In this paper, we verify that the method of [16] can be adapted to give a superpolynomial separation between $\mathrm{AC}^0[2]$-Frege and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, considered as systems for refuting CNF's. However, our main result goes in the other direction: 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. Our argument is inspired by another paper of Impagliazzo and Segerlind [17], which proves that AC^0-Frege with parity axioms simulates (in a specific technical sense) the algebraic proof system known as Nullstellensatz over $\mathbb{F}_2$. The idea of [17] is to show that if the

$$\begin{array}{c}
\mathrm{AC}^0[2]\text{-Frege} \\
\rotatebox{90}{$<$}^{p} \qquad \rotatebox{90}{$\leqslant$}^{qp} \\
\text{daglike } \mathrm{PK}^{O(1)}_{O(1)}(\oplus) \\
\rotatebox{90}{$\leqslant$}^{p} \qquad \rotatebox{90}{$\leqslant$}^{qp} \\
\text{treelike } \mathrm{PK}^{O(1)}_{O(1)}(\oplus) \\
\rotatebox{90}{$<$}^{p} \qquad \rotatebox{90}{$\equiv$}^{qp} \\
\mathrm{AC}^0\text{-Frege w/ parity axioms}
\end{array}$$

Figure 1: Known simulations between some subsystems of $\mathrm{AC}^0[2]$-Frege w.r.t proving formulas without $\oplus$. Quasipolynomial simulations denoted by $\leqslant^{qp}$. Use of $<$ rather than $\leqslant$ indicates simulation is known not to reverse.

algebraic translation of a Boolean formula φ has a low-degree Nullstellensatz refutation, then there is a small AC^0-Frege proof that φ implies the existence of perfect matchings on two sets differing by exactly one element. Our basic idea is similar, but the argument is considerably more complicated, as the definitions of the matchings now mimic the structure not of a Nullstellensatz proof (a single polynomial equation), but a treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ proof, which consists of multiple lines derived from one another by means of various inference rules formalizing both algebraic and Boolean reasoning.

The current state of knowledge about the (families of) subsystems of $\mathrm{AC}^0[2]$-Frege we consider, viewed as systems for proving formulas in the De Morgan language, can therefore be summarized in Figure 1. The striking feature of Figure 1 is the lack of any *superquasipolynomial* separations—at this point, even a quasipolynomial simulation of $\mathrm{AC}^0[2]$-Frege by AC^0-Frege with parity axioms has not yet been ruled out.

In the context of constant-depth systems, it often seems that the "right" notion of simulation is at least quasipolynomial rather than polynomial. For instance, the usual translation from bounded arithmetic to constant-depth proofs is quasipolynomial, so collapses in arithmetic translate into quasipolynomial simulations. In general the area abounds in results on quasipolynomial-size provability and/or simulation that are either open or false in the polysize case, e.g. [24, 15, 6, 1, 11]. Of course, if $\mathrm{AC}^0[2]$-Frege systems were to be quasipolynomially equivalent to, say, $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$—a seemingly farfetched scenario, but not disproved—then the known lower bounds for the latter would imply strong lower bounds also for the former. Thus, one of the aims of this paper is to encourage the study of the following open problem:

Open Problem. Prove a *superquasipolynomial* separation between $\mathrm{AC}^0[2]$-Frege and a subsystem containing AC^0-Frege with parity axioms on a family of formulas without $\oplus$.

The paper is organized as follows. We introduce the necessary definitions and background in Section 2. In Section 3, we prove some basic properties of $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, including the polynomial simulation of parity axioms and the fact that treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ can be balanced at the cost of allowing logarithmically rather than constantly many $\oplus$'s per line. Section 4 introduces and studies an auxiliary proof system which allows just one $\oplus$ per line, but has some additional rules.

Our main result, the quasipolynomial simulation of treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ by AC^0-Frege with parity axioms, is proved in Section 5. The proof is preceded by a rather detailed informal overview of the argument, which also explains the role played by the one-parity system.

In Section 6, we give easy proofs of exponential separations between treelike and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ and $\mathrm{AC}^0[2]$-Frege as refutation systems for formulas with $\oplus$. We also outline a proof of the superpolynomial separations without $\oplus$. Some more details on that argument are given in a separate Appendix.

In Section 7, we consider the loosely related topic of a refutation system introduced by Itsykson and Sokolov [18] in which proof lines are disjunctions of parities of literals. In [18], a lower bound on the treelike version of this system is proved. We point out that even a generalization of the treelike system in which the parities can have arbitrary constant-depth De Morgan formulas as inputs is unable to give short proofs of Count_3. We also comment on why such a result might be difficult to obtain for the daglike case.

2 Preliminaries

For a number n, the symbol $[n]$ stands for the set $\{1, \dots, n\}$. For a set S, the symbol $\binom{S}{2}$ stands for the set of two-element subsets of S. For a sequence S, the symbol $\mathrm{lh}(S)$ stands for the length of S.

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 (rather than set) 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 that of the disjunction $\bigvee\Gamma$.

The logical axioms are:

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

for a propositional variable x.

The inference rules are:

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

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

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

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

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

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 [23].

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-number* of P, denoted by $\mathbf{cn}(P)$, 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 $\mathrm{PK}_d^c(\oplus)$, where c, d are natural number constants. The systems $\mathrm{PK}_d^c(\oplus)$ are a variant of the family of proof systems introduced in [20][1]. A PK($\oplus$)-derivation is called a $\mathrm{PK}_d^c(\oplus)$*-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 $\oplus^b\Psi$, where $b \in \{0, 1\}$, and no other $\oplus$ connectives appear. In other words, a line in a $\mathrm{PK}_d^c(\oplus)$-derivation has the form

$$\varphi_1, \ldots, \varphi_\ell, \oplus^{b_1}\Psi_1, \ldots, \oplus^{b_k}\Psi_k,$$

where $k \leq c$, the φ_i's are formulas of depth $\leq d$ without $\oplus$, and the Ψ_j's are sequences of formulas of depth $\leq d-1$ without $\oplus$.

We can generalize the systems $\mathrm{PK}_d^c(\oplus)$ to $\mathrm{PK}_d^f(\oplus)$ for various functions $f\colon \mathbb{N} \to \mathbb{N}$. In a $\mathrm{PK}_d^f(\oplus)$ derivation P, each cedent may contain up to $f(\mathbf{s}(P))$ parities; the other syntactic conditions are as for $\mathrm{PK}_d^c(\oplus)$. For instance, if $\mathrm{id}\colon \mathbb{N} \to \mathbb{N}$ is the identity function, a line in the system $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$

[1] The systems $\mathrm{F}_d^c(\mathrm{MOD}_p)$ of [20] 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, [20] 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 15, 16, and 17—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.

may again look like

$$\varphi_1, \dots, \varphi_\ell, \oplus^{b_1}\Psi_1, \dots, \oplus^{b_k}\Psi_k,$$

with the same restrictions as for $\mathrm{PK}^c_d(\oplus)$, except that k can now be arbitrary. In a $\mathrm{PK}^{\log}_d(\oplus)$ derivation, we could use disjunctions/cedents of at most logarithmically (in the derivation size) many parities per line.

Subsystems of $\mathrm{PK}(\oplus)$ 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 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 elements of $\binom{[n]}{2}$, 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 since cedents are interpreted as disjunctions, the axiom says that the φ_e's do not define a partition of the odd-sized set $[n]$ into two-element subsets: either there is some element $i \in [n]$ such that $\neg\varphi_e$ holds for each subset e containing i, or there are two overlapping sets e and f such that φ_e and φ_f both hold.

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 total degree of a line in it as a formal polynomial. A refutation is a derivation whose last line is the constant polynomial 1.

3 Basic properties

The purpose of this section is to verify that the $\mathrm{PK}_d^c(\oplus)$ systems have some basic desirable properties.

3.1 Completeness

The first property we need to check is that each $\mathrm{PK}_d^c(\oplus)$ is a complete proof system, in a reasonably strong sense of the term.

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

Proof. Let $\mathcal{A}$ be a set $\mathrm{PK}_d^c(\oplus)$ cedents and Γ a cedent such that $\mathcal{A} \models \Gamma$. We prove that Γ can be derived from $\mathcal{A}$ using a somewhat technical claim.

Claim. There exists a $\mathrm{PK}_d^c(\oplus)$ derivation of Γ from some set of cedents $\{\Delta_1, \ldots, \Delta_n\}$ such that for each $i = 1, \ldots, n$, the cedent Δ_i has the following properties:

(i) Δ_i either is the logical axiom $\oplus^0\emptyset$ or contains no formulas with $\oplus$,

(ii) $\Gamma \models \Delta_i$,

(iii) at most one assignment to the variables of $\mathcal{A} \cup \{\Gamma\}$ falsifies Δ_i.

Proof of the Claim. Let $p_1, \ldots, p_m$ be all the variables appearing in $\mathcal{A} \cup \{\Gamma\}$. We build the claimed $\mathrm{PK}_d^c(\oplus)$ derivation backwards from the endcedent Γ.

First, using a series of cuts on $p_1, \ldots, p_m$, we derive Γ from the set of cedents of the form

$$\Gamma, p_1^{b(1)}, p_2^{b(2)}, \ldots, p_m^{b(m)}, \tag{1}$$

for all $b \in \{0,1\}^n$, where p_j^0 is p_j and p_j^1 is $\overline{p_j}$. Such cedents clearly satisfy property (ii), since they extend Γ, and property (iii), since there is only one assignment falsifying $p_1^{b(1)} \vee \ldots \vee p_m^{b(m)}$.

Using a series of MOD inferences (and exchanges), each cedent of the form (1) is derived from a set of $\mathrm{PK}_d^c(\oplus)$ cedents containing no formulas with $\oplus$ other than $\oplus^0\emptyset$ or $\oplus^1\emptyset$. The conclusion of a MOD inference implies each of its premises, so 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 preserves properties (ii), (iii) and also gives (i), thus proving the claim. □

Now assume $\mathcal{A} \models \Gamma$ and let $\Delta_1, \ldots, \Delta_n$ be a set of cedents with the properties from the claim. It suffices to describe a derivation of Δ_i from $\mathcal{A}$ for each i. Property (ii) implies $\mathcal{A} \models \Delta_i$, and property (iii) implies that Δ_i either is tautological or has exactly one falsifying assignment. In the former case, property (i) and the completeness of $\mathrm{PK}^0_d(\oplus)$ ensure that Δ_i has a $\mathrm{PK}^c_d(\oplus)$ derivation without any non-logical axioms.

In the latter case, since $\mathcal{A} \models \Delta_i$, the unique falsifying assigment for Δ_i must also falsify some cedent $\Xi \in \mathcal{A}$, and therefore $\Xi \models \Delta_i$. Let Ξ be $\xi_1, \ldots, \xi_k$. Using a series of cuts on $\xi_1, \ldots, \xi_k$, we can derive Δ_i from the cedent Δ_i, Ξ and cedents of the form $\Delta_i, \xi_1, \ldots, \xi_{j-1}, \bar{\xi_j}$ for $j = 1, \ldots, k$. This is a $\mathrm{PK}^c_d(\oplus)$ derivation because Δ_i contains no $\oplus$'s and there are at most c parities among $\xi_1, \ldots, \xi_k$. Now, Δ_i, Ξ follows from Ξ by weakening. On the other hand, the claim implies that each of the tautological cedents $\Delta_1, \xi_1, \ldots, \xi_{j-1}, \bar{\xi_j}$ 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. □

3.2 Short proofs of some tautologies

Another property we check is 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. In proving the proposition, we will make use of the following lemma, the proof of which is left as an exercise to the reader.

Lemma 2. *There is a polytime procedure which given a depth-d formula $\oplus^b(\Gamma, \varphi, \psi, \Delta)$ with no nesting of $\oplus$ and the index of φ among the inputs to $\oplus^b$ produces a treelike $\mathrm{PK}^3_d(\oplus)$ derivation of this formula from $\oplus^b(\Gamma, \psi, \varphi, \Delta)$.*

Proposition 3. *For each $d \geq 1$, there is a polynomial-time procedure which, given a $\mathrm{PK}^0_d(\oplus)$ cedent Γ which is a parity axiom, outputs a treelike $\mathrm{PK}^3_d(\oplus)$ derivation $P(\Gamma)$ of Γ.*

Proof. Let Γ be

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

Recall that the indices e and f range over $\binom{[n]}{2}$, the set of two-element subsets of $[n]$. Note also that each φ_e contains no parity connectives and that it has depth at most $d-1$ (otherwise Γ would not be a $\mathrm{PK}^0_d(\oplus)$ cedent).

We will show that treelike $\mathrm{PK}_d^3(\oplus)$ can carry out a standard double counting argument in polynomial size.

Informally, if $G = ([n], \{e \in \binom{[n]}{2} : \varphi_e \text{ is satisfied}\})$ is a graph, we can count the sum of the degrees of all its vertices in two different ways. First, we count edge-by-edge and we get an even number; in symbols, we obtain the cedent $\oplus^0(\varphi_e : e \in \binom{[n]}{2}, i \in e)$. Second, assuming Γ fails (which means the edges of G form a perfect matching on $[n]$), we count vertex-by-vertex and, because n is odd, we end up with an odd number; that is, we obtain the cedent $\Gamma, \oplus^1(\varphi_e : i \in [n], e \ni i)$. We conclude that Γ is true. The details of our formalization of this argument in treelike $\mathrm{PK}_d^0(\oplus)$ follow.

The final inference of $P(\Gamma)$ is:

$$\frac{\Gamma, \oplus^0(\varphi_e : i \in [n], e \ni i) \qquad \Gamma, \oplus^1(\varphi_e : i \in [n], e \ni i)}{\Gamma} \text{ Cut}$$

The cedent $\Gamma, \oplus^0(\varphi_e : i \in [n], e \ni i)$ follows by weakening from the cedent $\oplus^0(\varphi_e : i \in [n], e \ni i)$, which can be derived by permuting the formulas inside $\oplus^0$ (using Lemma 2 repeatedly) from $\oplus^0(\varphi_e : e \in \binom{[n]}{2}, i \in e)$. The latter cedent is obtained by a tree of additions from the cedents $\oplus^0(\varphi_e, \varphi_e)$ for $e \in \binom{[n]}{2}$, each of which has an easy treelike polynomial-size proof.

On the other hand, the cedent $\Gamma, \oplus^1(\varphi_e : i \in [n], e \ni i)$ can be derived by a tree of additions from the cedents of the form $\Gamma, \oplus^1(\varphi_e : e \ni i)$ for $i \in [n]$ (recall that n is odd!). Thus, for each fixed $i \in [n]$ we need to give a treelike polynomial-size proof of the cedent $\Gamma, \oplus^1(\varphi_e : e \ni i)$.

Each φ_e for $e \ni i$ is actually $\varphi_{\{i,j\}}$ for some $j \in [n] \backslash \{i\}$. The cedents $\overline{\varphi_{\{i,j\}}}, \oplus^1(\varphi_{\{i,j\}})$ and $\varphi_{\{i,j\}} \wedge \varphi_{\{i,k\}}, \overline{\varphi_{\{i,j\}}}, \oplus^0(\varphi_{\{i,k\}})$ for $k \neq j$ are very easy to derive. Some weakenings and a tree of additions give

$$\Gamma, \overline{\varphi_{\{i,j\}}}, \oplus^1(\varphi_e : e \ni i)$$

for each j. The desired $\Gamma, \oplus^1(\varphi_e : e \ni i)$ follows by an AND inference and a contraction. □

The following technical lemma shows that $\mathrm{PK}_d^0(\oplus)$ proves some basic tautologies in a reasonable size and height. The lemma will be used many times throughout the paper.

Lemma 4. *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 $\mathrm{PK}_d^0(\oplus)$-derivation of $\varphi, \overline{\varphi}$ of height $O(d)$ and size $O(S_\varphi^2)$,

(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 copies of $\oplus^0\Phi_1, \oplus^1\Phi_1$ and three of $\oplus^0\Phi_2, \oplus^1\Phi_2$. Iterating this $\lceil \log \ell \rceil$ times gives a height $O(\log \ell)$ size $O(S\ell^{\log 3})$ treelike $\mathrm{PK}^3_d(\oplus)$-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). □

3.3 Balancing

To conclude the section, 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.

Lemma 5. *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* s *and cedent-number* t*. Then there is a treelike* $\mathrm{PK}^{2c+\log t}_{d+1}(\oplus)$*-derivation of* Γ *from* $\mathcal{A}$ *which has height* $O(d + c\log s)$ *and size* $s^{\log O(c)}$*.*

With the help of Lemma 4 from the previous subsection, Lemma 5 follows easily from Lemma 8, which is stated and proved below. The proof of Lemma 8 is based on the usual divide-and-conquer idea dating back to Spira [27], but some effort is required to get the technical details right.

Definition 6. Let P be a $\mathrm{PK}^c_d(\oplus)$-derivation. Define $\mathsf{sa}(P)$ to be the set of cedents having one of the following forms:

(i) Λ, Ξ, where Ξ is an initial cedent of P and Λ is arbitrary,

(ii) $\Lambda, \overline{\oplus}^a\Phi, \Lambda', \Psi$, where Ψ is a cedent in P such that $\oplus^a\Phi$ is an element of Ψ, and Λ, Λ' are arbitrary,

(iii) $\Lambda, \bigwedge_{i=1}^{\ell} \overline{\gamma_i}, \Lambda', \Psi$, where Ψ is a cedent in P such that $\gamma_1, \ldots, \gamma_\ell$ is the subsequence of Ψ consisting of all formulas that do not contain any parity connective, and Λ, Λ' are arbitrary.

Definition 7. Define $P^{\textsf{-ic}}$ to be P with all initial cedents removed from it.

Recall that $\mathbf{s}(P)$, the size of the derivation P, denotes the number of symbols in it, and that $\mathbf{cn}(P)$, the cedent-number of P, denotes the number of occurrences of cedents in P.

Lemma 8. *Suppose that $c \geq 1$ and P is a treelike $\mathrm{PK}^c_d(\oplus)$-derivation of Γ from $\mathcal{A}$. Suppose that P has size s and cedent-number t. Then there is a treelike $\mathrm{PK}^{2c+\log(t-1)}_{d+1}(\oplus)$-derivation P' of Γ from $\mathsf{sa}(\mathrm{P})$ such that the height of P' is at most $(2c+6)\log(t-1)$, cedent-number of P' is at most $t^{\log(2c+2)}$, and the size of each cedent in P' is at most $s + 2\cdot\mathbf{s}(P^{\textsf{-ic}})$.*

Proof. We shall prove the lemma by induction on t. (In this proof we write $\log t$ to mean $\max\{1, \log_2 t\}$.)

The lemma is obviously true for $t \leq (2c+6)\log(t-1)$. In the inductive step, we take an inference I in P with the following properties:

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

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

(I can be found by constructing a path through P from the root upwards that eventually reaches an inference such that (b) is satisfied by its conclusion but not by any of its premises.) Denote

$$v := \log(2c+2)$$

By the induction hypothesis, for each premise Π of the inference I there is a treelike $\mathrm{PK}^{2c+\log(t-1)-1}_{d+1}(\oplus)$-derivation of Π from $\mathsf{sa}(P_\Pi)$ which has height $\leq (2c+6)(\log(t-1)-1)$, cedent-number $\leq \mathbf{cn}(P_\Pi)^v$, and has the size of each cedent bounded by $\mathbf{s}(P_\Pi) + 2\cdot\mathbf{s}(P_\Pi^{\textsf{-ic}})$. Put these derivations together with the inference I, and let Q denote the resulting derivation. Q is a treelike $\mathrm{PK}^{2c+\log(t-1)-1}_{d+1}(\oplus)$-derivation of Ψ from $\mathsf{sa}(P)$ with height

$$\leq (2c+6)(\log(t-1)-1)+1 = (2c+6)\log(t-1) - 2c - 5.$$

The size of each cedent in Q is bounded by $s + 2\cdot\mathbf{s}(P^{\textsf{-ic}})$, and we bound the cedent-number of Q as follows:

$$\mathbf{cn}(Q) \leq 1 + \sum_\Pi \mathbf{cn}(P_\Pi)^v \leq 1 + (\sum_\Pi \mathbf{cn}(P_\Pi))^v \leq 1 + (t - \mathbf{cn}(D))^v,$$

where Π in the sums ranges over the premises of the inference I.

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 $1 \leq \mathbf{cn}(D) \leq t/2$. In the case where $\mathbf{cn}(D) = 1$, we have $\Psi = \Gamma$, and we define the desired derivation P' to be just Q. Assume $\mathbf{cn}(D) > 1$. By the induction hypothesis, there is a treelike $\mathrm{PK}_{d+1}^{2c+\log(t-1)-1}(\oplus)$-derivation R of Γ from $\mathsf{sa}(D)$ of height $\leq (2c+6)(\log(t-1)-1)$, cedent-number $\leq \mathbf{cn}(D)^v$, and the size of each cedent in R is at most $\mathbf{s}(D) + 2 \cdot \mathbf{s}(D^{\text{-}\mathsf{ic}})$.

The idea now is to transform R into several derivations from the hypotheses $\mathsf{sa}(P)$ and combine them with Q by a repeated use of the cut rule to derive Γ. To this end, we introduce some notation. 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 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.

R' (resp. R_i, $i = 1, \ldots, k$) is constructed by adding the formula $\bigwedge \Delta'$ (resp. $\overline{\oplus}^{a_i}\Phi_i$) to the left of each cedent in R, and by applying one weakening and one exchange to the endcedent $\bigwedge \Delta', \Gamma$ (resp. $\overline{\oplus}^{a_i}\Phi_i, \Gamma$) to derive $\Gamma, \Theta_k, \bigwedge \Delta'$ (resp $\Gamma, \Theta_{i-1}, \overline{\oplus}^{a_i}\Phi_i$). Note that in the process of constructing R' and R_i, all initial cedents in R contained in $\mathsf{sa}(D) \setminus \mathsf{sa}(P)$ are replaced by cedents in $\mathsf{sa}(P)$.

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 2. Notice that just below the endcedent of Q we inserted some exchanges, a weakening, and an $\bigvee$-inference to make the cedent ready for applications of the cut rule.

It is easy to verify that P' is a treelike $\mathrm{PK}_{d+1}^{2c+\log(t-1)}(\oplus)$-derivation of Γ from $\mathsf{sa}(P)$ and the height of P' is at most $(2c+6)\log(t-1)$. Also, the size of each cedent in P' is at most $\mathbf{s}(P) + 2 \cdot \mathbf{s}(P^{\text{-}\mathsf{ic}})$ (due to the bound on the size of cedents in R and the fact that Ψ is in $P^{\text{-}\mathsf{ic}} \setminus D^{\text{-}\mathsf{ic}}$). The same conclusions are of course true in the case $\mathbf{cn}(D) = 1$ (which has $P' = Q$).

It remains to bound the cedent-number of P'. Each of R', R_i, $i = 1, \ldots, k$, has cedent-number $\leq \mathbf{cn}(R) + 2$. Hence, by construction,

$$\begin{aligned} \mathbf{cn}(P') &\leq \mathbf{cn}(Q) + (c+1)(\mathbf{cn}(R) + 2) + 2c + 4 \\ &\leq (t - \mathbf{cn}(D))^v + (c+1)\mathbf{cn}(D)^v + 4c + 7. \end{aligned} \tag{2}$$

This is true in the case $\mathbf{cn}(D) = 1$ as well. Recall that $v = \log(2c+2)$. Consider the last expression in (2) as a function of $\mathbf{cn}(D)$ and denote it

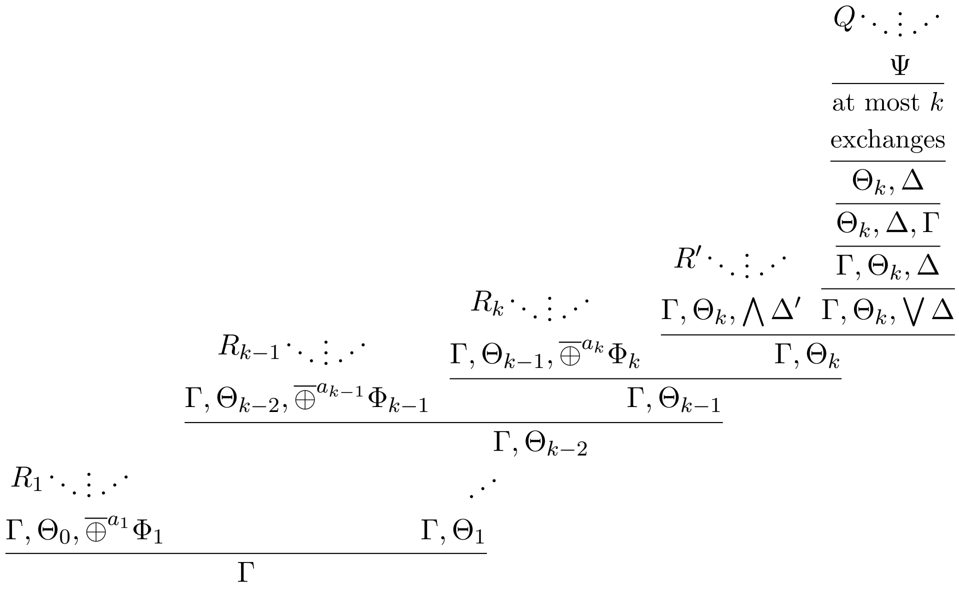

Q
Ψ
at most k exchanges
Θ_k, Δ
Θ_k, Δ, Γ
R'
Γ, Θ_k, Δ
R_k
$\Gamma, \Theta_k, \bigwedge \Delta'$
$\Gamma, \Theta_k, \bigvee \Delta$
R_{k-1}
$\Gamma, \Theta_{k-1}, \overline{\bigoplus}^{a_k} \Phi_k$
Γ, Θ_k
$\Gamma, \Theta_{k-2}, \overline{\bigoplus}^{a_{k-1}} \Phi_{k-1}$
Γ, Θ_{k-1}
Γ, Θ_{k-2}
R_1
$\Gamma, \Theta_0, \overline{\bigoplus}^{a_1} \Phi_1$
Γ, Θ_1
Γ

Figure 2: The derivation P'.

by $f(\mathbf{cn}(D))$. It is not difficult to verify that $f(\mathbf{cn}(D))$ is convex on the interval $[1, \ldots, t/2]$. Thanks to our assumption that t is large enough ($t > (2c+6)\log(t-1)$), another straightforward calculation gives $f(1) > f(t/2)$. It remains to obtain an upper bound on $f(1)$.

$$\begin{aligned} t^v - (t-1)^v \geq t^v - t^v + vt^{v-1} - \binom{v}{2} t^{v-2} &= t^{v-2} v(t-(v-1)/2) \\ &= t^{\log(2c+2)-2} \log(2c+2)\,(t - \log(c+1)/2) \geq 5c+8. \end{aligned}$$

The last inequality follows from our assumption on t, and it gives $f(1) \leq t^v$. Thus, $\mathbf{cn}(P') \leq t^v$. □

Proof of Lemma 5. Take the derivation P' provided by Lemma 8. We need to derive the initial cedents of P' in the required size and height. Definition 6 lists their possible forms. Form (i) is derived easily from $\mathcal{A}$, form (ii) follows from Lemma 4 part (c), and form (iii) is obtained using Lemma 4 part (a). □

4 The one-parity system

In this section, we introduce an auxiliary proof system that only allows one $\oplus$ per line. The idea is that the system treats $\oplus$'s of constant-depth formulas more or less as polynomials, and we do have to introduce extra rules, including one analogous to polynomial multiplication, to make the system strong enough. However, to make the interaction between the polynomial-style rules and traditional logical rules clearer, we allow the lines in the system to contain some constant-depth formulas outside the unique $\oplus$.

Below, we prove two main results about the one-parity system. The first is that a $\mathrm{PK}^{\mathrm{id}}_{O(1)}(\oplus)$ derivation can be translated into a derivation in the one-parity system with an increase in size that depends on the number of $\oplus$'s per line in the original derivation: it is polynomial for a $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ derivation and quasipolynomial for the $\mathrm{PK}^{\log}_{O(1)}(\oplus)$ derivations produced by balancing. Moreover, the translation preserves treelikeness and, in the cases of our interest, leads only to a modest increase in height. The second result is that, given a derivation of $\oplus\Phi$ in the one-parity system, we can "delay subtraction until the end", i.e. avoid using the Subtract rule and derive $\oplus(\Theta, \Theta, \Phi)$ for some sequence of formulas Θ, at the cost of a blowup which is exponential in the *height* of the original derivation; in particular, quasipolynomial if the derivation had polylog height.

The role of these two results in our simulation of treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ by parity axioms is explained in some detail at the beginning of Section 5. The curious reader may want to take a look at that explanation before reading on.

Definition 9 (Conjuncts of a formula, product of sequences of formulas)**.** Let φ be a formula. Define $\mathsf{cnj}(\varphi)$ (the *sequence of conjuncts* of φ) to be Θ if φ is of the form $\bigwedge\Theta$, and to be the one-element sequence (φ) otherwise.

If Φ and Ψ are $(\varphi_i)_{i\in I}$ and $(\psi_i)_{i\in J}$, respectively, define $\Phi\times\Psi$ to be the sequence $(\bigwedge(\mathsf{cnj}(\varphi_i), \mathsf{cnj}(\psi_j)))_{i\in I, j\in J}$. Further, define recursively

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

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

The idea is that for any assignment, the number of satisfied formulas in $\prod_{i=1}^{k}\Phi_i$ is even exactly if the number of satisfied formulas in Φ_i is even for at least one i.

Example. Let Φ be (p_1, p_2) and let Ψ be $(p_3 \wedge p_4)$, where all the p_i are variables. Then $\Phi \times \Phi$ is the sequence $(p_1 \wedge p_1, p_1 \wedge p_2, p_2 \wedge p_1, p_2 \wedge p_2)$, while $\Phi \times \Psi$ is the sequence $(p_1 \wedge p_3 \wedge p_4, p_2 \wedge p_3 \wedge p_4)$.

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:

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

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

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

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

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

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

where $\pi(\Phi)$ is a permutation of the formulas in Φ. Note that in MOD inferences, φ is required to have $\bigwedge$ as its topmost connective (formulas of other forms have to be turned into single-argument conjunctions by AND inferences before entering inside $\oplus$).

Lemma 10. *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) $\emptyset, \oplus\left(\bigwedge(\Gamma,\Phi), \bigwedge(\Gamma,\Phi,\Phi)\right)$

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

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

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

Proof. (a) is proved in the same way as Lemma 4 (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). $\square$

4.1 Translating derivations into $\mathrm{PK}_{O(1)}^{\mathrm{one}\oplus}$

Definition 11 (Translation of cedent to $\mathrm{PK}_{O(1)}^{\mathrm{one}\oplus}$)**.** We map each cedent Γ of the system $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$ 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 subsequence 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 $a_i = 1$.

Example. If Γ is $\oplus^0\Phi, \oplus^1\Phi$ for Φ equal to (p_1, p_2), then $(\Gamma)^{\mathrm{one}\oplus}$ is

$$\emptyset, \oplus(p_1 \wedge p_1, p_1 \wedge p_2, p_1, p_2 \wedge p_1, p_2 \wedge p_2, p_2).$$

Note the additional copies of the p_i's compared to the example after Definition 9: these are formally one-argument conjunctions, and they arise as conjunctions of p_i with the empty sequence of conjuncts of $\top$.

Lemma 12. *Let* P *be a (treelike)* $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$*-derivation of* Ω *from* $\mathcal{A}$*. Suppose that* P *has size* s*, height* h*, and at most* t *parities per line. 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* $O(s^{2t})$ *and height* $O(h + d + t\log s)$*.*

Proof. Let P' be the result of replacing each cedent Γ of P with its translation $(\Gamma)^{\mathrm{one}\oplus}$. Note that the logical axioms in P translate into logical axioms of $\mathrm{PK}^{\mathrm{one}\oplus}_{d+1}$. To make P' into a $\mathrm{PK}^{\mathrm{one}\oplus}_{d+1}$-derivation, we need to derive, for each inference rule of $\mathrm{PK}^{\mathrm{id}}_{d}(\oplus)$, 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

$$\frac{\Gamma^{\mathsf{M}}, \oplus\left(\left(\prod_{i=1}^{k-2}\Gamma^{\mathsf{P}}_i\right)\times\Gamma^{\mathsf{P}}_{k-1}\times\Gamma^{\mathsf{P}}_k\right)}{\Gamma^{\mathsf{M}}, \oplus\left(\left(\prod_{i=1}^{k-2}\Gamma^{\mathsf{P}}_i\right)\times\Gamma^{\mathsf{P}}_k\right)} \tag{3}$$

where $\Gamma^{\mathsf{P}}_{k-1}=\Gamma^{\mathsf{P}}_k$. Denote $\prod_{i=1}^{k-2}\Gamma^{\mathsf{P}}_i$ by Θ and let Γ^{P}_k be $(\varphi_j)_{j\in I}$. By Lemma 10 (b), for each γ in Θ and $j\in I$, there is a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+1}$-derivation of

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

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

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

Call this derivation Q_1. Similarly, for each γ in Θ and $i<j\in I$, there is, by Lemma 10 (c), a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+1}$-derivation of

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

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 (3) by:

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

One of the two MOD rules translates into

$$\frac{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus\left(\left(\prod_{i=1}^{k}\Gamma_i^{\mathsf{P}}\right)\times\Phi\right) \qquad \Gamma^{\mathsf{M}}, \varphi, \oplus\left(\left(\prod_{i=1}^{k}\Gamma_i^{\mathsf{P}}\right)\times(\Phi,\top)\right)}{\Gamma^{\mathsf{M}}, \oplus\left(\left(\prod_{i=1}^{k}\Gamma_i^{\mathsf{P}}\right)\times(\varphi,\Phi,\top)\right)} \tag{4}$$

The other MOD rule translates into

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

Denote $\prod_{i=1}^{k}\Gamma_i^{\mathsf{P}}$ by Θ. Let us derive (4) first. We use Lemma 10 (d) and (e) to get for each γ in Θ a treelike $\mathrm{PK}_{d+1}^{\mathrm{one}\oplus}$-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. Using each of these two sets of derivations and a balanced tree of Adds, we 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 (4) by $\textcircled{L}$ and the right premise by $\textcircled{R}$. We can derive (4) by:

$$\dfrac{\text{Perm.}\ \dfrac{\text{Add}\ \dfrac{\text{Weak.}\ \dfrac{\begin{array}{c} Q_3\ \vdots \\ \overline{\varphi}, \oplus(\Theta\times(\varphi,\top)) \end{array}}{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus(\Theta\times(\varphi,\top))} \quad \textcircled{L}}{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus(\Theta\times(\varphi,\top), \Theta\times\Phi)}}{\Gamma^{\mathsf{M}}, \overline{\varphi}, \oplus(\Theta\times(\varphi,\Phi,\top))} \qquad \dfrac{\dfrac{\dfrac{\begin{array}{c} Q_4\ \vdots \\ \varphi, \oplus(\Theta\times(\varphi)) \end{array}}{\Gamma^{\mathsf{M}}, \varphi, \oplus(\Theta\times(\varphi))}\ \text{Weak.} \quad \textcircled{R}}{\Gamma^{\mathsf{M}}, \varphi, \oplus(\Theta\times(\varphi), \Theta\times(\Phi,\top))}\ \text{Add}}{\Gamma^{\mathsf{M}}, \varphi, \oplus(\Theta\times(\varphi,\Phi,\top))}\ \text{Perm.}}{\Gamma^{\mathsf{M}}, \oplus(\Theta\times(\varphi,\Phi,\top))}\ \text{Cut}$$

To derive (5), 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 (4) (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 10 (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 $\mathrm{PK}_{d+1}^{\mathrm{one}\oplus}$-derivation P''. Note that for any cedent Γ in P of size S the size of $(\Gamma)^{\mathrm{one}\oplus}$ is $O(S^t)$. 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 $O(\log(S^t))$ and size $O(S^t \log(S^t))$. To the leaves of this tree we appended derivations of height $O(d)$ and of total size $O(S^{2t})$. Hence, what we added to the translation of J has height $O(d + t \log S)$ and size $O(S^{2t})$. Thus, P'' has the required properties. □

4.2 Delaying subtractions

Lemma 13. *Let P be a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-derivation of $\Gamma, \oplus\Phi$ from $\mathcal{A}$. Suppose that P has size s and height h. Then for some sequence of formulas 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 $O(s^h)$ and height $O(h)$.*

Proof. The idea is extremely simple: instead of using Subtract to derive $\oplus\Phi$ from $\oplus(\Phi, \Psi)$ and $\oplus\Psi$, we use Add and derive $\oplus(\Phi, \Psi, \Psi)$. Some additional syntactic manipulations are needed to make this work.

We proceed by induction on height. The lemma is clear for $h = 1$, so 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 small 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

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

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:

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

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. □

5 Simulation by parity axioms

We are now ready to describe our quasipolynomial simulation of treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ by AC^0-Frege with parity axioms. More specifically, we obtain a polynomial simulation of $\mathrm{PK}^{\mathrm{one}\oplus}_{O(1)}$ refutations with delayed subtraction (i.e. derivations not using the Subtract rule and ending in $\oplus(\Theta, \Theta, \top)$ for some Θ) by constant-depth proofs with parity axioms. The quasipolynomial blowup appears at earlier steps: first, in translating the derivations with logarithmically many parities per line that arise from balancing into $\mathrm{PK}^{\mathrm{one}\oplus}_{O(1)}$ (Lemma 12), and second, in delaying subtractions (Lemma 13).

Before we present the technical aspects, let us give a relatively detailed overview of the intuition behind the construction. The main idea is as follows. We are given a $\mathrm{PK}^{\mathrm{one}\oplus}_{O(1)}$ derivation P from some set of axioms $\mathcal{A}$ that do not contain $\oplus$. For each line C in P, we construct a small AC^0-Frege derivation (*without* parity axioms) that uses $\mathcal{A}$ as non-logical axioms and derives a constant-depth formula γ^C that "directly witnesses that C is true". For the last line of P, namely $\emptyset, \oplus(\Theta, \Theta, \top)$, this means that we will have used $\mathcal{A}$ to derive a constant-depth formula that witnesses an obvious falsehood—moreover, it does it in a way that can be seen to contradict a parity axiom. This yields a small refutation of $\mathcal{A}$ in AC^0-Frege with parity axioms.

Now, how does one write an AC^0-Frege formula that witnesses a cedent $C := \Omega, \oplus\Xi$, or in other words, witnesses that either Ω is satisfied or an even number of the formulas in Ξ are satisfied? The first disjunct is easy to express: we simply need a disjunction of all formulas in Ω. To deal with the second disjunct, we will want to say that there is a perfect matching on the set of satisfied elements of Ξ. To this end, for each $e \in \binom{[k]}{2}$, where k is the length of Ξ, we introduce a formula μ^C_e that intuitively says "the two formulas ξ_i, ξ_j in Ξ with $e = \{i, j\}$ are matched to one another". This will be an AC^0-Frege formula in the variables of P and its shape will depend on how C is derived in P. The formula γ^C will then say:

> "If all formulas in Ω are false, then:
> for each e, if μ^C_e holds, then ξ_i holds for $i \in e$;
> for each $i \in [k]$, if ξ_i holds then some μ^C_e for $e \ni i$ holds;
> finally, for $e \perp f$ at least one of μ^C_e, μ^C_f fails to hold".

Note that γ^C will be a constant-depth formula if the μ^C_e's are as well.

The difficulty of course is that we have to make both the γ^C's and their AC^0-Frege derivations have small size. This lets us explain our reason for translating $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ derivations into $\mathrm{PK}^{\mathrm{one}\oplus}_{O(1)}$. Consider the $\mathrm{PK}^2_1(\oplus)$ cedent $\oplus^0(p_1, \ldots, p_n), \oplus^1(p_1, \ldots, p_n)$, which has a small proof. What we *cannot* hope to do is witness the validity of this cedent by:

- writing down a small AC^0-Frege formula γ_0 that logically implies the existence of a perfect matching on the set of the satisfied p_i's,
- writing down a small AC^0-Frege formula γ_1 that logically implies the existence of a perfect matching on the set of the satisfied p_i's together with an extra element $\top$,
- and then giving a proof (of any sort) of $\gamma_0 \vee \gamma_1$.

If we could do that, γ_1 would be a small constant-depth formula for parity. However, it is easy to write down and prove a small formula witnessing the validity of the $(\cdot)^{\mathrm{one}\oplus}$ translation of our cedent (cf. the example below Definition 11): the formula says that if p_i is satisfied, then $p_i \wedge p_i$ is matched to p_i, and if both p_i and p_j are satisfied, then $p_i \wedge p_j$ is matched to $p_j \wedge p_i$. In general, to witness provable disjunctions of parities we have to think of them as single parities, and the translation makes that explicit.

To explain our reasons for avoiding the Subtract rule, we have to say something about how the matchings given by the μ^{C}_{e}'s are defined as we progress through P. The idea is that when a formula first enters inside $\oplus$, which can only happen at a MOD inference, it can be matched to the final $\top$ inside the same $\oplus$. So, for instance, for an inference

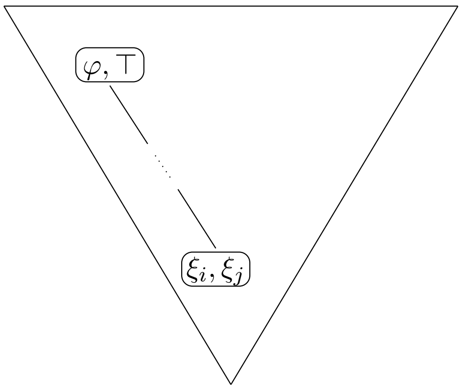
$$\frac{\Omega, \overline{\varphi}, \oplus\emptyset \qquad \Omega, \varphi, \oplus(\top)}{\Omega, \oplus(\varphi, \top)}$$

the formula γ^C corresponding to the conclusion of the inference will say that if φ holds, then φ is matched to $\top$; otherwise, there are no edges in the matching at all (in which case, γ^C claims that Ω is satisfied)—in other words, the formula $\mu^{C}_{\{1,2\}}$ corresponding to the unique possible edge in the conclusion is simply φ. Then, the matchings are propagated from premises to conclusions of inferences. For example, at a cut inference, say

$$\frac{\varphi, \oplus\Xi \qquad \overline{\varphi}, \oplus\Xi}{\emptyset, \oplus\Xi}\ \text{Cut}$$

we will want to use the matching from the right premise C_R if φ holds and the one from the left premise C_L if φ fails (recall that e.g. γ^{C_L} only claims to have a matching on Ξ if φ fails). Thus, the formula $\mu^{C}_{\{i,j\}}$ stating that two formulas $\xi_i, \xi_j \in \Xi$ are matched at the conclusion of the inference will have to be equivalent to $(\varphi \wedge \mu^{C_R}_{\{i,j\}}) \vee (\neg\varphi \wedge \mu^{C_L}_{\{i,j\}})$. At an Add inference, we will take the disjoint union of the matchings corresponding to the two premises, etc.

For all inferences except Subtract (and the special case of MOD outlined above), we are able to ensure that if two formulas ξ_i and ξ_j are to be matched at the conclusion of the inference, then there is a premise of the inference such that some canonical ancestors of ξ_i and ξ_j are *both* inside $\oplus$ in that premise and are already matched there (and some side conditions are satisfied telling us that this is the premise to look at—cf. the cut example). For this reason, if $C := \Omega, \oplus\Xi$ was derived without using subtraction, each potential scenario justifying that ξ_i and ξ_j are matched at C looks like this:



Here some ancestors of ξ_i, ξ_j having the form $\varphi, \top$ are first matched by an edge at a MOD inference and this edge propagates downwards along a path in P satisfying some conditions. There are at most as many such scenarios above as there are MOD inferences above C, and the formula $\mu^C_{\{i,j\}}$ is basically a disjunction over those scenarios.

The Subtract rule, however, presents a problem, as some arguments of $\oplus$ in the left premise disappear in the conclusion[2], and we have to find new matches for the formulas that were matched to those arguments. We could try something like:

$$\frac{\oplus^0(\xi_1, \xi_2, \boxed{\xi_3}, \boxed{\xi_4, \psi_1}, \boxed{\psi_2}, \psi_3, \psi_4) \qquad \oplus^0(\enclose{circle}{\psi_1, \psi_2}, \psi_3, \psi_4)}{\oplus^0(\xi_1, \xi_2, \enclose{circle}{\xi_3, \xi_4})}$$

In other words, match ξ_i, ξ_j at C if they were matched at C_L to formulas ψ_k, ψ_ℓ that were themselves matched at C_R. Still, this is problematic: there are many possibilities for what ψ_k, ψ_ℓ were, and for each possibility we are now faced with the task of justifying why some formulas were matched at C_L *and* why some other formulas were matched at C_R. If C_L, C_R were again derived using subtractions, and so on, the number of potential scenarios that could lead to ξ_i, ξ_j being matched at C becomes *doubly* exponential in the height of the derivation above C—in other words, exponential even if the derivation is balanced. Hence, the Subtract rule is best avoided, for instance in the way provided by Lemma 13.

We now give the technical details.

[2] Note that in PK($\oplus$) there are two rules that do not have the subformula property: Cut and Subtract. In the translation to the one-parity system, applications of these two rules to $\oplus$'s are both translated into Subtract.

Lemma 14. *Let $\mathcal{A}$ be a set of cedents consisting of formulas which do not contain parity connectives. Let P be a treelike $\mathrm{PK}_d^{\mathrm{one}\oplus}$-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 s. Then there is an AC^0-Frege with parity axioms derivation of Δ from $\mathcal{A}$ with size polynomial in s.*

Proof. For each cedent $C = \Omega, \oplus\Xi$ in P, let ξ_u^C be the u-th element of Ξ. As explained above, for each C we will give a polysize AC^0-Frege proof of the formula γ^C, which states that either Ω holds or the formulas μ_e^C, for $e \in \binom{[\mathrm{lh}(\Xi)]}{2}$, define a matching on the satisfied elements of Ξ.

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): if we write θ_u for the u-th element of Θ, the matching is defined by formulas $\nu_e = \theta_u$ for $e = \{u, u+\mathrm{lh}(\Theta)\}$, $u \in [\mathrm{lh}(\Theta)]$, and $\nu_e = \bot$ for all other e. 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 [17]—this is actually the only place where we use the parity axioms), which gives a proof of Δ.

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 poly(s).

First, we define the intuitively obvious concept of one formula inside $\oplus$ being a (tacitly: direct) predecessor of another such formula in the derivation P. The formula $\xi_{u'}^{C'}$ is a *predecessor* of ξ_u^C if C' is a premise and C the conclusion of some inference in P, the inference uses some rule J, and one of the following happens (referring to the rules as stated in Section 4):

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

Observe that a formula is identical to its predecessor with the exception of Multiply inferences, in which it is the conjunction of the predecessor (or the sequence of its conjuncts) with another formula. The shift by 1 in MOD inferences is related to the new formula entering $\oplus$ as ξ_1^C (the new formula has no predecessor).

A *matching witness branch* for a pair of formulas ξ_u^C, ξ_v^C (also referred to below as a matching witness branch for (C, e) where $e = \{u, v\}$) is a sequence

$$(\xi_{u_1}^{C_1}, \xi_{v_1}^{C_1}), (\xi_{u_2}^{C_2}, \xi_{v_2}^{C_2}), \ldots, (\xi_{u_m}^{C_m}, \xi_{v_m}^{C_m}), \tag{6}$$

where

- $(\xi_{u_m}^{C_m}, \xi_{v_m}^{C_m}) = (\xi_u^C, \xi_v^C)$,
- C_1 is obtained by a MOD inference, $u_1 = 1$, and $v_1 = \mathrm{lh}(\Xi_1)$,
- for $i < m$, $\xi_{u_i}^{C_i}$ and $\xi_{v_i}^{C_i}$ are the predecessors of $\xi_{u_{i+1}}^{C_{i+1}}$, $\xi_{v_{i+1}}^{C_{i+1}}$, respectively,
- for $i < m$, if C_{i+1} is obtained by Multiply applied to C_i and to some sequence of formulas Ψ, then $u_{i+1} \equiv v_{i+1} \pmod{\mathrm{lh}(\Psi)}$.

So, a matching witness branch for ξ_u^C, ξ_v^C (intuitively: a potential reason why ξ_u^C, ξ_v^C might be matched to each other) is a sequence of formula pairs that begins at a MOD with the formula entering $\oplus$ and the final $\top$ inside $\oplus$, then progresses along the proof from a given pair of formulas at a premise of an inference to a pair of its successors at the conclusion of the inference, and ends in ξ_u^C, ξ_v^C. If the $(i+1)$-th cedent along the branch is obtained using Multiply, the additional condition $u_{i+1} \equiv v_{i+1} \pmod{\mathrm{lh}(\Psi)}$ ensures that $\xi_{u_{i+1}}^{C_{i+1}}$ is equivalent to $\xi_{u_i}^{C_i} \wedge \psi$ and $\xi_{v_{i+1}}^{C_{i+1}}$ to $\xi_{v_i}^{C_i} \wedge \psi$ for the same $\psi \in \Psi$.

Let $\mathcal{B}$ be the set of all matching witness branches for ξ_u^C, ξ_v^C. We want the formula $\mu_{\{u,v\}}^C$ to be

$$\xi_u^C \wedge \xi_v^C \wedge \bigvee_{B \in \mathcal{B}} \beta^B,$$

where β^B is a formula describing some additional conditions on B, defined below. In other words, we want an edge to appear between a satisfied formula entering inside $\oplus$ at MOD and the last $\top$ inside that $\oplus$, and then we want this edge to propagate from premises to conclusions along B if some additional conditions on B are satisfied. To appreciate the last point, note that we should prefer propagating edges from premises in which the non-$\oplus$ part is false, since otherwise a perfect matching on the satisfied inputs to $\oplus$ might not even exist. Furthermore, even when there is no obvious reason to prefer some premise we must have a way of choosing just one of them in order to avoid propagating conflicting edges.

Let B be a matching witness branch as in (6). For $1 < i \leq m$, let J_i be the rule used to derive C_i in P. Define β^B to be:

$$\bigwedge_{1<i\leq m} \alpha_{J_i,C_{i-1}},$$

where $\alpha_{J_i,C_{i-1}}$ is a formula justifying the choice of premise C_{i-1} made by B. If J_i is one of the rules Cut, MOD, AND, the formula $\alpha_{J_i,C_{i-1}}$ is defined 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. If J_i is any rule other than Cut, MOD, AND, then $\alpha_{J_i,C_{i-1}}$ is $\top$. Roughly speaking, the conditions expressed by the $\alpha_{J_i,C_{i-1}}$ formulas make sure that the edges in a matching for C_i come from the correct premise: a premise for which it is most likely that the edges actually defined a matching (and from the first such premise in the case of AND).

This completes the description of the formulas μ_e^C and thus also of γ^C. Now it remains to show that for each line C of P the formula γ^C has a small AC^0-Frege proof from $\mathcal{A}$. If $C = \Omega, \oplus\Xi$ in P with $k = \mathrm{lh}(\Xi) \geq 2$, this comes down to proving the following cedents:

$$\Omega, \overline{\mu_e^C}, \xi_u^C \quad \text{for each } u \in e \in \binom{[k]}{2}, \tag{7}$$

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

$$\Omega, \overline{\mu_e^C}, \overline{\mu_f^C} \quad \text{for each } e, f \in \binom{[k]}{2} \text{ such that } e \perp f. \tag{9}$$

Using Lemma 4 (a) we derive $\overline{\xi_u^C}, \xi_u^C$ from which (7) follows easily. We construct polysize AC^0-Frege derivations of (8) and (9) from $\mathcal{A}$ by induction on C.

If C is an axiom, both (8) and (9) are obvious.

Suppose that (8), (9) hold for the premises of the rule used to derive line C. If C is obtained from $C' = \Omega', \oplus\Xi'$ by one of the rules Weakening, Exchange, OR, Contraction, then for $u \in [k]$ we have $\mu_e^C = \mu_e^{C'}$ for each $e \in \binom{[k]}{2}$ with $u \in e$, so (8) and (9) are easily obtained from the induction hypothesis.

Below, we focus on the case in which $C = \Gamma, \bigwedge_{i\in[n]} \varphi_i, \oplus\Xi$ is obtained by the AND inference

$$\frac{\Gamma, \varphi_1, \oplus\Xi \qquad \cdots \qquad \Gamma, \varphi_n, \oplus\Xi}{\Gamma, \bigwedge_{i\in[n]} \varphi_i, \oplus\Xi}$$

in the nontrivial situation where $n \geq 1$ and $k = \mathrm{lh}(\Xi) \geq 2$. The cases for Cut, Add, MOD, Multiply, Permute are easier and use similar ideas, so we leave them to the reader.

Let $u \in e \in \binom{[k]}{2}$. Notice that given any $\ell \in [n]$, to each matching witness branch B' of (C_ℓ, e) there is a unique matching witness branch B of (C, e) such that β^B is identical to $\beta^{B'} \wedge \alpha_{\mathrm{AND},C_\ell}$. Denote the set of matching witness branches of (C_ℓ, e) by $\mathcal{B}_\ell^e$.

To obtain (8), one may use, for each $\ell \in [n]$, the induction hypothesis

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

to derive

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

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

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

from which the desired

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

follows using OR rules and some easily derivable regrouping and distributivity properties.

To obtain (9), let $e = \{u, v\} \perp f = \{u, w\} \in \binom{[k]}{2}$ 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 $\bigwedge_{\ell \in [n]} \varphi_\ell$. □

Theorem 15. *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 s:

(i) apply Lemma 5 to obtain a treelike $\mathrm{PK}^{\log s + O(1)}_{d+1}(\oplus)$ refutation of $\mathcal{A}$ of size $s^{O(1)}$ and height $O(\log s)$,

(ii) apply Lemma 12 to obtain a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+2}$ refutation of $\{(\Xi)^{\mathrm{one}\oplus} : \Xi \in \mathcal{A}\}$ of size $s^{O(\log s)}$ and height $O(\log^2 s)$,

(iii) apply Lemma 13 to obtain a treelike $\mathrm{PK}^{\mathrm{one}\oplus}_{d+2}$ derivation of $\oplus(\Theta, \Theta, \top)$ from $\{(\Xi)^{\mathrm{one}\oplus} : \Xi \in \mathcal{A}\}$ which has size $s^{O(\log^3 s)}$, height $O(\log^2 s)$ and does not use the Subtract rule,

(iv) apply Lemma 14 to obtain a refutation of $\mathcal{A}$ in AC^0-Frege with parity axioms which has size polynomial in $s^{O(\log^3 s)}$.

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

Remark. By part (b) of Theorem 17, the simulation in Theorem 15 cannot be improved to polynomial.

6 Separations

In this section we study the question to what extent known methods can give separations between subsystems of $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ and AC^0[2]-Frege. It turns out that the answer depends on whether the separating formulas are allowed to contain parity connectives. It is easy to prove exponential separations between treelike and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ and AC^0[2]-Frege as refutation systems for sets of cedents containing $\oplus$. On the other hand, if we consider just refutations of CNF's, we verify that the Impagliazzo-Segerlind technique of [16] can be used to separate $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ from AC^0[2]-Frege—but this technique gives only superpolynomial separations, which, as Theorem 15 witnesses, leave quite a lot to be settled.

6.1 Exponential separations for formulas with $\bigoplus$

Theorem 16. *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}, $i < j \in [n]$:

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

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 $\mathbb{F}_2$:

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

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 $\mathrm{GOP}(G)$ [28].

On the other hand, any $\mathrm{PK}_d^c(\oplus)$ refutation of $\mathcal{A}_n$ requires size $2^{n^{\Omega(1/d)}}$. To see this, let P be a $\mathrm{PK}_d^c(\oplus)$ 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 $\rho_1 \ldots \rho_{d+1}$ 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. $\rho = \rho_1 \ldots \rho_{d+1}$ switches all $\oplus$-free subformulas of formulas appearing in P, as well as all the formulas $\bigvee \Gamma$ for $\Gamma, \oplus\Psi_1, \ldots, \oplus\Psi_c$ 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, $\rho_1 \ldots \rho_{d+1}$ 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 $\Gamma, \oplus\Psi_1, \ldots, \oplus\Psi_c$ in P gets mapped to a product of canonically defined polynomials $p_\Gamma p_{\Psi_1} \ldots p_{\Psi_c}$ obtained from the decision trees for $\bigvee \Gamma\restriction_{\rho\tau}$ and for the restrictions of elements of $\Psi_1, \ldots, \Psi_c$: p_Γ is 0 exactly if $\bigvee \Gamma\restriction_{\rho\tau}$ holds, and p_{Ψ_m} is 0 exactly if $\oplus(\Psi_m\restriction_{\rho\tau})$ holds for $m \in [c]$. Each polynomial $p_\Gamma p_{\Psi_1} \ldots p_{\Psi_c}$ 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 (up to renaming variables and replacing some $x_{ij\ell}$ by $1 + x_{ij\ell}$) 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 to replace variables with sums of variables in (negations of) tautologies that have polysize constant-degree refutations in Polynomial Calculus but require large degree to refute in the Nullstellensatz proof system of [3]—for instance, the housesitting principles of [13, 9]. Then, on the one hand, the obvious substitution into the original PC refutations gives polysize PC refutations (which can be easily translated into polysize $\mathrm{PK}_2^{O(1)}(\oplus)$ refutations) of $\mathcal{B}_n$. On the other hand, applying switching as above to a size-$2^{n^{a\cdot(1/d)}}$ treelike $\mathrm{PK}_d^c(\oplus)$ refutation of $\mathcal{B}_n$ would give rise to a size-$2^{O(n^{a\cdot(1/d)})}$, degree-$O(n^{a\cdot(1/d)})$ treelike PC refutation of the original principle. By the known translation of treelike PC into Nullstellensatz [10, Theorem 5.4], such a refutation does not exist if a is small enough. □

6.2 Separations without $\oplus$: the Impagliazzo-Segerlind argument

Theorem 17. *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* $\mathrm{poly}(n)$*-size refutation in* $\mathrm{AC}^0[2]$*-Frege, 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* $\mathrm{poly}(n)$*-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.*

Both parts of Theorem 17 are proved using a method of Impagliazzo and Segerlind presented in [16] and described in full detail in [26, Chapter VI]. We expect the reader to have those two texts at hand. Part (b) actually follows more or less directly from the work of [16, 26]. As for part (a), the idea is also essentially the same, but we need to verify that the method still applies. In the proof sketch below, we focus on describing the family of CNF's witnessing part (a). The upper bound in (a) is easy to prove. We explain some of the modifications and fixes to [26] needed to prove the lower bound in a separate appendix. Even in the appendix, the lower bound proof is only outlined, since the technical part of the argument is quite involved[3] but conceptually almost identical to the one described by Impagliazzo and Segerlind. The details are outsourced to [26, Chapter VI].

Proof sketch. Part (b) is witnessed by $\mathcal{B}_n = \mathrm{IS}(\mathcal{U})$, where $\mathcal{U}$ is what [16] would call an (n,n)-universe and the $\mathrm{IS}(\mathcal{U})$'s are the CNF's used in [16] 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 (with no explicit use of $\oplus$) that all of the cedents

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

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

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

are satisfied. Note that (12)-(14) are obtained by substituting parities of variables for individual variables in the obvious *propositional induction principle*: "p_0 holds, and if p_i holds then so does p_{i+1}, but p_n does not hold".

[3] Chapter VI of [26] is almost 60 pages.

If (12)-(14) are rewritten as a set of polynomial equations over $\mathbb{F}_2$, the set has polysize constant-degree Polynomial Calculus refutations, but by [12] requires logarithmic-degree Nullstellensatz refutations. This is used in [16, 26] to show that each $\mathcal{B}_n$ has size poly(n) and a polysize PC refutation (hence also a polysize $\mathrm{AC}^0[2]$-Frege refutation), but requires superpolynomial-size refutations in AC^0-Frege with parity axioms. However, it is also straightforward to give a polysize *treelike* $\mathrm{PK}_2^5(\oplus)$ 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.

To prove (a), we need to apply a modification similar to the one used to produce $\mathcal{B}_n$ from the propositional induction principle—however, this time, in analogy to Theorem 16 part (a), our starting point should be a family of tautologies that has small proofs in $\mathrm{AC}^0[2]$-Frege but no low-degree proofs in Polynomial Calculus. Once again, we want to replace individual variables by parities, but since the resulting formulas are not allowed to contain $\oplus$ connectives, every statement of the form "an even number of $\varphi_1, \ldots, \varphi_k$ are true" will have 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, we choose the weak pigeonhole principle PHP^{2m}_m. As a set of polynomials, this consists of:

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

The main reason for preferring weak PHP to the GOP formulas used in Theorem 16 is that the degree of weak PHP as a set of polynomials is only 2. Any higher degree would make the eventual formulas $\mathcal{A}_n$ quite hard to write down, not to mention work with. Moreover, weak PHP is defined in a very clear, explicit way, with no dependence on the structure of an expander graph (the restriction to a constant-degree expander in GOP is crucial to have any control over the degree of the polynomials). On the other hand, the fact that GOP has slightly smaller constant-depth refutations is no longer important 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 AC^0-Frege refutations of PHP^{2m}_m 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:

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

$$\oplus^0 \left(\{x_{i_1jk} \wedge x_{i_2j\ell} : k, \ell \in [n]\}\right), \qquad i_1 < i_2 \in [2m], j \in [m] \qquad (16)$$

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_1jk} and $x_{i_2j\ell}$ 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:

$$\begin{aligned}
&x_{ijk}, && i \in [2m], j \in [m], k \in [n],\\
&y_{ijkp}, && i \in [2m], j \in [m], k \in [n], p \in [mn+1],\\
&v_{ie}, && i \in [2m], e \in \binom{[mn+1]}{2},\\
&z_{i_1i_2jk\ell q}, && i_1 < i_2 \in [2m], j \in [m], k, \ell \in [n], q \in [n^2],\\
&w_{i_1i_2jf}, && i_1 < i_2 \in [2m], j \in [m], f \in \binom{[n^2]}{2}.
\end{aligned}$$

Intuitively, y_{ijkp} says that x_{ijk} is matched to the type-1 extra point p; v_{ie} says that the two type-1 extra points in e are matched for the given i; $z_{i_1i_2jk\ell q}$ says that the pair of x_{i_1jk} and $x_{i_2j\ell}$ is matched to the type-2 extra point q; and $w_{i_1i_2jf}$ says that the two type-2 extra points in f are matched for

(i_1, i_2, j). The clauses of $\mathcal{A}_n$ are:

$$\begin{array}{rr}
\overline{x_{ijk}} \vee \bigvee_{p\in[nm+1]} y_{ijkp} & i\in[2m], j\in[m], k\in[n], \\
\overline{y_{ijkp}} \vee x_{ijk} & i\in[2m], j\in[m], k\in[n], p\in[mn+1], \\
\bigvee_{j\in[m],k\in[n]} y_{ijkp} \vee \bigvee_{e\ni p} v_{ie} & i\in[2m], p\in[mn+1], \\
\overline{y_{ijkp}} \vee \overline{y_{ij'k'p}} & i\in[2m], (j,k)\neq(j',k'), p\in[mn+1], \\
\overline{y_{ijkp}} \vee \overline{y_{ijkp'}} & i\in[2m], j\in[m], k\in[n], p\neq p', \\
\overline{y_{ijkp}} \vee \overline{v_{ie}} & i\in[2m], j\in[m], k\in[n], p\in e, \\
\overline{v_{ie}} \vee \overline{v_{ie'}} & i\in[2m], e\perp e', \\
\overline{x_{i_1jk}} \vee \overline{x_{i_2j\ell}} \vee \bigvee_{q\in[n^2]} z_{i_1i_2jk\ell q} & i_1<i_2\in[2m], j\in[m], k,\ell\in[n], \\
\overline{z_{i_1i_2jk\ell q}} \vee x_{i_1jk} & i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\in[n^2], \\
\overline{z_{i_1i_2jk\ell q}} \vee x_{i_2j\ell} & i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\in[n^2], \\
\bigvee_{k,\ell\in[n]} z_{i_1i_2jk\ell q} \vee \bigvee_{f\ni q} w_{i_1i_2jf}, & i_1<i_2\in[2m], j\in[m], q\in[n^2], \\
\overline{z_{i_1i_2jk\ell q}} \vee \overline{z_{i_1i_2jk'\ell'q}} & i_1<i_2\in[2m], j\in[m], (k,\ell)\neq(k',\ell'), q\in[n^2], \\
\overline{z_{i_1i_2jk\ell q}} \vee \overline{z_{i_1i_2jk\ell q'}} & i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\neq q', \\
\overline{z_{i_1i_2jk\ell q}} \vee \overline{w_{i_1i_2jf}} & i_1<i_2\in[2m], j\in[m], k,\ell\in[n], q\in f, \\
\overline{w_{i_1i_2jf}} \vee \overline{w_{i_1i_2jf'}} & i_1<i_2\in[2m], j\in[m], f\perp f'.
\end{array}$$

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. For an outline of the lower bound on proof size in $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$, see the Appendix. □

Remark. The reader may have noticed that Theorem 17 does not contain an analogue of Theorem 16 part (b), that is, a superpolynomial separation between treelike and daglike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$ on refutations of CNF's. In all likelihood, applying the techniques of [16] to a suitable family of formulas that have polynomial size constant-degree Polynomial Calculus refutations but do not have logarithmic-degree Nullstellensatz refutations would give a family of CNF's with polysize refutations in daglike but not treelike $\mathrm{PK}^{O(1)}_{O(1)}(\oplus)$. However, we have not investigated this in enough detail to claim it as a result.

7 The Itsykson-Sokolov system

In [18], 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 [21] 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) $\mathrm{PK}^{\log}_{O(1)}(\oplus)$. This 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 $\mathrm{PK}^{\mathrm{id}}_{O(1)}(\oplus)$.

The following theorem is yet another instance of the well-known general phenomenon that the lines in a treelike derivation can be simplified at the cost of making the derivation daglike; our proof is modelled quite closely after those of [19, 22]. Here we want to trade treelikeness of a $\mathrm{PK}^{\mathrm{id}}_{d}(\oplus)$-refutation for a small number of parities per line.

Theorem 18. *Suppose that $\mathcal{A}$ is a set of $\mathrm{PK}^{\mathrm{id}}_{d}(\oplus)$ cedents each of which contains at most p formulas with an $\oplus$ connective. If $\mathcal{A}$ has a treelike $\mathrm{PK}^{\mathrm{id}}_{d}(\oplus)$-refutation of size s and cedent-number t, then it also has a $\mathrm{PK}^{\log t+p+3}_{d}(\oplus)$-refutation of size $\mathrm{poly}(s)$.*

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 $\mathrm{PK}^{\mathrm{id}}_{d}(\oplus)$-derivation P we define $\mathcal{A}^P$ to be the set of all cedents having one of the following forms:

- Θ, Ω, where $\Omega \in \mathcal{A}$ and each element of Θ occurs (possibly negated) as an element of a line in P,
- $\Theta, \varphi, \overline{\varphi}$, where φ and each element of Θ occurs (possibly negated) as an element of a line in P.

By induction on t we prove the following: if $\mathcal{A}$ is a set of cedents of $\oplus$-width p and $\Xi = \xi_1, \dots, \xi_k$ has a treelike $\mathrm{PK}^{\mathrm{id}}_{d}(\oplus)$-derivation P from $\mathcal{A}$ of size s and cedent-number t, then there is a $\mathrm{PK}_d^{\log t+p+3}(\oplus)$-refutation P' of $\mathcal{A}^P \cup \{(\overline{\xi}_i) : i \in [k]\}$ of cedent-number $5st$, such that the size of each cedent in P' is at most $2s$.

In the base case for $t = 1$ and $\Xi \in \mathcal{A}$, we obtain a refutation of $\oplus$-width at most p by cutting Ξ against each $(\overline{\xi}_i)$. The other base 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.

We will describe only how to treat the case where Ξ is derived by a MOD inference. The remaining cases (AND, OR, Cut, Add, Subtract) use similar ideas and are no harder to deal with. We write Γ to stand for the side formulas of Ξ and assume that $\Gamma = \gamma_1, \dots, \gamma_k$.

If $\Xi = \Gamma, \oplus^b(\Phi, \varphi)$ is derived by the MOD rule, let Q_1 and Q_2 denote, respectively, the subderivations of P with endcedent $(\Gamma, \overline{\varphi}, \oplus^{b-1}\Phi)$ and $(\Gamma, \varphi, \oplus^b\Phi)$. Suppose that $\mathbf{cn}(Q_1) \le \mathbf{cn}(Q_2)$ (the opposite case is treated analogously). Because P is treelike, $Q - 1$ and Q_2 are disjoint and so $\mathbf{cn}(Q_1) < t/2$ and $\mathbf{cn}(Q_2) < t - 1$. By the induction hypothesis, there is a $\mathrm{PK}_d^{\log t+p+2}(\oplus)$-refutation Q_1' of $\mathcal{A}^{Q_1} \cup \{(\overline{\gamma}_i) : i \in [k]\} \cup \{(\varphi), (\oplus^b\Phi)\}$, and there is a $\mathrm{PK}_d^{\log(t-1)+p+3}(\oplus)$-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'. Note that the $\oplus$-width of Q_1''' is one greater than the $\oplus$-width of Q_1', hence it is at most $\log t + p + 3$. Also, because the size of each cedent in Q_1' is bounded by $2\mathbf{s}(Q_1)$ (by the induction hypothesis) and it is increased when constructing Q_1'' by less than $2(\mathbf{s}(P) - \mathbf{s}(Q_1))$, the size of each cedent in Q_1'' is bounded by $2\mathbf{s}(P) = 2s$ (and similarly for Q_1'''). Attach Q_2' to Q_1'' and Q_1''', forming a $\mathrm{PK}_d^{\log t+p+3}(\oplus)$-refutation $\widetilde{P}$ of

$$\mathcal{A}^P \cup \{(\overline{\varphi}, \overline{\gamma}_i) : i \in [k]\} \cup \{(\oplus^{b-1}\Phi, \overline{\gamma}_i) : i \in [k]\} \cup \{(\overline{\varphi}, \oplus^b\Phi), (\oplus^{b-1}\Phi, \varphi)\}.$$

Since the axioms of the desired refutation P' can only be from the set $\mathcal{A}^P \cup \{(\overline{\gamma}_i) : i \in [k]\} \cup \{(\oplus^{b-1}(\Phi, \varphi))\}$, we need to add some derivations to $\widetilde{P}$.

Add at most $3k$ cedents to $\widetilde{P}$ so that each axiom $(\overline{\varphi}, \overline{\gamma}_i), i \in [k]$, and $(\oplus^{b-1}\Phi, \overline{\gamma}_i), i \in [k]$, of $\widetilde{P}$ is derived from $(\overline{\gamma}_i)$ by a weakening and exchange. Attach 9 cedents to the axiom $(\overline{\varphi}, \oplus^b\Phi)$ of $\widetilde{P}$ to form its derivation from the cedent $(\oplus^{b-1}(\Phi, \varphi))$, the cedent $(\oplus^b\Phi, \oplus^{b-1}\Phi)$ and the cedent $(\overline{\varphi}, \varphi)$. Similarly, attach 6 cedents to the axiom $(\oplus^{b-1}\Phi, \varphi)$ of $\widetilde{P}$ to form its derivation from the same set of cedents. Call the resulting refutation P'. Because of the assumption $\mathbf{cn}(Q_1) \le \mathbf{cn}(Q_2)$ and because P is treelike, we have

$\mathbf{cn}(P') \leq \mathbf{cn}(Q_2') + 2\mathbf{cn}(Q_1') + 3k + 15 \leq 5\mathbf{s}(Q_2)\mathbf{cn}(Q_2) + 10\mathbf{s}(Q_1)\mathbf{cn}(Q_1) + 3k + 15 \leq 5(\mathbf{s}(Q_2) + \mathbf{s}(Q_1) + \mathbf{s}(\Xi))(1 + \mathbf{cn}(Q_2) + \mathbf{cn}(Q_1)) = 5st$. The $\oplus$-width of the cedents added to $\widetilde{P}$ to form P' is not greater than the $\oplus$-width of $\widetilde{P}$. A bound on the sizes of these additional cedents can be obtained similarly to the bound on sizes of cedents in Q_1''. Hence P' is a $\mathrm{PK}_d^{\log t+p+3}(\oplus)$-refutation with the required properties. This concludes the inductive step.

Taking for P the refutation from the statement of the theorem, we thus obtain a $\mathrm{PK}_d^{\log t+p+3}(\oplus)$-refutation P' of $\mathcal{A}^P$ of cedent-number $5st$, such that each line has size at most $2s$. 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 4. The resulting $\mathrm{PK}_d^{\log t+p+3}(\oplus)$-refutation has size $\leq 5st \cdot 2s \cdot 3 + 5st \cdot O(s^4) = O(s^6)$. □

Theorem 18 implies:

Corollary 19. *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 [20] with the degree lower bounds of [8].

Namely, given a $\mathrm{PK}_d^{\log}(\oplus)$ refutation of Count_3^n smaller than 2^{n^δ} for sufficiently small δ, apply the switching lemma associated with Count_3 (cf. e.g. [2, Section 6.3]) to turn the refutation into a refutation of $\mathrm{Count}_3^{n^\gamma}$ (for some $\gamma > 0$) in Polynomial Calculus over $\mathbb{F}_2$ with degree $o(n^\gamma)$. (This part of the argument is analogous to [20], but a similar argument involving a simpler switching lemma is described in the proof of our Theorem 16 part (a).) However, by [8, Corollary 20], any refutation of $\mathrm{Count}_3^{n^\gamma}$ in PC over $\mathbb{F}_2$ must have degree $\Omega(n^\gamma)$. □

Remark. It would be possible to prove Corollary 19 via a quasipolynomial simulation of treelike $\mathrm{PK}_d^{\mathrm{id}}(\oplus)$ by daglike $\mathrm{PK}_{O(1)}^3(\oplus)$—in other words, by decreasing the number of $\oplus$'s per line to a constant at the cost of making the derivation somewhat bigger. To obtain the simulation, one first applies the simulation of Theorem 18 and then the translation $(\cdot)^{\mathrm{one}\oplus}$ of Section 4. The result is a sequence of $\mathrm{PK}_{O(1)}^{\mathrm{one}\oplus}(\oplus)$, which can also be viewed as $\mathrm{PK}_{O(1)}^1(\oplus)$ cedents in the obvious way. The sequence 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 19 can be easily obtained for daglike $\mathrm{PK}^{\mathrm{id}}_{O(1)}(\oplus)$. In fact, already bounds for a subsystem of $\mathrm{PK}^{\mathrm{id}}_{2}(\oplus)$ 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 $T^{2,\oplus\mathrm{P}}_{2}(\alpha)$ [11], 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 the two anonymous reviewers of the paper, whose detailed comments made it possible to improve the presentation considerably.

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. Technical Report UW-CSE-95-07-01, University of Washington, Department of Computer Science and Engineering, 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, Russell Impagliazzo, Jan Krajícek, Toniann Pitassi, Pavel Pudlák, and Alan R. Woods. Exponential lower bounds for the pigeonhole principle. In *Proc. 24th ACM Symposium on Theory of Computing*, pages 200–220. ACM, 1992.

[5] 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.

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

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] Samuel R. Buss and Toniann Pitassi. Good degree bounds on nullstellensatz refutations of the induction principle. *Journal of Computer and System Sciences*, 57(2):162–171, 1998.

[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 Jan Krajíček. A note on conservativity relations among bounded arithmetic theories. *Math. Log. Q.*, 48:375–377, 2002.

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

[17] 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.

[18] 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.

[19] Jan Krajíček. Lower bounds to the size of constant-depth propositional proofs. *The Journal of Symbolic Logic*, 59(1):73–86, 1994.

[20] Jan Krajíček. Lower bounds for a proof system with an exponential 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.

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

[22] Massimo Lauria. A note about k-DNF resolution. Manuscript submitted for publication, 2018.

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

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

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

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

[27] P. M. Spira. On time hardware complexity tradeoffs for boolean functions. In *Proceedings of the Fourth Hawaii International Symposium on System Sciences*, pages 525–527, 1971.

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

A The Impagliazzo-Segerlind lower bound argument: some details

We outline a proof of the lower bound in Theorem 17 part (a). For definitions of the family of formulas $\mathcal{A}_n$ and related notation, see Section 6.1. The proof is a switching lemma argument closely modelled after the one in [26, Chapter VI].

As usual in this sort of argument, we want to show that after a restriction of one kind or another, each constant-depth De Morgan formulas in a hypothetical polynomial-size $\mathrm{PK}_d^c(\oplus)$ proof can be translated into a low-depth decision tree. As in [26] (and in classical lower-bound proofs for e.g. PHP_n^{n+1} [4]), the decision trees we use are allowed to make queries other than just the truth value of variables and receive answers other than yes/no. More specifically, a decision tree can make the following queries:

- *Is x_{ijk} true?* This is answered with either a "yes" or a "no".
- *Which type-1 extra point is x_{ijk} is matched to?* This is answered either with "x_{ijk} is false" (and hence, it is not matched to anything) or "x_{ijk} is true and it is matched to p" (which means y_{ijkp} is also true).
- (For some given i) *What is the type-1 extra point p is matched to?* This is answered either by "p is matched to the extra point p'" (which means that v_{ie} is true, for $e = \{p, p'\}$) or by "p is matched to the variable x_{ijk}" (which means that both x_{ijk} and y_{ijkp} are true).
- *Which type-2 extra point is the pair $(x_{i_1jk}, x_{i_2j\ell})$ matched to?*. This is answered in one of four ways: "$x_{i_1jk}, x_{i_2j\ell}$ are both false", "x_{i_1jk} is true, but $x_{i_2j\ell}$ is false", "$x_{i_2j\ell}$ is true, but x_{i_1jk} is false", or "$x_{i_1jk}, x_{i_2j\ell}$ are both true and the pair is matched to q" (which means that $z_{i_1i_2jklq}$ is also true).
- (For some given i_1, i_2, j) *What is the type-2 extra point q matched to?* This is answered either by "q is matched to the extra point q'" (which means that $w_{i_1i_2jf}$ is true for $f = \{q, q'\}$) or by "q is matched to the pair $(x_{i_1jk}, x_{i_2j\ell})$" (which means that all of $x_{i_1jk}, x_{i_2j\ell}, z_{i_1i_2jklq}$ are true).

Each leaf of a decision tree corresponding to a formula in the hypothetical refutation is labelled by 0 or 1. A decision tree *strongly represents* a DNF φ if each 1-branch of the tree contains a term of φ and each 0-branch 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.).

As in [26], 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 $z_{i_1i_2jk\ell q}$, where q is the extra point assigned to (k, ℓ) in step (b); substitute $\neg x_{i_1jk}$

for the $w_{i_1 i_2 j f}$ such that f consists of the extra points assigned to (k,ℓ) and (k,ℓ') where $x_{i_2 j \ell'}$ has been set to 1 and $x_{i_2 j \ell}, x_{i_2 j \ell'}$ are matched in step (i)(b); set all $w_{i_1 i_2 j f'}$ 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).

The result of applying an (n, n^ϵ)-simplification ρ to $\mathcal{A}_n$ is a formula $\mathcal{A}_n\!\restriction_\rho$ which looks just like $\mathcal{A}_n$ except that n^ϵ now plays the role of n (also, some clauses of $\mathcal{A}_n$ become $\ell, \bar{\ell}$ for a literal ℓ instead of becoming 1 directly).

The aim then is to prove the following switching lemma.

Switching Lemma (analogous to Theorem 83 of [26]). Let r, t and $\epsilon \in (0,1]$ be constants. Let σ be an (n, n^ϵ)-simplification and let φ be an r-DNF in the variables of $\mathcal{A}_n\!\restriction_\sigma$. There exist constants $\delta \in (0,\epsilon)$ and h such that the probability, over randomly chosen (n, n^δ)-simplifications ρ extending σ, that $\varphi\!\restriction_\rho$ has no decision tree of height at most h, is at most n^{-t}.

To prove the switching lemma, one introduces a notion of independent terms in a DNF in the variables of $\mathcal{A}_n\!\restriction_\sigma$ w.r.t. ρ (very roughly, terms which after restriction by ρ can still be satisfied/falsified independently of each other; the actual definition is more subtle, see [26, Definition VI.G.10]). One then shows that for a sufficiently small δ and sufficiently large constant s, the probability that a given DNF has a set of independent terms of size at least s is below n^{-t}. This involves three main steps (the constants in the O and Ω notation below can be chosen independently of ϵ, δ, s):

- (analogous to [26, Lemma 89]) The probability, over random ρ and random $(n, n^\delta - 2rs)$-simplifications κ extending ρ, that a given procedure taking κ as input and outputting s literals of the form $x, \bar{x}, v$ or w actually outputs only literals satisfied by κ but not set in ρ, is $\leq (2rs/n^\epsilon)^{\Omega(s)}$.

- (analogous to [26, Lemma 90]) Given ρ and an independent set of size s w.r.t. ρ, the probability over κ as above that κ makes all terms in the independent set satisfied is $\geq (n^{-\delta r^2})^{O(s^2)}$ (a closer analogue to [26] would have $(n^{-\delta r})^{O(s)}$, but see Remark below).

- (analogous to [26, Lemma 91]) There is a randomized procedure which, for κ as above satisfying an independent set of size s w.r.t. ρ, outputs s literals of the form $x, \bar{x}, v$ or w satisfied by κ but not set in ρ with probability $\geq (1/r)^{O(s)}$.

Putting these three steps together, one concludes that the probability of selecting ρ with an independent set of size s is

$$\leq (2rs/n^{\epsilon})^{\Omega(s)}(n^{\delta r^2})^{O(s^2)}r^{O(s)},$$

which can be made smaller than n^{-t} by first choosing the constant s sufficiently large w.r.t. ϵ, t and then δ sufficiently small w.r.t. ϵ, s, r. Finally one shows (in analogy to [26, Theorem 93]) that an r-DNF formula with no set of independent terms of size more than s is strongly represented by a decision tree of height $O(r^4 s^2)$.

Remark. At this point, we have to mention two apparent issues with [26], at least one of which is not addressed in [16]. Firstly, the definition of simplification in [26] (more precisely, the definition of presimplification, Definition VI.G.2) is missing a step analogous to our (i)(b).

To use our setting and notation, this is as if instead of (iii)(c) we had the following (iii)(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 $z_{i_1i_2jk\ell q}$, where q is the extra point assigned to (k, ℓ) in step (b); pick a random matching $\mathcal{F}$ on the set of extra points q assigned to (k, ℓ) for some ℓ in step (b); for each $f \in \mathcal{F}$, substitute $\neg x_{i_1jk}$ for $w_{i_1i_2jf}$ and set all $w_{i_1i_2jf'}$ variables for $f \perp f'$ to 0. (And define (iii)(d') instead of (iii)(d) similarly—cf. [26, Definition VI.G.2, part 7].)

The effect is that the first paragraph of the proof of [26, 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): in the language of [26], this would mean ensuring that the matching on 1's induced by the partition $\mathcal{E}^j_\rho$ is the same for all unset X_j (and similarly for the partitions $\mathcal{F}^i_\rho$ for all unset X_i). This leads to a factor of $\frac{1}{2}$ appearing in the statement of the corrected version of [26, Lemma 89], but that has no bearing on the eventual statement of [26, Theorem 83].

The other issue is possibly fixed to some extent in [16], but some of the details are not present in that extended abstract. We explain the problem in the language of [26]. The proof of [26, Lemma 91], specifically the argument justifying the claim "there is a literal of T_{t+1} not set by B^{ρ}", does not seem to work under the current definitions of independent set and of s-encoding [26, 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 [16, 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 [16]). 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 [26, Lemma 90]. Eventually, this results in much worse bounds in [26, Lemma 84]: the term L^{Cr} has to be replaced by L^{Cr^2s}. Once again, though, no changes to the statement of [26, Theorem 83] are needed.

Now assume that P is a polysize $\mathrm{PK}_d^c(\oplus)$ refutation of $\mathcal{A}_n$. Choosing a sufficiently large t dependent on the size of P, we apply the switching lemma $d+1$ times to the $\oplus$-free formulas appearing in lines of P. Let $\rho = \rho_1 \ldots \rho_{d+1}$ stand for the simplification built during the iteration of the lemma. Then ρ induces a mapping from all $\oplus$-free 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. Given a bound $k \in \mathbb{N}$ on the height of the decision trees, it is possible to verify the mapping is a k-evaluation, in a sense analogous to [26, Definition VI.H.1].

As in the proof of Theorem 16 part (a), each line of $P{\restriction}_{\rho}$ can be viewed as a degree-k polynomial. Moreover, using the properties of k-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 and from $\mathcal{A}_n{\restriction}_{\rho}$. Apply a further substitution of variables by constants, τ, so that $\mathcal{A}_n{\restriction}_{\rho\tau}$ becomes identical to PHP_m^{2m}. Now $P{\restriction}_{\rho\tau}$ is a constant-degree Polynomial Calculus refutation of PHP_m^{2m}, which does not exist by [25].