Progress on $\mathrm{AC}^0[2]$-Frege Lower Bounds*

Yaroslav Alekseev
Technion

Svyatoslav Gryaznov
Imperial College London

Toniann Pitassi
Columbia University

Robert Robere
McGill University

Alexander Shekhovtsov
Columbia University

Iddo Tzameret
Imperial College London

November 3, 2026

Abstract

We prove superpolynomial size lower bounds for depth-3.5 $\mathrm{AC}^0[2]$-Frege proofs, whose lines are disjunctions of conjunctions of parities of conjunctions of at most b literals. The top two fan-ins are unbounded, and our bound holds for every $b = o(\sqrt{\log n})$. A normal-form theorem of Buss, Kołodziejczyk and Zdanowski reduces quasipolynomial-size $\mathrm{AC}^0[2]$-Frege proofs to this form with polylogarithmic b. Our result therefore proves quasipolynomial lower bounds for a restricted version of the normal form underlying the longstanding $\mathrm{AC}^0[2]$-Frege lower-bound problem.

Our main contribution is a stronger analysis of Pang's approximation method after replacing each variable of the input formula by a small majority function. The key is to control approximation errors using the algebraic hardness of the original formula, without requiring stronger starting degree lower bounds. This extends Pang's exponential lower bounds for resolution over polynomials, $\mathrm{Res}(\mathrm{PC}_r)$, from degree $r = O(\log n)$ to $r = (\log n)^{3/2}$. A second level of Razborov–Smolensky approximation, carried through the syntactic proof rules, yields the depth-3.5 result.

AI Methodology. The authors obtained the improved degree lower bounds by asking ChatGPT 6 to try to improve the degree bounds, with targeted prompts containing the main ideas of lifting with majorities (inspired by the recent correlation bounds for $\mathbb{F}_2$) plus the idea of decoupling the two uses of degree. They considerably simplified and rewrote the proofs it produced. AI tools (ChatGPT 5.6 and Claude (Fable 5.1)) were used to fine-tune the lower bounds. They were also used to proofread and assist with exposition.

*S.G., T.P., R.R., and I.T. would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Logical Foundations of Complexity Theory, where work on this paper was undertaken. This work was supported by EPSRC grant EP/Z000580/1. Yaroslav Alekseev was supported by Israel Science Foundation grant 507/24. Robert Robere was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). The work of Svyatoslav Gryaznov and Iddo Tzameret received funding from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 101002742, EPRICOT project). Iddo Tzameret was also supported by the Engineering and Physical Sciences Research Council under grant EP/Z534158/1.

1 Introduction

Proving superpolynomial lower bounds for $\mathrm{AC}^0[2]$-Frege has been a longstanding open problem in proof complexity. We have known exponential lower bounds for AC^0-Frege since the early 1990s. These bounds grew out of the structural insight underlying the lower bounds for AC^0 circuits: small constant-depth circuits simplify dramatically under random restrictions, and in particular their functions are constant on large subcubes [FSS84, Hås86]. Adapting this insight to proofs led to Ajtai's superpolynomial lower bound for bounded-depth Frege in 1988 [Ajt88], followed by the exponential lower bounds for the pigeonhole principle of Pitassi, Beame and Impagliazzo, and independently Krajíček, Pudlák and Woods [PBI93, KPW95, BIK+92].

For $\mathrm{AC}^0[p]$ circuits with mod p counting gates, celebrated exponential lower bounds were obtained by Razborov and Smolensky in 1987 [Raz87, Smo87]. Their method uses the polynomial method, where it is shown that any polynomial-size $\mathrm{AC}^0[p]$ Boolean circuit can be approximated by a low-degree polynomial over a finite field.

Although polynomial approximations were powerful enough to prove $\mathrm{AC}^0[p]$ circuit lower bounds, several problems remained challenging. First, while Razborov and Smolensky's lower bound is exponential, it was a longstanding open problem to obtain a strong *correlation* bound, showing that any small $\mathrm{AC}^0[p]$ circuit cannot compute Majority (or some other explicit function) correctly on more than $1/2 + 1/2^{n^\epsilon}$ of the inputs. This problem remained open for nearly 40 years and was recently solved in a major breakthrough [CHL+26].

Secondly, achieving similar exponential lower bounds, or even superpolynomial lower bounds, for $\mathrm{AC}^0[p]$-Frege has been a major challenge in proof complexity, and is perhaps the most important challenge in the field. There is, however, a structural fact that has no analogue for AC^0-Frege, and that makes the target much more concrete. By formalizing the Razborov–Smolensky argument inside low-depth $\mathrm{AC}^0[2]$-Frege, Buss, Kołodziejczyk and Zdanowski [BKZ15] proved a normal-form theorem: every quasipolynomial-size $\mathrm{AC}^0[2]$-Frege proof can be converted into a quasipolynomial-size proof whose lines have depth 3.5^+, that is, the form $\bigvee \circ \bigwedge \circ \oplus \circ \bigwedge_r$ with unbounded fan-in at the two upper levels and bottom fan-in $r = \mathrm{polylog}(n)$. The "plus" records that the bottom degree is polylogarithmic rather than constant, and the exponent of the polylogarithm depends on the depth d of the original proof [BKZ15, Theorems 42 and 43]: a depth-d proof collapses to bottom degree $(\log n)^{O(d)}$. So to prove superpolynomial lower bounds for depth-d $\mathrm{AC}^0[2]$-Frege it suffices to prove them for proofs of depth 3.5^+ with bottom degree $(\log n)^{O(d)}$.

The task is therefore "merely" to prove lower bounds for a single fixed depth, 3.5^+. Even that has been out of reach. Indeed, obtaining lower bounds for constant-depth Frege with mod-2 gates was an original motivation for introducing algebraic proof systems such as Nullstellensatz and polynomial calculus [BIK+96, CEI96, Pit97], which isolate some of the most basic forms of algebraic reasoning available in $\mathrm{AC}^0[2]$-Frege. In bounded-degree polynomial calculus, each line is a single low-degree polynomial equation over $\mathbb{F}_2$, or equivalently an equation involving a parity of small monomials. Another basic fragment is resolution over parities, $\mathrm{Res}(\oplus)$, whose lines are disjunctions of affine equations over $\mathbb{F}_2$. Both are restrictions of depth-3.5^+ $\mathrm{AC}^0[2]$-Frege to lines of depth two with bottom degree one. Yet, despite considerable effort, until very recently even $\mathrm{Res}(\oplus)$ resisted superpolynomial lower bounds for unrestricted DAG-like refutations.

Braun's recent result changed this picture [Bra26]. He proved an exponential lower bound for $\mathrm{Res}(\oplus)$ for the bit pigeonhole principle. The proof introduced a new algebraic technique with a surprisingly simple core: use one common polynomial multiplier to control the high-rank lines of an assumed short proof. These are the lines whose affine equations contain many independent linear

parts. Multiplication annihilates the polynomials representing their failure, allowing the proof to be converted into a low-degree polynomial-calculus derivation that can be ruled out algebraically.

While this is a significant breakthrough, it is not clear whether the ideas behind it can be useful for proving superpolynomial $\mathrm{AC}^0[p]$-Frege lower bounds. Lower bounds for PC_r and $\mathrm{Res}(\oplus)$ are in any case a prerequisite, since both are fragments of $\mathrm{AC}^0[p]$-Frege; and by the normal-form theorem above the distance from these fragments to the whole system is a single depth.

The semantic barrier. One of the key challenges in extending polynomial-calculus or resolution-over-parities lower bounds to all of $\mathrm{AC}^0[2]$-Frege is the *semantic barrier*. Here, a semantic inference may derive any allowed line from at most two earlier lines whenever the implication is sound. If the allowed lines can express conjunctions of input clauses, every unsatisfiable CNF has a polynomial-size semantic refutation: form the conjunction of its clauses and then infer a contradiction. There is a dual argument when the lines can express DNFs. The negation of an unsatisfiable CNF is a DNF tautology, which can be introduced by a semantic inference from the constant true line. The input clauses then eliminate its terms one at a time, ending with the empty disjunction. Thus, a lower bound for $\mathrm{AC}^0[2]$-Frege must separate syntactic proofs from short semantic proofs. It must use the inference rules in an essential way. Buss, Kołodziejczyk and Zdanowski put the same obstacle in the language of the circuit lower bounds: the Razborov–Smolensky approximation of a line by a low-degree polynomial "does not commute with inference rules" [BKZ15]. The common multiplier of Braun and Pang, described next, is exactly a way around this: it does not make the approximations commute with the rules, but it corrects, uniformly for all lines, the errors that the rules introduce.

Pang's extension. Pang gave a remarkable generalization of Braun's argument [Pan26]. His result applies to resolution over degree-r polynomial equations, denoted $\mathrm{Res}(\mathrm{PC}_r)$. A line has the form

$$(u_1 = 1) \vee \cdots \vee (u_s = 1), \qquad \deg u_i \le r,$$

with no bound on the number s of equations in the disjunction. Resolution over parities is the case $r = 1$.

Theorem 1.1 (Pang, [Pan26, Theorem 1.1])**.** *Let F be an unsatisfiable CNF on n variables, of maximum clause width $w(F)$. Suppose its polynomial encoding has no polynomial-calculus refutation over $\mathbb{F}_2$ of degree at most D, where $D \ge \max\{32r, 8w(F)\}$. Then every S-line syntactic* $\mathrm{Res}(\mathrm{PC}_r)$ *refutation of F satisfies*

$$\log S \ge \frac{D}{128r}\left(\frac{D}{16n}\right)^r.$$

In particular, a constant-width CNF requiring linear PC degree requires $\exp(\Omega(n))$ lines when r is constant, and $\exp(n^{\Omega(1)})$ lines for $r \le c_0 \log n$, where $c_0 > 0$ is sufficiently small. The theorem gives a general way to turn PC degree lower bounds into size lower bounds for a substantially more expressive system.

We observe that this system already encounters the semantic barrier. Let unsatisfiable $F = C_1 \wedge \cdots \wedge C_m$ have width at most w. The negation of each clause is a conjunction of at most w literals, and can be expressed as a degree-w polynomial equation: the product of those literals equals 1. Consequently, for $r \ge w$, the tautology $\neg F$ is a permitted $\mathrm{Res}(\mathrm{PC}_r)$ line. After one-shot deriving $\neg F$, one can then derive a contradiction by sequentially resolving $\neg F$ with each of $C_1, \ldots, C_m$. The argument above gives a semantic refutation with $O(m)$ lines and, for constant w, polynomial total size. Pang's theorem nevertheless gives exponential lower bounds for the syntactic system, including

for fixed $r \geq w$. His proof therefore already distinguishes the local syntactic rules from arbitrary sound semantic inference.

1.1 Our results

In this paper, we show how to extend Pang's method to larger polynomial degree. Our main improvement gives exponential lower bounds for Res(PC$_r$) with $r = (\log n)^{3/2}$, for polynomial-size CNFs obtained by a majority substitution. Moreover, by a different choice of parameters, we obtain superpolynomial lower bounds for all degrees $r \leq n^{1/3-\varepsilon}$, for any fixed $\varepsilon > 0$. The proof is inspired by the recent exponential correlation bounds for low-degree polynomials of Chattopadhyay, Hatami, Lee, Lovett, Tal and Viola [CHL+26]. The main application of the improved degree range is a superpolynomial lower bound for depth-3.5 AC^0[2]-Frege, a system one parameter away from the normal form of quasipolynomial-size AC^0[2]-Frege. We also obtain lower bounds for unrestricted-degree Res(PC) measured by total proof size.

Throughout, S denotes the number of proof lines, whereas T denotes total proof size, with polynomials written explicitly as sums of monomials. We write $|G|$ for the encoding length of the input CNF. This distinction is particularly important when we allow the polynomial degree, or the size of the substitution gadget, to grow.

Definition 1.2 (Majority-lifted CNFs)**.**
Let $F(y_1, \ldots, y_n)$ be a CNF of constant width $w(F)$ over n variables that requires PC degree at least $D = cn$ for a fixed $c > 0$. For odd ℓ, let $G := F \circ \mathrm{Maj}_\ell$ be the CNF over the $N = n\ell$ new variables $x_{i,a}$ ($i \in [n]$, $a \in [\ell]$) obtained by replacing each variable y_i of F by the majority $\mathrm{Maj}_\ell(x_{i,1}, \ldots, x_{i,\ell})$ of its own block of ℓ new variables and expanding each substituted clause into a CNF. Then G has $N = n\ell$ variables, width $O(\ell)$ and size $\mathrm{poly}(n)\, 2^{O(\ell)}$. Throughout, n is the number of original variables (equivalently, of blocks), ℓ the block length, and $N = n\ell$ the number of lifted variables.

Theorem 1.3 (Improved degree bounds via majority lifting)**.** *Let F be a constant-width CNF on n variables with no PC refutation of degree $D = cn$, and let $G = F \circ \mathrm{Maj}_\ell$ be its majority lift as in Definition 1.2. There are constants $\ell_0, \kappa, C, c_1 > 0$, depending only on c and $w(F)$, such that for all sufficiently large n, all odd $\ell \geq \ell_0$ and all $r \geq 1$ with $r + \ell \leq \kappa n$, every S-line syntactic* Res(PC$_r$) *refutation of G satisfies*

$$\log S \geq c_1 \frac{N}{\ell r} \exp\left(-C \frac{r}{\sqrt{\ell}}\right) \;=\; c_1 \frac{n}{r} \exp\left(-C \frac{r}{\sqrt{\ell}}\right). \tag{1}$$

Below are two main consequences of the theorem for different choices of the block length ℓ.

1. $\ell = \Theta(\log n)$: $|G| = \mathrm{poly}(N)$ and $S \geq \exp(N^{\Omega(1)})$ for all $1 \leq r \leq (\log N)^{3/2}$.
2. $\ell = \Theta(r^2)$ with $r = \lfloor n^{1/3-\delta} \rfloor$ for a fixed $0 < \delta < 1/3$: $\log |G| = O(n^{2/3-2\delta})$ and $S \geq \exp(\Omega(n^{2/3+\delta})) = |G|^{\Omega(n^{3\delta})}$ (Corollary 4.23).

1.1.1 Corollaries of main theorem.

Superpolynomial size lower bounds for depth-3.5 AC^0[2]-Frege. Our main result concerns a proof system that is, up to one parameter, the normal form in which all of quasipolynomial-size AC^0[2]-Frege lives. Recall from above that by [BKZ15] every quasipolynomial-size AC^0[2]-Frege proof can be converted into a depth-four proof whose lines have the form $\bigvee \circ \bigwedge \circ \oplus \circ \bigwedge_r$ with

$r = \text{polylog}(n)$. Define depth-3.5 $\text{AC}^0[2]$-Frege to be the system whose lines have exactly this form with the bottom fan-in r replaced by a parameter b: a line is an OR of *terms*, where a term is an AND of polynomial equations $u = 1$ with $\deg u \le b$; that is, a line has the form $\bigvee \circ \bigwedge \circ \oplus \circ \bigwedge_b$ with unbounded fan-in at the OR and the AND. The bounded bottom level counts as half a level, hence the name. The system is formalized in Appendix D, where we also prove (Lemma D.2) that it is polynomially equivalent to the depth-3.5^+ system used to prove the normal form theorem [BKZ15]. Consequently a quasipolynomial-size depth-d $\text{AC}^0[2]$-Frege refutation of a CNF yields a quasipolynomial-size depth-3.5 refutation in our system with $b = (\log n)^{O(d)}$. Our main result, in Section 5, proves superpolynomial lower bounds for this system whenever $b = o(\sqrt{\log n})$.

Theorem 1.4 (Superpolynomial size bounds for depth-3.5 AC^0 [2]-Frege)**.** *Let F be a constant-width CNF on n variables with no PC refutation of degree cn, and let $G = F \circ \text{Maj}_\ell$ with $\ell = \Theta(\log n)$ chosen as in Corollary 4.21, so that G has polynomial size. There is a constant $\alpha > 0$, depending only on c, such that for all sufficiently large n and all $1 \le b \le \sqrt{\ell}$, every depth-3.5 $\text{AC}^0[2]$-Frege refutation of G with bottom degree b has total size*

$$\log T \;\ge\; \alpha \frac{(\log n)^{3/2}}{b}.$$

Since $|G|$ is polynomial in n, the bound is superpolynomial in $|G|$ for all $b = o(\sqrt{\log n})$, and equals $\exp(\Omega((\log n)^{3/2}))$ for constant b. Three remarks on what this theorem does and does not say.

First, by the simulation just described, the gap between this theorem and superpolynomial lower bounds for depth-d $\text{AC}^0[2]$-Frege is quantitative and lies in a single parameter: the bottom degree is $o(\sqrt{\log n})$ where it would have to be $(\log n)^{O(d)}$.

Second, the semantic barrier described above is at its most severe here. Already with $b = 1$ a single term can be the conjunction of the negated literals of a clause, so the negation of the input CNF, a DNF, is literally a line of the system, and the short semantic refutation described above (introduce $\neg F$ and eliminate its terms one at a time with the input clauses) uses only depth-3.5 lines. A lower bound for this system is therefore necessarily a statement about the rules, not about the lines, and the proof has to follow the rules one at a time; this is what Section 5 and Appendix D do.

Third, this is the one place where the majority lift is a prerequisite rather than an improvement. The argument approximates each term by a polynomial of degree $R = O(b \log T)$ and then approximates each line by a polynomial in these; Pang's method can handle the second step only when $R = O(\log n)$, which yields the trivial bound $T \ge n^{\Omega(1/b)}$. It is the improved decay $e^{-O(R/\sqrt{\ell})}$ given by the majority lift and more refined analysis that makes $R = \Theta((\log n)^{3/2})$ affordable and gives (38).

Consequences for total proof size. Finally, we derive *size* lower bounds for Res(PC) with no degree restriction as corollaries of our *line size* lower bounds. In particular using a simple restriction argument applied to the CNF $G' = G \circ \text{Sel}_q$ where Sel_q is a selector gadget (Definition 6.1), a line-size lower bound for G becomes a total size bound of q^d for G', for a well-chosen choice of parameter settings (Theorem 6.3). Applied to Theorem 1.3 this gives the following.

Theorem 1.5 (Size lower bounds for Res(PC) without a degree restriction)**.** *Let F and $G = F \circ \text{Maj}_\ell$ be as in Theorem 1.4, let $N = n\ell$ be the number of variables of G, and let $q = N$. Then $G \circ \text{Sel}_q$ is an unsatisfiable CNF of size polynomial in n, and every* Res(PC) *refutation of it, with no bound on the degree of its polynomials, has total size*

$$T \ge \exp(\Omega((\log n)^{5/2})).$$

We note that starting instead with Pang's lower bound for F, the same argument gives only $T \geq \exp(\Omega(\log^2 n))$.

1.2 The main idea

Pang's Proof. Write every line as $L = (u_1 = 1) \vee \cdots \vee (u_s = 1)$ with $\deg u_i \leq r$, so that L is true exactly when some $u_i = 1$, and false exactly when its complement $\overline{L} = \prod_i (1 + u_i)$ equals 1. Pang assigns to each line a low-degree *approximator*

$$P_L = \prod_{\nu=1}^{t} q_{L,\nu}, \qquad q_{L,\nu} = 1 + \sum_{i=1}^{s} a_{\nu i} u_i, \quad a_{\nu i} \in \mathbb{F}_2,$$

built exactly as in the Razborov–Smolensky approximation of an OR. If L is false then every u_i is zero, hence every $q_{L,\nu}$ equals 1, and $P_L = 1$. If $P_L = 0 \implies$ *some* $q_{L,\nu} = 0 \implies$ *some* $u_i = 1 \implies$ *L is true*. So P_L can only err in one direction: $P_L = 1$ while L is true, i.e. while some $u_i = 1$. The polynomials $P_L u_i$ record precisely these errors. One then looks for a single *multiplier* f of degree at most K that cancels every error for every line, in the algebraic sense

$$f P_L u_i \in C_D \quad \text{for all } L, i, \qquad \text{but} \qquad f \notin C_D,$$

where C_D is the space of polynomials derivable from F in PC degree D. The first family of conditions lets one follow the refutation line by line and derive $f P_L \in C_D$ for every L (Subsection 3.1); at the empty final line $P_L = 1$, giving $f \in C_D$ and contradicting the second condition. Everything therefore reduces to a counting question: each error condition is a homogeneous linear equation on the coefficients of f, and a suitable f exists as soon as the total number of independent equations, summed over all S lines, is less than the number h of degrees of freedom in choosing f. Denoting V_K to be the vector space generated by all monomials of degree at most K, h is the codimension of $C_D \cap V_K$ in V_K.

The degree barrier. Adding a factor q to P_L makes every candidate polynomial f of the form $f = (1 + q)v$ with $\deg v \leq K - r$ a solution of the new error conditions, since $fq = 0$ on the Boolean cube. Pang's matrix lemma (Theorem A.5 in Appendix A) shows that a good choice of q turns these into many *new* independent solutions, and the net effect of one factor is to cut the number of independent equations by the factor

$$1 - \eta, \qquad \eta = \tfrac{1}{2} \frac{M_n(K - r)}{M_n(K)},$$

where $M_n(j) = \sum_{k=0}^{j} \binom{n}{k}$ counts the monomials of degree at most j. After adding t factors to P_L, the line L contributes at most $e^{-\eta t} h$ linear equations on coefficients of f, making a total of $S e^{-\eta t} h$ linear equations. If $S e^{-\eta t} h < h$, then the desired $f \notin \mathcal{C}_D$ exists and so the lower bound is $\log S \geq \eta t$. The approximating polynomials P_L have degree rt and all of f, P_L, u_i must be multiplied together inside PC degree $D = cn$, which forces $K + 2rt \lesssim D$. This is the obstacle: with K a constant fraction of n below $n/2$, each step down in degree shrinks $M_n(\cdot)$ by a constant factor, so $\eta \approx (K/n)^r$ is exponentially small in r and the bound is nontrivial only for $r = O(\log n)$. To make η close to 1 one would need K close to $n/2$, where the binomial coefficients are flat; but then f has degree far above D and PC cannot multiply by it.

Our Improvement. We prove lower bounds for substantially larger degrees by lifting F with majority gadgets. Substituting majorities gives us a way to decouple the two roles of "degree". In the lifted variables, every polynomial has a unique expansion

$$f(x) = \sum_{A \in \mathcal{A}} x^A f_A(y), \qquad y_i = \mathrm{Maj}_\ell(x_{i,1}, \ldots, x_{i,\ell}),$$

where A selects at most $(\ell - 1)/2$ bits from each block and each coefficient f_A is a multilinear polynomial in the *original* variables (Lemma 4.5). We measure f in two ways: its ordinary degree $\deg_x f$, which is what the monomial counting sees, and its coefficient degree $\beta(f) = \max_A \deg_y f_A$, which is what PC sees. The key example is a monomial that selects exactly $(\ell-1)/2$ bits from every block: its ordinary degree is about $N/2$, while $\beta = 0$ because its only coefficient polynomial is the constant 1. More generally, $\beta(f)$ is the largest number of blocks in which a monomial of f selects *more* than half the bits (Lemma 4.8). Multiplication by a polynomial v raises β by at most $\deg_x v$ (Lemma 4.10), and the lifted consequence space $\widehat{C}_D = \{f : f_A \in C_D \text{ for all } A\}$ is closed under such multiplications as long as the coefficient degree stays below D. Pang's simulation therefore goes through verbatim provided we impose $\beta(f) \le Q := cn/16$, and this restriction says nothing about $\deg_x f$.

Why the gain is $\exp(-O(r/\sqrt{\ell}))$. We may now take the ordinary degree of f close to $N/2$, namely

$$K = \left\lfloor \tfrac{N}{2} - an\sqrt{\ell} \right\rfloor.$$

The offset $an\sqrt{\ell}$ is forced by the restriction $\beta(f) \le Q$: a uniformly random monomial of degree K selects on average $\ell/2 - a\sqrt{\ell}$ bits from each block, which is about $2a$ standard deviations below the majority threshold, so a given block is "heavy" with probability $e^{-\Omega(a^2)}$, and a union bound over sets of $Q+1$ blocks shows that all but an $e^{-\Omega(n)}$ fraction of the monomials of each degree through K satisfy $\beta \le Q$ (Lemma 4.19). Thus the extra restriction costs only an exponentially small fraction of the available coefficients of f, which is what the modified rank argument of Appendix B absorbs. On the other hand, at this K the ratio of consecutive binomial coefficients is

$$\binom{N}{K-1} \Big/ \binom{N}{K} = \frac{K}{N-K+1} \;\approx\; 1 - \frac{4a}{\sqrt{\ell}},$$

so stepping down r degrees costs only a factor $\exp(-O(r/\sqrt{\ell}))$ rather than $\exp(-O(r))$ (Lemma C.1). This replaces $\eta \approx (K/n)^r$ by $\eta \ge \frac{1}{2}\exp(-Cr/\sqrt{\ell})$, and the line bound becomes $\log S \gtrsim \frac{n}{r}\, e^{-Cr/\sqrt{\ell}}$. Nothing else in the argument changes: the number of factors is still $t = \Theta(D/r)$, and $D = cn$ is still the PC degree of the *original* formula, because PC only ever interacts with the coefficient polynomials.

The price. Block length ℓ buys degree $r \approx \sqrt{\ell}\log n$ before the bound degenerates, but the lifted CNF has $N = n\ell$ variables, width $O(\ell)$ and encoding length $\exp(\Theta(\ell))$ once $\ell \gg \log n$. Taking $\ell = \Theta(\log n)$ keeps G polynomial-size and reaches $r = (\log n)^{3/2}$; taking $\ell = \Theta(r^2)$ reaches any $r = o(n^{1/3})$ at the cost of a formula of size $\exp(\Theta(r^2))$, against which the bound $\exp(\Omega(n/r))$ is still superpolynomial. Note that already at $r \ge w(G) = \Theta(\ell)$ a single line can express the negation of any input clause of G, so a semantic refutation always exists in the whole range we consider. Therefore the lower bounds here are necessarily syntactic.

1.3 Recent related work

In a recent manuscript, Conneryd [Con26] proves that a random k-CNF on n variables with $n2^k$ clauses, for $\omega(1) \le k \le O(\log n)$, requires polynomial-calculus degree $n/2 - O(n/\sqrt{k})$ over every field. Pang's degree–size relation needs a degree this close to $n/2$ to say anything about $\mathrm{AC}^0[p]$-Frege (Conneryd points out that even at depth two it requires $n/2 - o(n)$), and for $k = \Theta(\log n)$ he obtains size lower bounds of $\exp(\Omega((\log n)^{3/2}))$ for depth-two $\mathrm{AC}^0[p]$-Frege [Con26, Corollary 1.2]. His work is independent of ours, and we learned of it only after obtaining our main results. The two are closely related: applying Pang's argument directly to a random $(\log n)$-CNF, with his degree bound in place of the lift, gives the $\mathrm{Res}(\mathrm{PC}_r)$ lower bounds of Section 4 for $\ell = \Theta(\log n)$; and running the proof of Theorem 1.4 with a random $(\log n)$-CNF in place of $F \circ \mathrm{Maj}_\ell$ would give the same lower bound for depth-3.5 $\mathrm{AC}^0[2]$-Frege. We stress that this depth-3.5 statement is not in [Con26], whose results concern depth two.

We nevertheless regard the lift as the more promising approach, for two reasons. First, it is a black-box transfer: it starts from degree $\Omega(n)$, which is known for many formula families, rather than from a degree within $o(n)$ of $n/2$ proved for one family. Second, it is not a degree bound in disguise. By substitution, $F \circ \mathrm{Maj}_\ell$ has polynomial-calculus degree at most $\ell \cdot \deg_{\mathrm{PC}}(F) + O(w\ell)$, a fraction of N no larger than $\deg_{\mathrm{PC}}(F)/n$, so Pang's relation applied to the lifted formula itself gives nothing; the argument runs instead over the space $\widehat{C}_D$, which is the consequence space of no formula. Pang's framework uses only a few properties of the target space, and Section 4 is the first instance in which that space is not the consequence space of the formula being refuted. This separates two routes for $\mathrm{AC}^0[2]$-Frege: One is a degree bound: if for every constant C there is a polynomial-size CNF of width $\mathrm{polylog}(n)$ with PC degree at least $n/2 - n/(\log n)^C$ over $\mathbb{F}_2$ (the open problem stated in [Con26]), then the proof of Theorem 1.4, run without the lift, gives $\log T \ge \Omega((\log n)^{C+1}/b)$ for bottom degree b, hence superpolynomial lower bounds for all $b = \mathrm{polylog}(n)$ and so, by the normal form of [BKZ15], for $\mathrm{AC}^0[2]$-Frege. The other is constructing a better target space: the restriction $\beta(f) \le D$ imposed by the input clauses is what pins the multiplier's degree at $N/2 - \Theta(N/\sqrt{\ell})$, and whether a different construction can do better is a question that arises only on this route.

Secondly, a recent paper [AGIPRT26] obtains exponential line lower bounds for $\mathrm{Res}(\mathrm{PC}_r)$ over $\mathbb{F}_2$ when $r \le \varepsilon \log n$. This paper also proves superpolynomial lower bounds for semantic depth-four polynomial calculus with unrestricted bottom fan-in, and semantic depth-d $\mathsf{AC}^0[2]$-Frege lower bounds for unsatisfiable formulas of depth-d. The main distinction is that [AGIPRT26] is stronger in the sense that all of the results obtained there hold for semantic versions of the proof systems, whereas the results of this paper only hold for syntactic systems, which is a necessary restriction for proving $\mathrm{AC}^0[2]$-Frege lower bounds.

Organization. Section 2 gives the definitions. Section 3 presents Pang's proof, and Section 4 proves the majority improvement. Section 5 proves the lower bound for depth-3.5 $\mathrm{AC}^0[2]$-Frege, and Section 6 removes the degree restriction for Res(PC) by a random-restriction argument. Appendix A proves the one-factor selection argument, Appendix B supplies the rank argument with the coefficient-degree restriction, Appendix C the counting estimates, and Appendix D the rules of depth-3.5 $\mathrm{AC}^0[2]$-Frege and the verification that the two-level approximation preserves them.

2 Definitions

Definition 2.1 (Boolean polynomials and degree)**.**
Throughout this paper all polynomials are over $\mathbb{F}_2$. We identify a function on a Boolean cube with its unique multilinear polynomial. Products used as identities of functions are multilinearized, using $z^2 = z$ for each variable z. For any formal polynomial p, the notation $\mathrm{ML}(p)$ denotes its unique multilinear representative on the Boolean cube, obtained by replacing every positive power of a variable by that variable. The degree of such a function is the degree of its multilinear polynomial.

Definition 2.2 (Encoding a CNF by polynomial equations)**.**
Let F be a width-$w(F)$ CNF in the Boolean variables $y_1, \ldots, y_n$. We encode F by polynomial assertions with right-hand side one, whose Boolean solutions are exactly the satisfying assignments of F. The standard convention in polynomial calculus and in Res(PC) is that a polynomial p asserts $p = 0$. Throughout this paper we instead write every assertion with right-hand side one, $p = 1$. This is purely a matter of notational convenience: $p = 1$ is the same assertion as $1 + p = 0$, and the reader may translate at any time by passing to the complement $1 + p$, which changes no degree and at most doubles the number of monomials. We adopt it because the lines of Res(PC) below are disjunctions of equations $u_i = 1$, and with this convention the polynomial u_i is the truth indicator of its own equation, which keeps the approximation argument of Sections 3 and 4 free of complements.

For a literal λ, write $\lambda(y)$ for its truth polynomial: y_i for the positive literal y_i, and $1 + y_i$ for the negative literal $\neg y_i$. For each clause J, define its truth polynomial and assert

$$a_J(y) = 1 + \prod_{\lambda \in J} (1 + \lambda(y)), \qquad a_J(y) = 1.$$

The polynomial a_J is one exactly when J is satisfied; its complement $\overline{J} := 1 + a_J = \prod_{\lambda \in J} \overline{\lambda}$, where $\overline{\lambda} := 1 + \lambda$, is one exactly when J is falsified. The *polynomial encoding of* F consists of these assertions, together with the Boolean axioms $y_i^2 = y_i$, for $i \in [n]$, which in the right-hand-side-one convention read $1 + y_i^2 + y_i = 1$.

Definition 2.3 (Polynomial calculus)**.**
A polynomial-calculus (PC) derivation starts with the assertions in the polynomial encoding of F.

To perform the algebraic calculations, represent an assertion $p = 1$ by the polynomial $1 + p$, obtained by moving the right-hand side to the left. We call $\overline{p} := 1 + p$ the *complement* of the assertion; it vanishes exactly when the assertion is true. (Over $\mathbb{F}_2$, the field used throughout, $1 + p$ is the negation of the Boolean function p.) The same bar is used for variables and literals: $\overline{y}_i = 1 + y_i$ is the truth polynomial of the negative literal $\neg y_i$.

PC may take a scalar linear combination of two previously derived complements, or multiply a previously derived complement by a variable. A complement a represents the assertion $1 + a = 1$. The degree of a derivation is the maximum formal degree appearing in it. A *refutation* derives $0 = 1$, whose complement is the constant polynomial 1.

Definition 2.4 (The space C_D)**.**
Let $F(y_1, \ldots, y_n)$ have width $w(F) \le D$, and let G_F be its polynomial encoding. The vector space C_D consists of the multilinear complements obtainable from G_F by PC derivations of degree at most D. Thus

$$p \in C_D \quad \Longleftrightarrow \quad \text{PC derives } 1 + p = 1 \text{ from } F \text{ in degree at most } D.$$

In particular, if F has no degree-D PC refutation, then $1 \notin C_D$. The linear spaces and matrices below use these complements; the asserted proof lines have right-hand side one.

Multilinearization can be implemented with the Boolean equations without increasing the degree of the polynomial being reduced. We record the following simple multiplication rule for the vector space C_D.

Fact 2.5 (Multiplication rule)**.**

$$u \in C_D, \qquad \deg_y u + \deg_y v \leq D \quad \Longrightarrow \quad uv \in C_D. \tag{2}$$

Definition 2.6 (Res(PC))**.**
Res(PC) generalizes Res($\oplus$) by replacing disjunctions of linear equations with disjunctions of polynomial equations. A line has the form

$$L = (u_1 = 1) \vee \cdots \vee (u_s = 1),$$

where the u_i's are polynomials over $\mathbb{F}_2$. We call $u_i = 1$ the *equations* of L and the u_i its *polynomials*. On a Boolean assignment the equation $u_i = 1$ holds exactly when the polynomial u_i takes the value one, so u_i is the truth indicator of its own equation, and L is true exactly when some $u_i = 1$. (Recall from Definition 2.2 that we write equations as $u = 1$ rather than the standard $u = 0$; Pang [Pan26] uses the standard form, and the two are interchanged by complementing every displayed polynomial, with no change in the number of lines or the degree bound.) For a line L, let $\overline{L}$ denote the *complement of* L:

$$\overline{L} := \prod_i \overline{u}_i = \prod_i (1 + u_i).$$

So $\overline{L}$ is the product of the complements of its equations; as a Boolean function $\overline{L} = 1 + L$, i.e. $\overline{L}$ is one exactly when L is false.

The rules are weakening, contraction, deletion of the false constant equation $0 = 1$, linear combination

$$\frac{A \vee (u = 1) \qquad B \vee (v = 1)}{A \vee B \vee (w = 1)}, \qquad \overline{w} = \overline{u} + \overline{v},$$

and multiplication

$$\frac{A \vee (u = 1)}{A \vee (w = 1)}, \qquad \overline{w} = \overline{u}\, v.$$

Both are sound because $\overline{u} = \overline{v} = 0$ forces $\overline{w} = 0$; they are the usual rules applied to the complements $\overline{u}$ and $\overline{v}$. The default axioms are the true equation $1 = 1$ and, for each variable, the line $(y_i = 1) \vee (\overline{y}_i = 1)$. The latter is of course a tautology, but it is not the same line as $1 = 1$: its two equations are non-constant and can be acted on by the polynomial rules. It is what gives the system the Boolean relation $y_i^2 = y_i$, which the formal multiplication rule does not know by itself (multiply the first equation by y_i, the second by $\overline{y}_i$, and contract to obtain $y_i \overline{y}_i = 0$). It plays the role of $x \vee \neg x$ in resolution over parities. Input clauses are expressed as disjunctions of degree-one equations $(\lambda = 1)$, one per literal.

A refutation of F is a sequence of such lines, each of which is an axiom, an input clause, or follows from earlier lines by a rule, and whose final line is the empty disjunction. Its degree is the maximum formal degree of a polynomial in it. The system Res(PC_r) restricts every polynomial in the refutation, including the conclusions of rules, to degree at most r. Without this restriction we write Res(PC).

Definition 2.7 (Line Size and Total Proof Size)**.**
For a Res(PC) refutation Π, its *line size*, $S(\Pi)$ is the number of lines in Π. Its *total size*, $T(\Pi)$ is the number of monomial occurrences in Π, where every polynomial in Π is written as an explicit sum of monomials. This is the usual monomial-size measure for polynomial calculus; it differs from the size of Π in bits by a factor polynomial in the number of variables, which is immaterial for the superpolynomial bounds proved here. Note that $S \le T$.

Sections 3 and 4 count only lines and therefore use S, regardless of the number of monomials in an equation. Sections 5 and 6 count the total size T.

Definition 2.8 (Monomial count)**.**
Let $M_n(K) := \sum_{j=0}^{K} \binom{n}{j}$ denote the number of multilinear monomials over n variables of degree at most K. Let $V_{K,n}$ denote the space of all multilinear polynomials over n variables of degree at most K. Note that $\dim(V_{K,n})$, the dimension of $V_{K,n}$ is equal to $M_n(K)$.

3 Pang's proof

Fix an S-line Res(PC_r) refutation Π of $F(y_1, \dots, y_n)$, and assume F has no PC refutation of degree at most D. Recall that C_D records complement polynomials as in Definition 2.4, and $1 \notin C_D$. We will obtain a contradiction if S is too small. Products and polynomial identities are interpreted after multilinearization, as in Section 2.

We choose a low-degree product P_L for every proof line L, and *one common polynomial* f for the entire proof. The assertion $P_L = 0$ approximates L. Multiplication by f corrects its errors in the sense of membership in C_D. We choose f outside C_D; the local rules will force f into C_D, giving the contradiction.

Definition 3.1 (Lower Bound Conditions and Approximators)**.**
Let Π be an alleged Res(PC_r) refutation, and write each line as $L = \bigvee_{i=1}^{s}(u_{L,i} = 1)$. When the line is fixed, we write u_i for its polynomials. For each line choose t factors

$$q_{L,\nu} = 1 + \sum_{i=1}^{s} a_{\nu i} u_{L,i}, \qquad a_{\nu i} \in \mathbb{F}_2, \qquad \nu = 1, \dots, t, \tag{3}$$

and define its *line approximator* by

$$P_L = \prod_{\nu=1}^{t} q_{L,\nu}, \qquad \text{so that } \deg P_L \le tr.$$

At the final empty line every factor is one and $P_L = 1$. The coefficients $a_{\nu i}$ are free parameters of the definition; they are chosen in Subsection 3.2, where the choice is what makes the counting work, and nothing in Subsection 3.1 depends on it. This is exactly the Razborov–Smolensky approximation of the OR of the u_i. We want to show that there exists line approximators $\{P_L \mid L \in \Pi\}$ plus a single polynomial f, the *multiplier*, that together satisfies:

(0) (Degrees) $\deg f \le K$, where the parameters satisfy

$$r \le K \le n, \qquad t \ge 1, \qquad K + 2rt + 2r + 2w(F) \le D. \tag{4}$$

(1) (Error conditions) $fP_L u_{L,i} \in C_D$ for every line L and every equation $u_{L,i} = 1$ of that line.

(2) (Condition on the multiplier) $f \notin C_D$.

3.1 Simulation: what the conditions say, and what the induction will show

Before the formal argument it is worth being explicit about the meaning of the three objects P_L, f and fP_L, because the induction below is nothing but this meaning written out in polynomial calculus.

1. *P_L is a one-sided approximation of $\overline{L}$.* Recall that $\overline{L} = \prod_i (1 + u_i)$ is one exactly when L is false. If L is false at an assignment then every u_i is zero there, so every factor $q_{L,\nu}$ equals one and $P_L = 1$. Thus P_L never claims that a false line is true. It can only err in the other direction: $P_L = 1$ while L is true. The *error set* of P_L (with respect to L) is

$$\mathrm{Err}(P_L) := \{a \in \{0,1\}^n : P_L(a) = 1 \text{ and } L \text{ is true at } a\} = \bigcup_i \{a : P_L(a)u_i(a) = 1\},$$

   so the products $P_L u_i$ are exactly the indicators of the errors that P_L makes with respect to L.

2. *f is the indicator of the assignments on which we trust the approximation.* Consider the error conditions (1) pointwise for a moment: $fP_L u_i = 0$ as a function says that f vanishes on $\mathrm{Err}(P_L)$. In other words, on the support $\operatorname{supp} f = \{a : f(a) = 1\}$ no approximator ever errs, so on $\operatorname{supp} f$ each P_L *is* the complement $\overline{L}$. The actual condition (1) replaces "$fP_L u_i$ vanishes as a function" by "PC derives the assertion $1 + fP_L u_i = 1$ from F in degree D". Neither implies the other in any useful sense: since F is unsatisfiable, every polynomial is semantically a consequence of F, so membership in C_D says nothing about the values of $fP_L u_i$ and everything about degree. It is what allows the derivations below to use the condition. (In the terminology of Appendix A, the pointwise reading is the case $C = \{0\}$ of the one-factor argument and the actual condition is the case $C = C_D$.)

3. *$fP_L \in C_D$ says that degree-D PC proves L on $\operatorname{supp} f$.* Since P_L is exact on $\operatorname{supp} f := f^{-1}(1)$, the assertion $1 + fP_L = 1$ says: wherever $f = 1$, the line L is not false. So a PC derivation of it from F is a PC proof of the *relativised* line "L holds on $\operatorname{supp} f$", or equivalently a derivation of L from $F \wedge (f = 1)$; and $fP_L \in C_D$ says more than that L holds there, namely that polynomial calculus proves it in degree D. The induction below converts the given refutation Π, line by line, into such relativised derivations, all in degree D. At the empty line $P_L = 1$ and the relativised line reads $f = 0$: the conversion ends with a degree-D refutation of $F \wedge (f = 1)$. In words, Π yields a PC refutation that is correct on every assignment outside the error sets, and we never have to look at the remaining assignments because f is zero there. Condition (2) says precisely that no degree-D derivation from F ends with $f = 0$. (Semantically $f = 0$ *does* follow from F, since F is unsatisfiable; the content of (2) is entirely about degree.) The contradiction between the converted proof and (2) is the lower bound.

Proposition 3.2 (Simulation)**.** *Suppose f and the approximators P_L satisfy conditions (0) and (1) of Definition 3.1. Then $fP_L \in C_D$ for every line L of Π. In particular, taking L to be the final empty line, $f \in C_D$.*

Corollary 3.3. *No f satisfies all three conditions of Definition 3.1 simultaneously. Consequently, if Subsection 3.2 produces approximators and an f satisfying (0)–(2) whenever S is small, the refutation Π cannot have that few lines.*

The rest of this subsection proves Proposition 3.2. Throughout we use the multiplication rule (2): $u \in C_D$ and $\deg u + \deg v \le D$ imply $uv \in C_D$. All identities are between multilinear polynomials, i.e. between Boolean functions; in particular $\overline{u}\, u = 0$ for every u.

Three facts about a one-sided approximator

The induction rests on one algebraic identity, which is the exact polynomial form of "P_L is one-sided", and on two consequences of it.

Lemma 3.4 (One-sidedness as an identity)**.** *For every line $L = \bigvee_{i=1}^{s}(u_i = 1)$ there are polynomials $U_{L,1}, \dots, U_{L,s}$ of degree at most $(t-1)r$ with*

$$1 + P_L = \sum_{i=1}^{s} U_{L,i}\, u_i. \tag{5}$$

Proof. Multiplying out one factor at a time, $\prod_{\nu\le\mu} q_{L,\nu} = \prod_{\nu<\mu} q_{L,\nu} + (1+q_{L,\mu}) \prod_{\nu<\mu} q_{L,\nu}$, and summing over $\mu = 1, \dots, t$ telescopes to

$$1 + P_L = \sum_{\mu=1}^{t} (1 + q_{L,\mu}) \prod_{\nu<\mu} q_{L,\nu}.$$

Substituting $1 + q_{L,\mu} = \sum_i a_{\mu i} u_i$ and collecting the coefficient of each u_i gives (5) with $U_{L,i} = \sum_\mu a_{\mu i} \prod_{\nu<\mu} q_{L,\nu}$. □

The identity says that $1 + P_L$ lies in the ideal generated by the polynomials of L. In particular, if L is false, so that every u_i is zero, then $P_L = 1$: the approximator never declares a false line true. This is the one-sidedness of P_L, in a form that can be multiplied by other polynomials.

The first consequence concerns a line L obtained from L' by a structural rule (weakening, contraction, deletion of $0 = 1$): then every equation of L' with nonzero polynomial occurs in L, and semantically L' implies L, so "L false" implies "L' false". The lemma is this implication, $P_L \Rightarrow P_{L'}$, relativised to $\operatorname{supp} f$, where the approximators are exact.

Lemma 3.5 (Transfer between lines)**.** *Let L, L' be lines such that every equation of L' with nonzero polynomial also occurs in L. If f satisfies the error conditions for L, then*

$$f P_L (1 + P_{L'}) \in C_D,$$

with all products of degree at most $K + 2rt$.

Proof. By Lemma 3.4 for L', $f P_L (1 + P_{L'}) = \sum_i U_{L',i}\, (f P_L u_{L',i})$. Each nonzero $u_{L',i}$ is a polynomial of L, so $f P_L u_{L',i} \in C_D$ by the error conditions for L; multiply by $U_{L',i}$, of degree at most $(t-1)r$, using $K + rt + r + (t-1)r = K + 2rt \le D$. □

Lemma 3.6 (The surviving disjunct is true)**.** *Let $L' = A \vee (u = 1)$ be a line with $f P_{L'} \in C_D$, and let L be a line containing every equation of A (and possibly others). If f satisfies the error conditions for L, then*

$$\overline{u}\, f P_L \in C_D. \tag{6}$$

The derivation uses products of degree at most $K + 2rt + r$, and the polynomial in (6) has degree at most $K + rt + r$.

Proof. Apply Lemma 3.4 to L', separating the term for the equation $u = 1$:

$$f P_L (1 + P_{L'}) = \sum_{i \in A} U_{L',i}\, (f P_L u_i) \;+\; U_{L',u}\, f P_L u.$$

The sum over A lies in C_D exactly as in Lemma 3.5, since every u_i with $i \in A$ is a polynomial of L. Adding $P_L \cdot (f P_{L'}) \in C_D$ (degree $K + 2rt$) therefore gives

$$f P_L + U_{L',u}\, u\, f P_L \in C_D.$$

Multiply by $\overline{u}$, of degree at most r. Since $\overline{u}\, u = 0$, the second term disappears and (6) remains. □

Semantically: L' holds on $\operatorname{supp} f$ (induction hypothesis); where L is false all of A is false; so on $\operatorname{supp} f \cap \{P_L = 1\}$ the only disjunct of L' that can be true, $u = 1$, must be true. That is $\overline{u}\, f P_L = 0$.

The last fact is what turns the error conditions into the conclusion $f P_L \in C_D$. The error conditions alone do not give it: they say that $f P_L u_j \in C_D$ for each equation, i.e. they control the points of $\operatorname{supp} f \cap \{P_L = 1\}$ at which some equation of L is *true*, and say nothing about the points at which the equations are all false. The extra hypothesis below supplies exactly that: for a set J of equations of L, it says that on $\operatorname{supp} f \cap \{P_L = 1\}$ the equations in J cannot all be false. Together the two cover every point, and the telescoping identity in the proof is this case split written algebraically. In the simulation J is a single equation for the inference rules, and the set of all equations of the clause for an input clause.

Lemma 3.7 (Closing a line)**.** *Let L be a line, let J be a set of equations $u_j = 1$ of L, and put $g = \prod_{j \in J} \overline{u}_j$. If $g\, f P_L \in C_D$ and f satisfies the error conditions for L, then $f P_L \in C_D$. The products used have degree at most $K + rt + \deg g + \max_{j \in J} \deg u_j$.*

Proof. Order J as $j_1, \ldots, j_k$ and expand $1 = \prod_m (u_{j_m} + \overline{u}_{j_m})$ by the telescoping identity

$$1 = \prod_{m=1}^{k} \overline{u}_{j_m} + \sum_{m=1}^{k} \Big(\prod_{m'<m} \overline{u}_{j_{m'}} \Big) u_{j_m}.$$

Multiplying by $f P_L$, the first term is $g\, f P_L \in C_D$ by hypothesis, and the m-th summand is a multiple of $f P_L u_{j_m} \in C_D$ by the error conditions. □

Proof of Proposition 3.2

We prove $f P_L \in C_D$ by induction along Π, one case per rule. In each case the strategy is the same: the premises give relativised lines $f P_{L'} \in C_D$; Lemma 3.6 turns each premise into the statement that its active equation is true on $\operatorname{supp} f \cap \{P_L = 1\}$; the rule's soundness turns these into the statement that the new equation of L is true there; and Lemma 3.7 contradicts this with the error conditions for L, which say that it is false there. All degree bounds quoted are at most $K + 2rt + 2r + 2w(F) \le D$ by (4).

Default axioms. For $L = (1 = 1)$ the only polynomial is the constant 1, so the error condition reads $f P_L \in C_D$ directly. For $L = (y = 1) \vee (\overline{y} = 1)$, the polynomial of the second equation is $\overline{y}$, so the error condition for it reads $\overline{y}\, f P_L \in C_D$; this is the hypothesis of Lemma 3.7 with $J = \{y = 1\}$. Hence $f P_L \in C_D$. (Degree $K + rt + 2$.)

Input clauses. $L = (\lambda_1 = 1) \vee \cdots \vee (\lambda_s = 1)$ with $s \le w(F)$, whose complement $\overline{L} = \prod_j \overline{\lambda}_j$ is exactly the polynomial $\overline{C}$ of Definition 2.2 for this clause C; since $a_C = 1$ is an input assertion, $\overline{L} \in C_D$. Multiplying by $f P_L$ gives $\overline{L}\, f P_L \in C_D$ (degree $K + rt + w(F)$), which is the hypothesis

of Lemma 3.7 with J the set of all equations of L (degree $K + rt + w(F)$). Hence $fP_L \in C_D$. Semantically: on $\operatorname{supp} f \cap \{P_L = 1\}$ every literal of the clause is false, so the clause is falsified, which the input assertion forbids.

Weakening, contraction, deletion of $0 = 1$. Every equation of the premise L' with nonzero polynomial occurs in the conclusion L (the deleted equation $0 = 1$ has polynomial 0). By Lemma 3.5, $fP_L(1 + P_{L'}) \in C_D$, and the induction hypothesis gives $P_L \cdot (fP_{L'}) \in C_D$. Their sum is fP_L. (Degree $K + 2rt$.)

Linear combination. Premises $A \vee (u = 1)$ and $B \vee (v = 1)$, conclusion $L = A \vee B \vee (w = 1)$ with $\overline{w} = \overline{u} + \overline{v}$. Both A and B are contained in L, so Lemma 3.6 applied to each premise gives

$$\overline{u}\, fP_L \in C_D, \qquad \overline{v}\, fP_L \in C_D, \qquad \text{hence} \qquad \overline{w}\, fP_L = \overline{u}\, fP_L + \overline{v}\, fP_L \in C_D.$$

This is the hypothesis of Lemma 3.7 with $J = \{w = 1\}$, so $fP_L \in C_D$. (Degree $K + 2rt + r$.)

Multiplication. Premise $A \vee (u = 1)$, conclusion $L = A \vee (w = 1)$ with $\overline{w} = \overline{u}\, v$, where $\deg \overline{w} \le r$ but v is unrestricted. Lemma 3.6 gives $\overline{u}\, fP_L \in C_D$. Multiply by the polynomial $\overline{w}$ of the conclusion, not by v; since $\deg \overline{w} \le r$, the product has degree at most $K + rt + 2r$. Since $\overline{u}\,\overline{w} = \overline{u}\,\overline{u}\, v = \overline{w}$, this gives $\overline{w}\, fP_L \in C_D$. Lemma 3.7 with $J = \{w = 1\}$ then gives $fP_L \in C_D$. Note that the multiplier v never enters the calculation; only the conclusion does.

The empty line. It is obtained by one of the rules above, and $P_L = 1$; the corresponding case gives $f = fP_L \in C_D$. □

Remark (What was used). The induction used exactly three things: the one-sidedness identity (5), which is a property of the *form* of P_L and holds for any choice of the factors; the error conditions (1), always for the *conclusion* of the rule being simulated; and the degree budget (4). The choice of the factors matters only in Subsection 3.2, where it is made so that the error conditions impose few independent linear conditions on f and an $f \notin C_D$ survives. The simulation itself is indifferent to that choice. This separation is what makes the argument portable: in Section 4 the same induction is run verbatim over the lifted consequence space $\widehat{C}_D$, with the multiplication rule (2) replaced by Lemma 4.10 and deg replaced by the coefficient degree β.

3.2 One multiplier for all lines

In this subsection we prove the existence of approximators P_L and a multiplier f satisfying the conditions of Definition 3.1. Precisely, we prove Corollary 3.13 below: if

$$\log S < \eta t, \qquad \eta = \tfrac{1}{2}\,\frac{M_n(K - r)}{M_n(K)},$$

then there are approximators P_L, one for each line, and an $f \in V_K \setminus C_D$ satisfying all the error conditions. Together with Corollary 3.3 this gives $\log S \ge \eta t$ for every refutation, which is Pang's bound once the parameters are set in Subsection 3.3.

The argument is linear algebra. Recall that C_D is a linear subspace of the polynomials, so for a fixed line and a fixed approximator the error conditions $fP_L u_i \in C_D$ are homogeneous linear equations in

the coefficients of f. We choose the approximators so that the sum, over all lines, of the numbers of independent equations is small; the common solution space then has dimension larger than $V_K \cap C_D$, and hence contains a polynomial $f \notin C_D$. Definition 3.8 specifies exactly how many independent equations we can afford.

Definition 3.8 (Multiplier space and the threshold h)**.**
For a degree bound j, let V_j be the vector space of multilinear polynomials in $y_1, \ldots, y_n$ of degree at most j. Its dimension is $M_n(j)$, as defined in Definition 2.8. Put

$$B = V_K \cap C_D, \qquad h = M_n(K) - \dim B.$$

Thus B consists of the possible multipliers already belonging to C_D. Since $1 \notin C_D$, we have $h > 0$.

For each L, we will show that the subspace of good f's can be described by the solutions to a linear system of less than h/S equations (where the underlying variables correspond to the coefficients of the polynomial f). Therefore by a union bound, the subspace of f's that are simultaneously good for all lines can be described by solutions to a linear system with less than h equations. Since fewer than h independent homogeneous equations on the $M_n(K)$ coefficients of f leave a solution space of dimension greater than $\dim B$, there must exist a single $f \in V_K$ that is good for all lines in Π and that is not in C_D, thus satisfying condition (2) as well.

Definition 3.9 (Error matrices)**.**
Choose a basis of the subspace C_D in reduced row echelon form with respect to the monomial basis of the polynomials: each basis element has a distinct leading monomial, its *pivot*, and the pivots do not occur in the other basis elements. Extend it to a basis of all multilinear polynomials by adding the monomials that are not pivots, the *nonpivot monomials*. Every polynomial p is then uniquely an element of C_D plus a combination of nonpivot monomials; call the latter combination $\mathrm{rem}_D(p)$, the *remainder of p modulo C_D*. The map $p \mapsto \mathrm{rem}_D(p)$ is linear, and $\mathrm{rem}_D(p) = 0$ exactly when $p \in C_D$.

Now fix a line $L = \bigvee_{i=1}^{s}(u_i = 1)$ and an intermediate approximator $P = \prod_\nu q_\nu$ for it. For $0 \le j \le K$, let

$$T_{L,P,j} \colon V_j \longrightarrow (\text{polynomials})^s, \qquad f \longmapsto (\mathrm{rem}_D(fPu_1), \ldots, \mathrm{rem}_D(fPu_s)),$$

the linear map that sends a multiplier f of degree at most j to the remainders of the s products that the error conditions for L require to lie in C_D. As a matrix, $T_{L,P,j}$ has one column for each monomial $y^A = \prod_{a \in A} y_a$ with $|A| \le j$, and one row for each pair (i, μ) of an equation $u_i = 1$ of L and a nonpivot monomial μ, with entry the coefficient of μ in $\mathrm{rem}_D(y^A P u_i)$. Its kernel and rank are

$$\begin{aligned} Z_P(j) &= \{f \in V_j : fPu_i \in C_D \text{ for every } i\} = \ker T_{L,P,j}, \\ c_P(j) &= \operatorname{rank} T_{L,P,j} = M_n(j) - \dim Z_P(j). \end{aligned}$$

We drop L from the notation when the line is fixed.

Thus $f \in V_j$ satisfies the error conditions for L with approximator P exactly when $f \in \ker T_{L,P,j}$, and $c_P(j)$ is the number of independent linear equations that this membership imposes on the $M_n(j)$ coefficients of f. Our goal is therefore to choose an approximator P_L for every line L and then to find a single $f \in V_K$ with $f \notin C_D$ and $T_{L,P_L,K}(f) = 0$ for all lines L. Being in the kernel of all these maps simultaneously imposes at most $\sum_L c_{P_L}(K)$ independent equations on f, while $h = M_n(K) - \dim B$ is the number we can afford: if fewer than h independent homogeneous equations are imposed on

V_K, the solution space has dimension greater than $\dim B$ and so is not contained in $B = V_K \cap C_D$. Thus we want to choose the approximators so that

$$\sum_L c_{P_L}(K) < h. \tag{7}$$

Choosing the factors of each approximator. We work on one line at a time, always at the same fixed degree K, and build the approximator of the line one factor at a time, starting from $P = 1$. The next lemma is the key step: each new factor can be chosen to remove a fixed fraction of the remaining independent conditions. Applied t times to each line, it makes every $c_{P_L}(K)$ exponentially small in t, and then (7) follows by summing over the S lines. For the starting point, let $L = \bigvee_i (u_i = 1)$ be a line in Π and $P = 1$. Every $f \in B$ satisfies $fu_i \in C_D$, by multiplication within degree $K + r \le D$, so $B \subseteq Z_1(K)$ and

$$c_1(K) \le h. \tag{8}$$

Lemma 3.10 (One-factor gain)**.** *Let $L = \bigvee_i (u_i = 1)$ be a line of Π, so that $\deg u_i \le r$, and let $P = \prod_{\nu=1}^{j} q_\nu$, with $j \le t - 1$, be an intermediate approximator for L; recall that $K + 2rt + 2r \le D$ by (4). Then some factor $q = 1 + \sum_i a_i u_i$, with $a_i \in \mathbb{F}_2$, has the following property. Setting $P' = Pq$, we have*

$$c_{P'}(K) \le c_P(K) - \tfrac{1}{2} c_P(K - r). \tag{9}$$

The update multiplies each product fPu_i by q, so it preserves every old good multiplier and makes every polynomial $(1 + q)v$, with $v \in V_{K-r}$, a good multiplier. The main task is to choose q so that these candidates contribute at least $c_P(K-r)/2$ independent solutions beyond the old solution space. Appendix A proves Lemma 3.10. It does so by proving a more general statement (Theorem A.2) in which C_D is replaced by an arbitrary subspace C of the multilinear polynomials that is closed under multiplication within the degree budget D, as in Fact 2.5; nothing else about C_D is used. The lemma is the case $C = C_D$.

The gain involves degree $K - r$, whereas the current conditions are counted at degree K. Lemma 3.11 compares the two.

Lemma 3.11 (Degree comparison)**.** *For every intermediate P,*

$$c_P(K - r) \ge \frac{M_n(K - r)}{M_n(K)} c_P(K). \tag{10}$$

Proof. Order monomials by degree, breaking ties lexicographically; the *leading monomial* of a nonzero polynomial is its largest monomial in this order. Row reduction of a basis of $Z_P(K)$ shows that $c_P(j)$ counts the monomials of degree at most j that are not leading monomials of elements of $Z_P(K)$. Here we use $Z_P(j) = Z_P(K) \cap V_j$.

These counted monomials are closed under taking divisors. Indeed, if a monomial is leading for a solution, multiplying that solution by a variable absent from the leading monomial makes the enlarged monomial leading for another solution, as long as its degree is at most K. The products fPu_i remain in C_D because $K + \deg P + r \le D$.

Let b_j count the counted monomials of degree exactly j. Each counted degree-$(j + 1)$ monomial has $j + 1$ counted divisors, while each degree-j monomial has at most $n - j$ extensions. Therefore

$$(j + 1) b_{j+1} \le (n - j) b_j.$$

It follows that $b_j/\binom{n}{j}$ is nonincreasing in j. Its cumulative weighted average

$$\frac{c_P(j)}{M_n(j)} = \frac{\sum_{a=0}^{j} b_a}{\sum_{a=0}^{j} \binom{n}{a}}$$

is therefore also nonincreasing in j. □

Apply both lemmas and iterate. Choose q using Lemma 3.10 and set $P' = Pq$. Substituting Lemma 3.11 into the term being subtracted gives the one-factor inequality

$$c_{P'}(K) \le c_P(K) - \tfrac{1}{2}\frac{M_n(K-r)}{M_n(K)} c_P(K) = (1-\eta)\, c_P(K), \qquad \eta = \frac{1}{2}\,\frac{M_n(K-r)}{M_n(K)}, \tag{11}$$

where η is the fraction of conditions removed per step. We now iterate this, separately for each line.

Proposition 3.12 (The approximation estimate)**.** *Suppose $t, r \ge 1$, $r \le K$, and $K + (t+2)r \le D$, and let $\eta = \frac{1}{2}M_n(K-r)/M_n(K)$. For each line there is an approximator P_L built from t factors of the form $1 + \sum_i a_i u_i$ such that*

$$c_{P_L}(K) \le (1-\eta)^t\, h \le e^{-\eta t}\, h. \tag{12}$$

Proof. Fix a line $L = \bigvee_i (u_i = 1)$. Define approximators $P^{(0)}, P^{(1)}, \dots, P^{(t)}$ for L inductively: $P^{(0)} = 1$, and $P^{(j+1)} = P^{(j)} q_j$, where $q_j = 1 + \sum_i a_i u_i$ is the factor that Lemma 3.10 provides for the line L and the approximator $P^{(j)}$. Before the last factor is added, $\deg P^{(j)} \le (t-1)r$, so $K + (t+2)r \le D$ implies the degree hypotheses of Lemmas 3.10 and 3.11, and (11) gives $c_{P^{(j+1)}}(K) \le (1-\eta)\, c_{P^{(j)}}(K)$ for each $j < t$. By induction, starting from $c_{P^{(0)}}(K) = c_1(K) \le h$ (8), we get $c_{P^{(t)}}(K) \le (1-\eta)^t h$. Put $P_L = P^{(t)}$. □

Corollary 3.13 (A common multiplier)**.** *Under the hypotheses of Proposition 3.12, if*

$$t \ge \frac{\log S}{\eta}, \tag{13}$$

then there are approximators and an $f \in V_K$ with $f \notin C_D$ and $f P_L u_i \in C_D$ for every line L and polynomial u_i of L. Consequently every S-line refutation satisfies

$$\log S \ge \eta t. \tag{14}$$

Proof. Stack the error matrices $T_{L,P_L,K}$ of all S lines. By (12) their total rank is at most $\sum_L c_{P_L}(K) \le S e^{-\eta t} h$, which is less than h when (13) holds strictly; the common solution space then has dimension greater than $\dim B$, so it contains an $f \notin B$, i.e. $f \in V_K \setminus C_D$. Such an f contradicts Proposition 3.2, so $t < \log S/\eta$, which is (14). □

3.3 Setting the parameters and obtaining the bound

It remains to choose K and t within the budget (4) and to estimate η. The estimate rests on the elementary bound

$$\frac{M_n(j-1)}{M_n(j)} \ge \frac{j}{n+1}, \tag{15}$$

which follows from $j\binom{n}{j} = (n-j+1)\binom{n}{j-1} \le (n-j+1) M_n(j-1)$.

Theorem 3.14 (Pang's degree–size relation; Theorem 1.1 restated)**.** *Let F be an unsatisfiable CNF on n variables with no PC refutation of degree at most D, where* $\max\{32r, 8w(F)\} \le D \le n$*. Then every S-line syntactic* $\mathrm{Res}(\mathrm{PC}_r)$ *refutation of F satisfies*

$$\log S \ge \frac{D}{128r}\left(\frac{D}{16n}\right)^r. \tag{16}$$

Proof. Take $K = \lfloor D/4 \rfloor$ and $t = \lfloor D/(16r) \rfloor$. These satisfy (4) and $K+(t+2)r \le D$, and $t \ge D/(32r)$. Multiplying (15) for $j = K - r + 1, \dots, K$, where $j \ge 7D/32$, gives $\eta \ge \frac{1}{2}(D/(16n))^r$. Substituting into (14) gives (16). □

For $D = \Theta(n)$ and constant input width, this is a line-size lower bound of $\exp(\Omega(n))$ for constant r, and $\exp(n^{\Omega(1)})$ for r a sufficiently small constant multiple of $\log n$.

4 Majority substitution with variable block length

Pang's argument has one quantitative bottleneck, already identified in Section 1: each new factor of P_L removes a fraction $\eta = \frac{1}{2}M_n(K-r)/M_n(K)$ of the remaining error conditions, and because the degree K of the multiplier must stay a constant fraction below $n/2$, this fraction is about $(K/n)^r$, exponentially small in r. The useful degree range is therefore $r = O(\log n)$. Near degree $N/2$ the binomial coefficients are flat and the same fraction would be close to $1/2$; but a multiplier of degree $N/2$ is far outside the range in which polynomial calculus can multiply by it.

Majority substitution resolves this by giving a polynomial in the lifted variables *two* degrees. We expand it in a basis whose coefficients are polynomials in the original variables (Subsection 4.1). The ordinary degree $\deg_x f$, which is what the monomial counting sees, may then approach $N/2$; the *coefficient degree* $\beta(f)$, which is what polynomial calculus sees, stays below the original threshold D. The rest of the section does three things: it defines the lifted consequence space $\widehat{C}_D$ and checks that the simulation of Section 3.1 runs over it (Subsection 4.2); it shows that imposing $\beta(f) \le Q$ costs an exponentially small fraction of the available multipliers, so that Pang's rank argument survives with an additive loss (Subsection 4.3); and it sets the parameters and proves Theorem 4.20 (Subsection 4.4). The majority basis and the counting ideas come from the majority-based correlation arguments [Smo87, CHL+26]. Appendix B contains the modified rank argument and Appendix C the counting estimates.

4.1 The lifted formula and the majority basis

Definition 4.1 (Majority)**.**
Let ℓ be odd and put $m = (\ell-1)/2$. On $\{0,1\}^\ell$, let H_0 be the set of inputs with at most m ones and H_1 the set of inputs with at least $m+1$ ones. The function Maj_ℓ is zero on H_0 and one on H_1.

Definition 4.2 (The lifted CNF)**.**
For each original variable y_i introduce a block $x_i = (x_{i,1}, \dots, x_{i,\ell})$ of ℓ new variables, and put

$$h_i(x) = \mathrm{Maj}_\ell(x_i), \qquad N = n\ell;$$

the polynomial h_i is the lifted variable y_i, and it is used below to expand polynomials in the lifted variables. The CNF G is F with y_i replaced by h_i: replace each positive literal y_i of F by the CNF $\bigwedge_{|A|=m+1}\bigvee_{a\in A} x_{i,a}$ of $\mathrm{Maj}_\ell(x_i)$, each negative literal by the CNF $\bigwedge_{|A|=m+1}\bigvee_{a\in A}\neg x_{i,a}$ of $\neg\,\mathrm{Maj}_\ell(x_i)$,

and distribute each clause into CNF. The result is $G = F \circ \mathrm{Maj}_\ell$. Throughout, C_D continues to denote the consequence space of F in the *original* variables.

A clause of width w in F produces at most $2^{w\ell}$ clauses of width at most $w(m+1)$ in G. For a polynomial-size, constant-width F this gives

$$N = n\ell, \qquad w(G) \le w(\ell+1)/2, \qquad \log|G| = O(\log n + \ell),$$

where $|G|$ is the written encoding length. So $\ell = \Theta(\log n)$ keeps G polynomial-size, and larger blocks are paid for in input length.

The ideal of the majority. Fix the ℓ variables $x_1, \dots, x_\ell$ of a single block, and write $h = \mathrm{Maj}_\ell(x_1, \dots, x_\ell)$ for its majority. For this definition we will be working in the Boolean ring of multilinear polynomials over this block of variables. Call a monomial $x^\sigma = \prod_{a\in\sigma} x_a$ *light* if $|\sigma| \le m$ and *heavy* if $|\sigma| \ge m+1$; since $\ell = 2m+1$, there are $2^{\ell-1}$ of each. Let $V_{\le m}$ be the span of the light monomials, i.e. the polynomials of degree at most m. The following lemma is the algebraic form of the splitting of a Boolean function along the majority used in Smolensky's argument.

Lemma 4.3 (The ideal of the majority)**.** *In the Boolean ring of one block:*

(i) The multilinear polynomial h is a sum of heavy monomials, and $x^\sigma h = x^\sigma$ for every heavy monomial x^σ. Consequently the ideal $\langle h\rangle$ generated by h is exactly the span of the heavy monomials, which is also the space of functions vanishing on H_0; and the light monomials form a basis of the quotient by $\langle h\rangle$.

(ii) Every nonzero polynomial vanishing on H_1 has degree at least $m+1$. Consequently $v \mapsto hv$ is injective on $V_{\le m}$, and $\{hx^\sigma : |\sigma| \le m\}$ is a basis of $\langle h\rangle$.

(iii) Every polynomial p has a unique expansion $p = u + hv$ with $\deg u, \deg v \le m$.

Proof. (i) The coefficient of x^σ in the multilinear expansion of a Boolean function g is $\sum_{\tau\subseteq\sigma} g(\mathbf{1}_\tau)$ modulo 2, where $\mathbf{1}_\tau$ is the input with ones exactly on τ. If $|\sigma| \le m$ then every $\mathbf{1}_\tau$ with $\tau \subseteq \sigma$ has at most m ones, so $h(\mathbf{1}_\tau) = 0$ and the coefficient vanishes: h is a sum of heavy monomials. If x^σ is heavy and $x^\sigma(x) = 1$ then x has at least $m+1$ ones and $h(x) = 1$, so $x^\sigma h = x^\sigma$ as functions. Hence every heavy monomial lies in $\langle h\rangle$; conversely the span of the heavy monomials is an ideal (multiplying a heavy monomial by a variable gives a heavy monomial or itself) that contains h, so it contains $\langle h\rangle$. Heavy monomials vanish on H_0, and they are $2^{\ell-1}$ independent elements of the space of functions vanishing on H_0, which has dimension $2^\ell - |H_0| = 2^{\ell-1}$; so they span it. The light monomials are the remaining elements of the monomial basis, hence a basis of the quotient.

(ii) Substituting $1 + x_a$ for each x_a exchanges H_0 and H_1 and preserves degree, so by (i) the functions vanishing on H_1 are spanned by the products $\prod_{a\in\sigma}(1+x_a)$ with $|\sigma| \ge m+1$. Each such product has x^σ as its unique monomial of top degree $|\sigma|$, and these top monomials are distinct for distinct σ; so in a nonzero combination the monomials x^σ of the largest $|\sigma|$ present do not cancel, and the combination has degree at least $m+1$. If now $hv = 0$ with $\deg v \le m$, then v vanishes on H_1, where $h = 1$, so $v = 0$. Thus $v \mapsto hv$ is injective on $V_{\le m}$, and its image, of dimension $2^{\ell-1}$, is all of $\langle h\rangle$.

(iii) Split p into its light and heavy monomials, $p = u + p_{\mathrm{heavy}}$. By (i) and (ii), $p_{\mathrm{heavy}} = hv$ for a unique $v \in V_{\le m}$. If $u + hv = u' + hv'$ with all four of degree at most m, then $u + u' = h(v + v')$ is both a combination of light monomials and, by (i), of heavy ones, so it is zero, and $v = v'$ by injectivity. □

We now pass from one block to all n. The Boolean ring in the N lifted variables is the tensor product of the n one-block rings, so a basis of each block ring gives, by taking products, a basis of the whole. In block i, Lemma 4.3(iii) gives the basis consisting of the light monomials x^σ and the products $h_i x^\sigma$, $|\sigma| \le m$. Taking one basis element from each block and multiplying gives the elements $x^A h^J$: here A collects the light monomials chosen in the blocks, so A selects at most half the bits of every block, and $J \subseteq [n]$ is the set of blocks in which the factor h_i was chosen. The reason this is the right basis for us is that the $h_i = \mathrm{Maj}_\ell(x_i)$ are the lifted variables y_i. So h^J is the monomial y^J in the original variables, and once a polynomial in the lifted variables is written in this basis, grouping its terms by A expresses it as a combination of the x^A with coefficients that are polynomials in the original variables, which is what polynomial calculus for F can act on. This is the expansion of Lemma 4.5 below; the next definition names this basis.

Definition 4.4 (The majority basis)**.**
For $A \subseteq [n] \times [\ell]$ let $A_i = \{a : (i,a) \in A\}$ be its part in block i, and $x^A = \prod_{(i,a)\in A} x_{i,a}$. Call the monomial x^A *light* if it selects at most half the bits of every block, and put

$$\mathcal{A} = \{A : |A_i| \le m \text{ for every } i\}, \qquad U = \mathrm{span}_{\mathbb{F}_2}\{x^A : A \in \mathcal{A}\}.$$

The *majority basis* of the Boolean ring in the lifted variables is $\{x^A h^J : A \in \mathcal{A},\ J \subseteq [n]\}$, where $h^J = \prod_{i\in J} h_i$; by the preceding paragraph it is a basis.

Lemma 4.5 (Majority expansion)**.** *Every polynomial p in the lifted variables has a unique expansion*

$$p(x) = \sum_{A\in\mathcal{A}} x^A\, p_A\big(h_1(x), \dots, h_n(x)\big), \tag{17}$$

with every $p_A(y_1, \dots, y_n)$ multilinear.

Proof. Write p in the majority basis and collect, for each A, the terms $x^A h^J$ into $x^A p_A(h)$; uniqueness is uniqueness of coordinates in a basis. □

Definition 4.6 (Scalar coefficients, coefficient polynomials, and the two degrees)**.**
A polynomial p in the lifted variables has two expansions,

$$p(x) = \sum_T c_T x^T = \sum_{A\in\mathcal{A}} x^A p_A(h_1, \dots, h_n).$$

The $c_T \in \mathbb{F}_2$ are its *scalar coefficients* and the $p_A(y)$ its *coefficient polynomials*. The *coefficient degree* of p is

$$\beta(p) = \max_{A\in\mathcal{A}} \deg_y p_A, \qquad \beta(0) = -\infty,$$

and $\deg_x p$ is its ordinary degree in the lifted variables. For an integer Q write $W_Q = \{p : \beta(p) \le Q\}$.

The two degrees are independent of one another. A monomial that selects exactly m bits from every block has $\deg_x = nm$, close to $N/2$, and $\beta = 0$: it is itself a basis element x^A with coefficient polynomial 1. Conversely $h_i = \mathrm{Maj}_\ell(x_i)$ has $\beta = 1$ and ordinary degree ℓ. The following lemma says exactly which scalar coefficients β sees: $\beta(p)$ is the largest number of blocks in which a monomial of p selects *more* than half the bits.

Definition 4.7 (Heavy blocks of a monomial)**.**
For a monomial x^T, let $J(T) = \{i : |T_i| > m\}$ be the set of blocks in which it selects more than half the bits.

Lemma 4.8 (Coefficient degree in scalar coordinates)**.** *If $p = \sum_T c_T x^T \neq 0$, then*

$$\beta(p) = \max\{|J(T)| : c_T \neq 0\}.$$

Consequently $p \in W_Q$ if and only if $c_T = 0$ whenever $|J(T)| > Q$.

Proof. On one block, by Lemma 4.3(i) and (ii), the heavy monomials and the products $h\,x^\sigma$ with $|\sigma| \leq m$ span the same space $\langle h \rangle$, while the light monomials span $V_{\leq m}$. Taking products over blocks, the ordinary monomials x^T with $J(T) = J$ span exactly $h^J U$; the coefficient polynomials of a nonzero element of $h^J U$ are nonzero multiples of y^J; and the spaces $h^J U$ for different J form a direct sum by Lemma 4.5. Hence components with different J cannot cancel, and $\beta(p)$ is the largest $|J|$ present. □

Thus, for the unknown multiplier $f = \sum_{|T| \leq K} c_T x^T$, the condition $\beta(f) \leq Q$ is simply a set of coordinate restrictions: the scalars c_T with more than Q heavy blocks are zero. All the error conditions of Section 3 are linear in the same scalars, so the two kinds of condition can be counted together.

4.2 The lifted consequence space and the lifted simulation

We now say what plays the role of C_D for the lifted formula. Rather than the degree-D consequences of G itself, we use the space of polynomials in the lifted variables all of whose coefficient polynomials, in the majority basis, are consequences of F.

Definition 4.9 (The space $\widehat{C}_D$)**.**

$$\widehat{C}_D = \{p(x) : p_A(y) \in C_D \text{ for every } A \in \mathcal{A}\}, \tag{18}$$

with p_A the coefficient polynomials of (17).

This is a linear subspace of the polynomials in the lifted variables, since the expansion (17) is linear. We make no claim that it coincides with the degree-D consequences of G. What we need of it is what the argument of Section 3 needed of C_D. The simulation of Subsection 3.1 used two properties: the multiplication rule (Fact 2.5) and the fact that the complements $\overline{C}$ of the input clauses lie in C_D. We re-prove both for $\widehat{C}_D$, as Lemmas 4.10 and 4.11, and the simulation then transfers verbatim. The counting of Subsection 3.2 used one more, $1 \notin C_D$, and its replacement is the following. A nonzero element of U has at least one coefficient polynomial equal to the constant 1, and $1 \notin C_D$, so

$$U \cap \widehat{C}_D = \{0\}. \tag{19}$$

Since U contains all light monomials, including the 2^{-n}-fraction of all monomials counted in Lemma 4.19, this says that $\widehat{C}_D$ misses a very large space; this is where the room for the multiplier will come from.

Lemma 4.10 (Multiplication in majority coordinates)**.** *For all polynomials p, v in the lifted variables, $\beta(pv) \leq \beta(p) + \deg_x v$. Moreover,*

$$p \in \widehat{C}_D, \qquad \beta(p) + \deg_x v \leq D \quad \Longrightarrow \quad pv \in \widehat{C}_D.$$

Both statements hold with $\deg_x v$ *replaced by the number of blocks on which* v *depends.*

Proof. Multiplying a one-block basis element x^{A_i}, $|A_i| \le m$, by a bit $x_{i,a}$ of the same block gives either another such element (if $a \in A_i$ or $|A_i| < m$) or a heavy monomial, which by Lemma 4.3(ii) equals $h_i v$ with $\deg v \le m$. In either case the coefficient polynomials of the product are the old ones multiplied by polynomials of degree at most one in y_i. Multiplying by a monomial of degree d repeats this d times; the multiplying coefficient polynomials then have degree at most d and involve only the original variables of the blocks that the monomial touches. After collecting terms, each coefficient polynomial of pv is a sum $\sum_B p_B q_{A,B}(y)$ with $\deg q_{A,B}$ bounded as stated, and Fact 2.5 applied to each product $p_B q_{A,B}$ gives the second assertion. □

Note that β is not submultiplicative: on a three-bit block $\beta(x_1) = \beta(x_2) = 0$ but $\beta(x_1 x_2) = 1$. The bound $\beta(p) + \deg_x v$ is the one we use, with p the polynomial whose membership in $\widehat{C}_D$ is known and v a factor of degree $O(rt)$.

Lemma 4.11 (Lifted input clauses)**.** *Let* $1 + a(x) = 1$ *be the encoding of an input clause of* G*, with* $a = \overline{C}$ *its complement. If* $\beta(f) \le Q$ *and* $Q + 2w(F) \le D$*, then* $fa \in \widehat{C}_D$*.*

Proof. The clause of G comes from a clause of F with complement $R(y)$. If the clause of G is false at x, then one clause of each substituted literal's CNF is false, so each substituted literal is false and the clause of F is false at $y = h(x)$. Hence $a = R(h)\,a$ as functions. Also a involves at most $w(F)$ blocks, so by Lemma 4.10 the expansion $fa = \sum_A x^A v_A(h)$ has $\deg_y v_A \le Q + w(F)$. From $fa = R(h) fa$ and uniqueness of the expansion, $v_A = R v_A$ for every A. Since $R \in C_D$ is an input complement and $\deg R + \deg v_A \le 2w(F) + Q \le D$, Fact 2.5 gives $v_A = R v_A \in C_D$, i.e. $fa \in \widehat{C}_D$. □

With these two lemmas in hand, the simulation of Section 3.1 transfers:

Proposition 4.12 (Lifted simulation)**.** *Let* Π *be a* Res(PC$_r$) *refutation of* G*, with each* P_L *of the form in Definition 3.1 using* $t \ge 1$ *factors of degree at most* r*. Suppose* $\beta(f) \le Q$*, the error conditions* $f P_L u_{L,i} \in \widehat{C}_D$ *hold for all lines and equations, and*

$$Q + \max\{2rt + 2r,\ rt + w(G)\} \le D, \qquad Q + 2w(F) \le D. \tag{20}$$

Then $f P_L \in \widehat{C}_D$ *for every line of* Π*, and in particular* $f \in \widehat{C}_D$*.*

Proof. Run Lemmas 3.4–3.7 and the case analysis in the proof of Proposition 3.2 with C_D replaced by $\widehat{C}_D$, using Lemma 4.10 for every multiplication and tracking β in place of the degree. Each step multiplies a polynomial known to be in $\widehat{C}_D$, of coefficient degree at most $Q + rt + r$, by a factor of ordinary degree at most $rt + r$; the largest coefficient degree reached is $Q + 2rt + 2r$. For an input clause with complement a, Lemma 4.11 gives $fa \in \widehat{C}_D$, and Lemma 3.7 with J the set of all equations of the clause gives $f P_L \in \widehat{C}_D$ within coefficient degree $Q + rt + w(G)$. Both are covered by (20). □

4.3 One multiplier for all lines, after substitution

We need a common f and one approximator per line such that

$$\beta(f) \le Q, \qquad f \notin \widehat{C}_D, \qquad f P_L u_i \in \widehat{C}_D \ \text{ for every line } L \text{ and equation } (u_i = 1) \in L. \tag{21}$$

The first condition is new: it is what Proposition 4.12 and Lemma 4.10 need, and nothing in the counting would otherwise impose it. By Lemma 4.8 it is a collection of zero-coefficient restrictions on the scalars of f. The plan is to run Pang's rank argument on the full space of polynomials of degree at most K, exactly as before, and to account separately for the dimensions lost each time a multiplication forces us to restrict to W_Q.

Definition 4.13 (Multiplier space, threshold h, and excluded coefficients)**.**
Let $V_K(x)$ be the space of multilinear polynomials of degree at most K in the N lifted variables, so $\dim V_K(x) = M_N(K)$; the multiplier will be sought in $V_K(x) \cap W_Q$. As in Definition 3.8, put

$$B = V_K(x) \cap \widehat{C}_D, \qquad h = M_N(K) - \dim B.$$

For $0 \le j \le K$, let E_j be the number of monomials x^T with $|T| \le j$ and $|J(T)| > Q$, so that $E_j = \dim V_j(x) - \dim(V_j(x) \cap W_Q)$. Let ε be the largest excluded fraction in any single degree:

$$\varepsilon = \max_{0 \le j \le K} \frac{\#\{T : |T| = j,\ |J(T)| > Q\}}{\binom{N}{j}}, \qquad \text{so that } E_j \le \varepsilon M_N(j).$$

As before, h is the number of independent linear conditions on $f \in V_K(x)$ that we can afford: fewer than h leave a solution space larger than B, hence a solution outside $\widehat{C}_D$. The number E_K is the price of the restriction to W_Q: for any subspace $Z \subseteq V_K(x)$,

$$\dim(Z \cap W_Q) \ge \dim Z - E_K, \tag{22}$$

since W_Q is cut out of $V_K(x)$ by E_K coordinate equations, and likewise restricting a linear map to $V_K(x) \cap W_Q$ lowers its rank by at most E_K. The whole point of the parameter choice in Subsection 4.4 is to make E_K, and in fact $\varepsilon M_N(K)$, exponentially smaller than h.

Definition 4.14 (Error matrices for the lifted formula)**.**
Definition 3.9 carries over with C_D replaced by $\widehat{C}_D$ and the variables y by x: fix a basis of $\widehat{C}_D$ completed by monomials, let $\widehat{\operatorname{rem}}_D(p)$ be the component of p outside $\widehat{C}_D$, and for a line with polynomials $u_1, \dots, u_s$ and an intermediate approximator P let $T_{P,j}$ be the linear map sending $f \in V_j(x)$ to the scalar coefficients of $(\widehat{\operatorname{rem}}_D(fPu_1), \dots, \widehat{\operatorname{rem}}_D(fPu_s))$. Its kernel and rank are

$$Z_P(j) = \{f \in V_j(x) : fPu_i \in \widehat{C}_D \text{ for every } i\}, \qquad c_P(j) = \operatorname{rank} T_{P,j} = M_N(j) - \dim Z_P(j).$$

The one-factor gain in the lifted setting. The next lemma is the version of the one-factor gain, Lemma 3.10, that we need here. Recall how that gain is obtained (Lemma A.3): appending a factor $q = 1 + \sum_i a_i u_i$ keeps every old solution $f \in Z_P(K)$, because $fPu_i \in C_D$ may be multiplied by q, and creates the new solutions $(1+q)v$, $v \in V_{K-r}$, because $q(1+q) = 0$; the gain is then the number of new solutions independent of the old ones, which Pang's matrix lemma (Theorem A.5) bounds below by $\frac{1}{2} c_P(K-r)$.

Over $\widehat{C}_D$ the same two steps are available, with one proviso. Each multiplication is justified by Lemma 4.10 only when the polynomial being multiplied has coefficient degree at most Q, and a general element of $Z_P(K)$ or of $V_{K-r}(x)$ need not. So before multiplying we restrict both spaces to W_Q:

$$Z_{Pq}(K) \supseteq (Z_P(K) \cap W_Q) + (1+q)(V_{K-r}(x) \cap W_Q), \tag{23}$$

valid when $Q + \deg_x P + 2r \le D$. By (22) the first restriction discards at most E_K dimensions of old solutions and the second at most E_{K-r} dimensions of candidates, after which the matrix lemma

applies exactly as before. The restriction is internal to the proof: the solutions it produces need not lie in W_Q, and $Z_{Pq}(K)$ is still counted in the full space $V_K(x)$. (The requirement $f \in W_Q$ on the final multiplier is imposed once, at the end, in Corollary 4.18.) The result is Lemma 3.10 with the two discarded dimensions as additive terms.

Lemma 4.15 (One-factor gain with coefficient-degree loss)**.** *Suppose $K \geq r$ and $Q+\deg_x P+3r \leq D$. Then some factor $q = 1 + \sum_i a_i u_i$, $a_i \in \mathbb{F}_2$, satisfies*

$$c_{Pq}(K) \leq c_P(K) - \tfrac{1}{2}\, c_P(K-r) + E_K + \tfrac{1}{2}\, E_{K-r}. \tag{24}$$

The gain is again measured at degree $K-r$, and Lemma 3.11 compared that count with $c_P(K)$ by showing that the uncounted monomials are closed under taking divisors. Over $\widehat{C}_D$ the closure step is a multiplication by one variable, justified only on $Z_P(K) \cap W_Q$, so the comparison also acquires a loss.

Lemma 4.16 (Degree comparison with coefficient-degree loss)**.** *Suppose $r \leq K < N/2$ and $Q + \deg_x P + r + 1 \leq D$. Then*

$$c_P(K-r) \geq \frac{M_N(K-r)}{M_N(K)}\, c_P(K) - N^2 \varepsilon M_N(K-r). \tag{25}$$

Both lemmas are proved in Appendix B. Substituting (25) into (24) and using $E_j \leq \varepsilon M_N(j)$ gives the one-factor inequality

$$c_{Pq}(K) \leq (1-\eta)\, c_P(K) + 2N^2 \varepsilon M_N(K), \qquad \eta = \frac{1}{2}\,\frac{M_N(K-r)}{M_N(K)}, \tag{26}$$

which is the one-factor inequality (11) with the new number of variables and an additive leak. At the start, $B \cap W_Q \subseteq Z_1(K)$ when $Q + r \leq D$, so $c_1(K) \leq h + E_K \leq h + \varepsilon M_N(K)$, in place of (8). Iterating (26) t times from this starting point gives the replacement for Proposition 3.12.

Proposition 4.17 (The approximation estimate with coefficient-degree losses; compare Proposition 3.12)**.** *Suppose $t, r \geq 1$, $r \leq K < N/2$, and $Q + (t+2)r \leq D$, and let $\eta = \frac{1}{2} M_N(K-r)/M_N(K)$ as in (26). For each line there is an approximator P_L built from t factors of the form $1 + \sum_i a_i u_i$ such that*

$$c_{P_L}(K) \leq e^{-\eta t}\big(h + \varepsilon M_N(K)\big) + \frac{2N^2 \varepsilon M_N(K)}{\eta}. \tag{27}$$

The first term of (27) is Pang's geometric decrease, now starting from $h + \varepsilon M_N(K)$ rather than h. The second is the accumulated additive loss: each factor leaks at most $2N^2 \varepsilon M_N(K)$ conditions back in, and because only a fraction η of the conditions is removed per step, the leaks sum to at most $1/\eta$ times the per-step loss.

Corollary 4.18 (A common multiplier after substitution; compare Corollary 3.13)**.** *Under the hypotheses of Proposition 4.17, suppose*

$$t \geq \frac{\log(4S)}{\eta}, \qquad \varepsilon M_N(K) \leq \frac{\eta h}{16 S N^2}. \tag{28}$$

Then there are approximators and an f with

$$f \in V_K(x) \cap W_Q, \qquad f \notin \widehat{C}_D, \qquad f P_L u_i \in \widehat{C}_D \ \text{ for all } L, i. \tag{29}$$

Proof. Stack the error matrices of all S lines and append the E_K coordinate equations defining W_Q. The total rank is at most

$$Se^{-\eta t}\big(h+\varepsilon M_N(K)\big)+\frac{2SN^2\varepsilon M_N(K)}{\eta}+\varepsilon M_N(K). \tag{30}$$

Under (28), and since $\eta \le 1/2$ gives $\varepsilon M_N(K) \le h/32$, the three terms are at most $(h+h/32)/4$, $h/8$ and $h/32$, whose sum is less than h. The common solution space therefore has dimension greater than $\dim B$, so it contains some $f \notin B$; as $f \in V_K(x)$, this means $f \notin \widehat{C}_D$. □

Two things remain: to choose K and Q so that $\varepsilon M_N(K)$ is tiny relative to h (not merely relative to $M_N(K)$), and to see that at this K the fraction η is much larger than in Pang's setting.

4.4 Setting the parameters and obtaining the bound

Throughout, F has n variables, width at most w, and no PC refutation of degree $D = \lfloor cn \rfloor$, for fixed $0 < c < 1$; $G = F \circ \mathrm{Maj}_\ell$ has $N = n\ell$ variables. All constants below depend only on c, w, not on ℓ or r.

The choice of Q and K. Take

$$Q = \lfloor cn/16 \rfloor, \qquad K = \lfloor N/2 - an\sqrt{\ell} \rfloor, \tag{31}$$

with $a \ge 1$ a sufficiently large constant depending only on c. The reason for the offset $an\sqrt{\ell}$ is the following. A uniformly random monomial of degree K selects about $\ell/2 - a\sqrt{\ell}$ bits in each block, with standard deviation about $\sqrt{\ell}/2$; so it is roughly $2a$ standard deviations below the majority threshold in a typical block, and a given block is heavy with probability $e^{-\Omega(a^2)}$. For the monomial to be excluded from W_Q it needs more than $Q = \Omega(n)$ heavy blocks, which happens with probability $e^{-\Omega(a^2 n)}$. The offset is thus just large enough to make the restriction $\beta(f) \le Q$ almost free, and no larger, so that K stays as close to $N/2$ as possible.

Lemma 4.19 (The two counts needed for the dimension argument)**.** *For the choices* (31)*, whenever* $K \ge 0$*,*

$$h \ge 2^{-n} M_N(K), \tag{32}$$

$$\varepsilon \le \exp\big[-(ca^2/8 - \log 2)\, n\big]. \tag{33}$$

Taking $ca^2/8 \ge 2\log 2 + 12$ *therefore gives*

$$\frac{\varepsilon M_N(K)}{h} \le e^{-12n}. \tag{34}$$

Proof. By (19), $U \cap V_K(x)$ and $B = V_K(x) \cap \widehat{C}_D$ intersect trivially, so $h \ge \dim(U \cap V_K(x))$. Map each monomial of degree at most K to the monomial obtained by complementing its selected set inside every heavy block. The image lies in U, has no larger degree, and each monomial of U has at most 2^n preimages (one per choice of which blocks were complemented). Hence $\dim(U \cap V_K(x)) \ge 2^{-n} M_N(K)$.

For (33), fix a degree $j \le K$ and $Q+1$ blocks. A uniform monomial of degree j selects on average at most $(Q+1)\ell \cdot K/N \le (Q+1)\ell/2 - (Q+1)a\sqrt{\ell}$ bits in the union of these blocks, whereas being heavy in all of them requires at least $(Q+1)\ell/2$. The tail bound for sampling without replacement in Appendix C bounds this deviation by $e^{-2(Q+1)a^2}$, and a union bound over the $\binom{n}{Q+1} \le 2^n$ choices of blocks gives (33). Dividing by (32) gives (34). Neither estimate needs an upper bound on ℓ. □

Why each factor now removes a large fraction. At $K = N/2 - an\sqrt{\ell}$, consecutive binomial coefficients satisfy $\binom{N}{K-1}/\binom{N}{K} = K/(N-K+1) \approx 1 - 4a/\sqrt{\ell}$, and the cumulative counts $M_N(j)$ inherit this: stepping down one degree loses only a fraction $O(1/\sqrt{\ell})$, so stepping down r degrees loses a factor $\exp(-O(r/\sqrt{\ell}))$ rather than Pang's $\exp(-O(r))$. This is the replacement for the elementary estimate (15); Lemma C.1 in Appendix C makes it precise:

$$\eta^{-1} \le 2\exp\Big(64(a+2)\frac{r}{\sqrt{\ell}}\Big) \qquad (\sqrt{\ell} \ge 128(a+2),\ 1 \le r \le n). \tag{35}$$

Consequently a line of a refutation with S lines needs only $O\big(e^{O(r/\sqrt{\ell})}\log S\big)$ factors, whereas the degree budget $Q + (t+2)r \le D$ allows $t = \Theta(n/r)$ of them. The lower bound is the comparison of these two numbers. Note that nothing here requires $r \le \ell$: the degree of P_L is compared with D only through coefficient degrees.

Theorem 4.20 (Majority extension with variable block length)**.** *Fix $0 < c < 1$ and a positive integer w. There are constants $\ell_0, \kappa, c_1, C > 0$, depending only on c, w, with the following property. Let F_n be an unsatisfiable polynomial-size CNF on n variables of width at most w with no PC refutation over $\mathbb{F}_2$ of degree $D = \lfloor cn \rfloor$. Let $\ell \ge \ell_0$ be odd and $r \ge 1$ an integer with $r + \ell \le \kappa n$. Then for all sufficiently large n, every S-line syntactic* $\mathrm{Res}(\mathrm{PC}_r)$ *refutation of $G_n = F_n \circ \mathrm{Maj}_\ell$ satisfies*

$$\log S \ge c_1 \frac{n}{r}\exp\Big(-C\frac{r}{\sqrt{\ell}}\Big). \tag{36}$$

The formula G_n has $N = n\ell$ variables, width at most $w(\ell+1)/2$, and $\log|G_n| = O(\ell + \log n)$. Hence the bound on lines, and a fortiori on total size, is superpolynomial in $|G_n|$ whenever

$$\frac{n}{r}\exp\Big(-C\frac{r}{\sqrt{\ell}}\Big) = \omega(\ell + \log n). \tag{37}$$

Proof. Choose a with $ca^2/8 \ge 2\log 2 + 12$, and put

$$C = 64(a+2), \qquad \ell_0 = \lceil [128(a+2)]^2 \rceil, \qquad \kappa = \min\{c/128,\ c/(32w),\ 1/2\}.$$

Take Q, K as in (31) and $t = \lfloor D/(16r) \rfloor$. For large n these give $Q \le D/8$, $r \le D/64$, $w(G_n) \le D/8$ and $D/(32r) \le t \le D/(16r)$, while $\ell \ge \ell_0$ ensures $r \le n \le K < N/2$. The degree conditions of Proposition 4.17 and (20) follow:

$$Q + (t+2)r \le \tfrac{D}{8} + \tfrac{D}{16} + \tfrac{D}{32}, \qquad Q + 2rt + 2r \le \tfrac{D}{8} + \tfrac{D}{8} + \tfrac{D}{32}, \qquad Q + rt + w(G_n) \le \tfrac{D}{8} + \tfrac{D}{16} + \tfrac{D}{8},$$

all less than D, and $Q + 2w(F_n) \le D$ for large n.

By (35), $\eta t \ge \frac{D}{64r}\exp(-Cr/\sqrt{\ell})$. If $\eta t < 1$ then $S \ge 2$ already gives $\log S \ge \eta t/4$. So assume $\eta t \ge 1$ and, for a contradiction, $\log S \le \eta t/4$. Dividing the combined rank (30) by h and using (34), it is at most

$$(1 + e^{-12n})e^{-3\eta t/4} + \frac{2N^2}{\eta}\exp(\eta t/4 - 12n) + e^{-12n}.$$

The first term is at most $(1 + e^{-12n})e^{-3/4}$. For the second, $\eta t \le n/32$, $N \le n^2$, and $\eta^{-1} \le 2e^{r/2} \le 2e^{n/2}$ by the choice of ℓ_0, so it is at most $4n^4 e^{-11n}$. The sum is below 1 for large n, so a common multiplier with $\beta(f) \le Q$, $f \notin \widehat{C}_D$ and all error conditions exists, and Proposition 4.12 gives $f \in \widehat{C}_D$, a contradiction.

Thus $\log S \ge \eta t/4$ in both cases, and since $D \ge cn/2$ for large n this is (36) with $c_1 = c/512$. The size and width of G_n are from Definition 4.2, and $T \ge S$ gives the last assertion. □

We record three natural choices of parameter settings:

Corollary 4.21 (Polynomial-size formulas)**.** *Under the hypotheses of Theorem 4.20, let ℓ be a sufficiently large constant times $\log n$, rounded up to an odd integer. Then G_n has polynomial size in $N = \Theta(n \log n)$, and every syntactic $\mathrm{Res}(\mathrm{PC}_r)$ refutation has $\exp(n^{\Omega(1)})$ lines for all $1 \le r \le (\log n)^{3/2}$. For $r = o((\log n)^{3/2})$ the bound is $\exp(n^{1-o(1)})$.*

Corollary 4.22 (Larger polylogarithmic degrees)**.** *Under the same hypotheses, fix $\gamma \ge 1$ and let ℓ be $(\log n)^\gamma$ rounded up to an odd integer. There is $\delta > 0$, depending only on c, w, such that every syntactic $\mathrm{Res}(\mathrm{PC}_r)$ refutation of G_n with $1 \le r \le \delta(\log n)^{1+\gamma/2}$ has $\exp(n^{\Omega(1)})$ lines, while $\log |G_n| = O((\log n)^\gamma)$. The bound is superpolynomial in $|G_n|$.*

Corollary 4.23 (Polynomial degrees)**.** *Under the same hypotheses, fix $0 < \delta < 1/3$, let $r = \lfloor n^{1/3-\delta} \rfloor$, and let ℓ be r^2 rounded up to an odd integer. Every syntactic $\mathrm{Res}(\mathrm{PC}_r)$ refutation of G_n has*

$$S \ge \exp(\Omega(n^{2/3+\delta})), \qquad \log |G_n| = O(n^{2/3-2\delta}),$$

so $S \ge |G_n|^{\Omega(n^{3\delta})}$. In terms of $N = \Theta(n^{5/3-2\delta})$, the degree is $r = \Theta(N^{(1/3-\delta)/(5/3-2\delta)})$.

5 Depth-3.5 $\mathrm{AC}^0[2]$-Frege

Recall from the introduction that a line of depth-3.5 $\mathrm{AC}^0[2]$-Frege is an OR of ANDs of polynomial equations of degree at most b. A *term* $\mathcal{T} = \bigwedge_j (u_{\mathcal{T},j} = 1)$ is an AND of polynomial equations with $\deg u_{\mathcal{T},j} \le b$, and a line is an OR of terms; written out, a line has the form $\mathrm{OR} \circ \mathrm{AND} \circ \mathrm{PARITY} \circ \mathrm{AND}_{\le b}$, with unbounded fan-in at the OR and at the AND and fan-in at most b at the bottom. This is exactly the shape of a cedent of $\mathrm{PCK}^1_{\mathbb{F}_2}$ in [BKZ15], a disjunction of polynomial assertions and of conjunctions of them (Appendix D). The lines are asserted and the refutation ends with the empty disjunction, which is the form in which the approximation argument of Sections 3 and 4 is stated.

Appendix D gives the precise rules (Definition D.1) and shows (Lemma D.2) that our system and degree-b $\mathrm{PCK}^1_{\mathbb{F}_2}$ in the sense of [BKZ15] simulate each other with polynomial overhead. Size means total size T, counting every occurrence of a term and of a monomial.

Recall from the introduction that by [BKZ15] any quasipolynomial-size depth-d $\mathrm{AC}^0[2]$-Frege refutation of a CNF can be converted into a quasipolynomial-size refutation of depth 3.5^+ with bottom degree $(\log n)^{O(d)}$, and that by Lemma D.2 in Appendix D our system is polynomially equivalent to theirs at the same bottom degree. So the difference between the theorem below and a lower bound for quasipolynomial-size depth-d $\mathrm{AC}^0[2]$-Frege is that we handle $b = o(\sqrt{\log n})$ rather than $b = (\log n)^{O(d)}$.

Throughout the section F is a constant-width CNF on n variables with no PC refutation of degree $D = \lfloor cn \rfloor$, and $G = F \circ \mathrm{Maj}_\ell$ with $\ell = \Theta(\log n)$ odd; then G has polynomial size; $N = n\ell$ is the number of variables of G.

From Section 4 we use the lifted consequence space $\widehat{C}_D$, the coefficient-degree bound $Q = \lfloor cn/16 \rfloor$, the multiplier degree $K = \lfloor N/2 - an\sqrt{\ell} \rfloor$ with its subspace W_Q and quantities h, ε, E_K (Definition 4.13), and only the following three facts about them, all for sufficiently large n:

(i) $E_K \le \varepsilon M_N(K)$ and $\varepsilon M_N(K) \le e^{-12n} h$ (Lemma 4.19);

(ii) Let $\eta(d) = \frac{1}{2} M_N(K-d)/M_N(K)$. Then for $1 \le d \le n$: $\eta(d)^{-1} \le 2e^{O(d/\sqrt{\ell})}$.

(iii) $n \le K < N/2$, $N \le n^2$, and $w(G) = O(\log n)$.

Theorem 1.4 (restated)**.** *There is a constant $\alpha > 0$, depending only on c, such that for all sufficiently large n and all $1 \le b \le \sqrt{\ell}$, every depth-3.5 $\mathrm{AC}^0[2]$-Frege refutation of G with bottom degree b has total size*

$$\log T \;\ge\; \alpha \,\frac{(\log n)^{3/2}}{b}. \tag{38}$$

In particular the bound is superpolynomial for all $b = o(\sqrt{\log n})$, and equals $\exp(\Omega((\log n)^{3/2}))$ for constant b.

5.1 The idea

Lines and terms. A line of the system is

$$L = \mathcal{T}_1 \vee \cdots \vee \mathcal{T}_m,$$

an OR of *terms*, and each term is an AND of polynomial equations,

$$\mathcal{T}_k = (u_{k,1} = 1) \wedge \cdots \wedge (u_{k,s_k} = 1), \qquad \deg u_{k,j} \le b.$$

We use calligraphic letters for terms, so that $\mathcal{T}$ is a term and $u_{\mathcal{T},1}, \ldots, u_{\mathcal{T},s}$ are its polynomials; the Roman T remains the total size of the proof.

We will approximate each line of a depth-3.5 $\mathrm{AC}^0[2]$-Frege proof bottom-up, in the manner of Razborov–Smolensky. To fix the notation for the approximators at each level, we next describe the two-level approximation of a line.

Approximating a term. In Section 3 we approximated a line $L = (u_1 = 1) \vee \cdots \vee (u_s = 1)$ of Res(PC) by a polynomial P_L, a product of random affine combinations of the u_i. It equals one whenever every u_i is zero, that is, whenever L is false, so P_L is a one-sided approximation of the negation of L, and the approximating assertion is $P_L = 0$. Now consider a term $\mathcal{T} = (u_{\mathcal{T},1} = 1) \wedge \cdots \wedge (u_{\mathcal{T},s} = 1)$ of a line of our system. It can be written as a negated OR,

$$\mathcal{T} \;\equiv\; \neg\big[(1 + u_{\mathcal{T},1} = 1) \vee \cdots \vee (1 + u_{\mathcal{T},s} = 1)\big],$$

whose disjuncts are the complementary equations. Applying the same construction to this OR, with the complements $1 + u_{\mathcal{T},j}$ in place of the u_i, gives a polynomial $A_{\mathcal{T}}$ that approximates $\mathcal{T}$ itself: $A_{\mathcal{T}} = 1$ whenever $\mathcal{T}$ is true, and $A_{\mathcal{T}}$ can be wrong only at points where $A_{\mathcal{T}} = 1$ while some complement $1 + u_{\mathcal{T},j}$ is 1. The approximating assertion for a term is $A_{\mathcal{T}} = 1$.

Approximating a line. Given $A_{\mathcal{T}_1}, \ldots, A_{\mathcal{T}_m}$, the line L is, up to these approximations, the OR of the equations $A_{\mathcal{T}_1} = 1, \ldots, A_{\mathcal{T}_m} = 1$. So we build the line approximator P_L from $A_{\mathcal{T}_1}, \ldots, A_{\mathcal{T}_m}$ exactly as P_L was built from $u_1, \ldots, u_s$ in Section 3: a product of random affine combinations of the $A_{\mathcal{T}_k}$, one-sided for the falsity of L.

One multiplier for both levels. As before we look for a single multiplier f that corrects all the errors, now at both levels:

$$f A_{\mathcal{T}}(1+u_{\mathcal{T},j}) \in \widehat{C}_D \quad \text{for every term } \mathcal{T} \text{ and each } j, \qquad f P_L A_{\mathcal{T}_k} \in \widehat{C}_D \quad \text{for every line } L \text{ and each } k.$$

On $\operatorname{supp} f$ the first family makes every $A_{\mathcal{T}}$ exact, and the second makes every P_L exact relative to its $A_{\mathcal{T}_k}$. The rank count of Proposition 4.17, applied once to each term and once to each line, shows that such an $f \notin \widehat{C}_D$ exists when the proof is small, and the simulation then forces $f \in \widehat{C}_D$.

The one new point in the simulation. The term rules relate a conjunction $\mathcal{T} \wedge \mathcal{U}$ to its two parts, but the three approximants $A_{\mathcal{T}}$, $A_{\mathcal{U}}$ and $A_{\mathcal{T}\wedge\mathcal{U}}$ were chosen independently of one another. What saves the induction is that the first family of error conditions by itself forces

$$f(A_{\mathcal{T}\wedge\mathcal{U}} + A_{\mathcal{T}}A_{\mathcal{U}}) \in \widehat{C}_D,$$

that is, on $\operatorname{supp} f$ the approximant of a conjunction is the product of the approximants of its parts. With this the argument of Section 3.1 goes through for the term rules as well (Proposition 5.2).

Where the bound comes from. The weakness of the bound is a matter of degree. A term approximant $A_{\mathcal{T}}$ needs about $\log T$ factors of degree b, so its degree is $R = O(b \log T)$. The line approximator P_L is built from polynomials of degree R, and by (35) it then needs about $e^{O(R/\sqrt{\ell})} \log T$ factors, so its degree is of order $b(\log T)^2 e^{O(b \log T/\sqrt{\ell})}$. All of this must stay below a small constant times $D = cn$. With $\ell = \Theta(\log n)$ the exponential factor is $n^{O(\alpha)}$ as long as $\log T \le \alpha\sqrt{\ell}\log n/b$, and the polynomial factor is then harmless; so a refutation with $\log T \le \alpha(\log n)^{3/2}/b$, for a small constant α, is one that the degree budget can handle, and that is the threshold in (38).

5.2 Approximants and error conditions

Definition 5.1 (Approximants and their error conditions)**.**
Choose integers $t_{\text{in}}, t_{\text{out}} \ge 1$, and put $R = bt_{\text{in}}$. For each distinct term and each line choose

$$A_{\mathcal{T}} = \prod_{\nu=1}^{t_{\text{in}}} q_{\mathcal{T},\nu}, \quad q_{\mathcal{T},\nu} = 1 + \sum_{j \in I_{\mathcal{T}}} a_{\mathcal{T},\nu,j}(1 + u_{\mathcal{T},j}),$$
$$P_L = \prod_{\nu=1}^{t_{\text{out}}} q_{L,\nu}, \quad q_{L,\nu} = 1 + \sum_{\mathcal{T} \in L} a_{L,\nu,\mathcal{T}} A_{\mathcal{T}}.$$

All coefficients $a_{\mathcal{T},\nu,j}$ and $a_{L,\nu,\mathcal{T}}$ lie in $\mathbb{F}_2$. Use the same $A_{\mathcal{T}}$ at every occurrence of $\mathcal{T}$. For a singleton use the exact approximant $A_{(u=1)} = u$, padding with factors $q = 1$; for the empty term use $A_{\top} = 1$. Then $\deg A_{\mathcal{T}} \le R$, $\deg P_L \le Rt_{\text{out}}$, and $P_{\varnothing} = 1$. The approximating assertions are $A_{\mathcal{T}} = 1$ and $P_L = 0$, the latter equivalently written as $1 + P_L = 1$. The two families of error conditions on a common f are

$$fA_{\mathcal{T}}(1 + u_{\mathcal{T},j}) \in \widehat{C}_D \qquad \text{for every term and each of its equations,} \tag{39}$$
$$fP_LA_{\mathcal{T}} \in \widehat{C}_D \qquad \text{for every line and each of its terms.} \tag{40}$$

The inner conditions do not involve P_L, so they can be used at every occurrence of the term.

Proposition 5.2 (Preservation of the rules)**.** *Suppose $\beta(f) \le Q$, both error families hold, and*

$$Q + \max\{2Rt_{\text{out}} + 2R,\ Rt_{\text{out}} + w(G)\} \le D, \qquad Q + 2w(F) \le D. \tag{41}$$

Then every line in a refutation satisfies $fP_L \in \widehat{C}_D$. The empty final line consequently gives $f \in \widehat{C}_D$.

The proof is the induction of Section 3.1, run over $\widehat{C}_D$ as in Proposition 4.12, with one additional ingredient: from the inner conditions one derives, for every term $\mathcal{V} = \mathcal{T} \wedge \mathcal{U}$,

$$f(A_{\mathcal{V}} + A_{\mathcal{T}} A_{\mathcal{U}}) \in \widehat{C}_D,$$

and this identity, together with its analogue for projection, carries the three term rules. The rule-by-rule verification, including the degree bookkeeping behind (41), is in Appendix D.

5.3 One multiplier for both levels

For $1 \le d \le K$ write

$$\eta(d) = \frac{1}{2} \frac{M_N(K-d)}{M_N(K)}.$$

We choose all inner approximants first, then all outer approximants, and finally impose the resulting equations on the coefficients of a single $f \in V_K(x) \cap W_Q$.

Corollary 5.3 (The combined rank criterion)**.** *For a refutation of total size T, assume $R \le K < N/2$ and (41). The approximants can be chosen so that all error conditions, together with $f \in W_Q$, have combined rank at most*

$$\begin{aligned} &T\big(e^{-\eta(b)t_{\mathrm{in}}} + e^{-\eta(R)t_{\mathrm{out}}}\big)\big(h + \varepsilon M_N(K)\big) \\ &\qquad + 2TN^2 \varepsilon M_N(K)\big(\eta(b)^{-1} + \eta(R)^{-1}\big) + E_K. \end{aligned} \tag{42}$$

If this is less than h, such a refutation cannot exist.

Proof. Apply Proposition 4.17 to each nonsingleton term, with supplied polynomials $1 + u_{\mathcal{T},j}$, degree b, and t_{in} factors, and use the resulting product as $A_{\mathcal{T}}$; its error conditions are exactly the inner conditions (39). Singletons and empty terms impose no inner constraints. Fix these choices, then apply the same proposition to each line, with supplied polynomials $A_{\mathcal{T}}$, degree R, and t_{out} factors; this gives P_L and the outer conditions (40). The hypotheses $Q + (t_{\mathrm{in}} + 2)b \le D$ and $Q + (t_{\mathrm{out}} + 2)R \le D$ follow from (41). There are at most T terms and T lines, each contributing the bound (27) with $\eta(b)$ and t_{in}, respectively $\eta(R)$ and t_{out}; stacking their error matrices and the E_K coordinate restrictions defining W_Q gives (42). If the total rank is less than h, the common solution space in $V_K(x)$ has dimension greater than $\dim(V_K(x) \cap \widehat{C}_D)$, so it contains an $f \in V_K(x) \cap W_Q$ with $f \notin \widehat{C}_D$ satisfying all error conditions, and Proposition 5.2 gives the contradiction $f \in \widehat{C}_D$. □

5.4 Choosing the parameters

Proof of Theorem 1.4. All constants below depend only on c and w, and n is taken large enough for (i)–(iii) and for the finitely many inequalities marked "for large n". Let $\alpha > 0$ be a small constant, fixed in a moment, and suppose for a contradiction that a refutation of total size T has $\log T \le \alpha (\log n)^{3/2}/b$. Put $L = \log(64T)$, so that $L = O(\alpha(\log n)^{3/2}/b)$, and choose

$$t_{\mathrm{in}} = \Big\lceil \frac{L}{\eta(b)} \Big\rceil, \qquad R = b t_{\mathrm{in}}, \qquad t_{\mathrm{out}} = \Big\lceil \frac{L}{\eta(R)} \Big\rceil.$$

The inner degree. Since $b \le \sqrt{\ell}$, (ii) gives $\eta(b)^{-1} = O(1)$, so $t_{\mathrm{in}} = O(L)$ and

$$R = O(bL) = O\big(\alpha(\log n)^{3/2}\big), \qquad \frac{R}{\sqrt{\ell}} = O(\alpha \log n),$$

the second because $\ell = \Theta(\log n)$. In particular $R \le n$ for large n, so (ii) applies to $d = R$ and gives $\eta(R)^{-1} \le 2n^{O(\alpha)}$. Fix α so small that $\eta(R)^{-1} \le n^{1/4}$.

The outer degree. Then $t_{\text{out}} = O(n^{1/4}L)$ and

$$Rt_{\text{out}} = O(n^{1/4}bL^2) = O(n^{1/4}(\log n)^3) \le \frac{cn}{32}$$

for large n. With $Q \le cn/16$, $R \le Rt_{\text{out}}$ and $w(G) = O(\log n)$, both inequalities in (41) follow: $Q + 2Rt_{\text{out}} + 2R \le cn/16 + cn/8 < D$, $Q + Rt_{\text{out}} + w(G) < D$, and $Q + 2w(F) \le D$. Together with $R \le n \le K < N/2$ from (iii), the hypotheses of Corollary 5.3 hold.

The rank. By the choice of t_{in} and t_{out}, both $e^{-\eta(b)t_{\text{in}}}$ and $e^{-\eta(R)t_{\text{out}}}$ are at most $e^{-L} = 1/(64T)$. Divide (42) by h and use (i), $T \le e^L$, $N^2 \le n^4$, $\eta(b)^{-1} = O(1)$ and $\eta(R)^{-1} \le n^{1/4}$: the combined rank is at most

$$\frac{2(1+e^{-12n})}{64} + n^{O(1)}e^{O((\log n)^{3/2})}e^{-12n} + e^{-12n} \;<\; 1$$

times h for large n. By Corollary 5.3 no such refutation exists, so every refutation has $\log T > \alpha(\log n)^{3/2}/b$, which is (38). The polynomial-overhead simulation of $\mathrm{PCK}^1_{\mathbb{F}_2}$ changes only the constant. □

6 Removing the degree restriction

The bounds of Sections 3–5 restrict the degree of every polynomial in the proof. This section removes that restriction for Res(PC). Theorem 6.3 shows that a small Res(PC) refutation of a slightly larger formula, with no bound on the degrees of its polynomials, restricts to a low-degree refutation of the original formula with few lines; so a line lower bound for $\mathrm{Res}(\mathrm{PC}_{2d})$ becomes a lower bound on the total size (Definition 2.7) of Res(PC) refutations of the larger formula, and Subsection 6.2 applies it to the lifted line bound of Corollary 4.21 to prove Theorem 1.5. The price is that size is measured by the total number of monomial occurrences T rather than by the number of lines S.

The mechanism is a random restriction. Replace each variable of the original formula by the selector function of Definition 6.1, which reads a value off a block of $2q$ bits of which exactly one is on, and then choose uniformly at random one bit of each row in each block to carry the variable, setting all other bits to zero. Any such restriction returns the original formula. A monomial that touches two bits of one block is killed outright, and a monomial that touches j blocks survives with probability exactly q^{-j}. So a proof with fewer than q^d monomial occurrences has, for some restriction ρ, no surviving monomial of degree d or more, by a union bound. Now restrict the entire refutation by ρ, replacing each polynomial u in the proof by the multilinearized version of the restricted polynomial $u|_\rho$. This is still a degree-d refutation Π_ρ, but now of the restricted formula, which is just F' itself. (It needs to be checked that the lines $u|_\rho$ still fit together into a legitimate refutation. The only step that needs a bit of care is the multiplication rule, which we show can be simulated after the restriction at a cost of $(n')^{O(d)}$ lines per multiplication step, negligible for our parameter settings.)

The selector function was chosen to make q as large as possible. The final bound is q^d, so we want $\log q = \Theta(\log n)$, that is, q polynomial, while the substituted formula stays polynomial-size. The selector function encodes a variable in one-hot form, with q bits standing for each truth value, so that both the variable and its negation become disjunctions of q bits; every clause then stays a single clause and the formula grows only linearly in q.

6.1 The selector substitution, its restriction, and the transfer theorem

Definition 6.1 (The selector function and the selector substitution)**.**
Let $q \geq 2$. The *selector function* Sel_q takes $2q$ bits, which we arrange in two rows of q,

$$X = (X_{1,1}, \ldots, X_{1,q};\ X_{0,1}, \ldots, X_{0,q}),$$

and is defined on the *one-hot* inputs, those with exactly one bit equal to 1: $\mathrm{Sel}_q(X)$ is the row $b \in \{0,1\}$ of that bit. The column of the bit carries no information. On one-hot inputs the function and its negation are both disjunctions,

$$\mathrm{Sel}_q(X) = \bigvee_{a=1}^{q} X_{1,a}, \qquad \neg\,\mathrm{Sel}_q(X) = \bigvee_{a=1}^{q} X_{0,a}.$$

Now let F' be a CNF on the variables $y_1, \ldots, y_{n'}$. The CNF $F' \circ \mathrm{Sel}_q$ has, for each $i \in [n']$, a block X_i of $2q$ new variables $X_{i,b,a}$, and consists of two kinds of clauses.

- For each clause C of F', the clause C^{Sel} obtained by replacing each literal y_i by $\bigvee_a X_{i,1,a}$ and each literal $\neg y_i$ by $\bigvee_a X_{i,0,a}$. By the display above, on one-hot inputs C^{Sel} is the clause $C(\mathrm{Sel}_q(X_1), \ldots, \mathrm{Sel}_q(X_{n'}))$.
- For each block i, the *one-hot clauses* C'_i, which force the input of the i-th selector to be one-hot: the clause $\bigvee_{b,a} X_{i,b,a}$ (at least one bit of the block is 1) and the clauses $\neg X_{i,b,a} \vee \neg X_{i,b',a'}$ for every two distinct bits of the block (at most one is).

The formula $F' \circ \mathrm{Sel}_q$ is unsatisfiable: a satisfying assignment makes every block one-hot, and then $y_i := \mathrm{Sel}_q(X_i)$ satisfies F'. Since both literals of y_i are replaced by disjunctions, every clause of F' becomes a single clause C^{Sel}, of width at most $q\,w(F')$. Also $F' \circ \mathrm{Sel}_q$ has $2n'q$ variables and $|F'| + n' + n'\binom{2q}{2}$ clauses where $|F'|$ denotes the number of clauses in F' .

Definition 6.2 (The selector restriction)**.**
With F' and $F' \circ \mathrm{Sel}_q$ as above, the *selector restriction* ρ chooses, independently for each block i and each row $b \in \{0,1\}$, a uniformly random column $A_{i,b} \in [q]$, keeps the two chosen bits of the block and kills the rest:

$$X_{i,1,A_{i,1}} \mapsto y_i, \qquad X_{i,0,A_{i,0}} \mapsto 1 + y_i, \qquad X_{i,b,a} \mapsto 0 \ \ (a \neq A_{i,b}).$$

Every selector restriction returns F': a clause C^{Sel} becomes C together with false equations $0 = 1$; the at-least-one clause of a block contains both y_i and $1 + y_i$; and an exclusion clause contains the constant 1 unless both its bits were chosen, in which case it too contains both y_i and $1 + y_i$.

The next theorem shows that, given a small Res(PC) refutation Π of $F' \circ \mathrm{Sel}_q$ with no restriction on degree, a random selector restriction ρ will, with high probability, convert Π into a refutation of F' of low degree and small line size.

Theorem 6.3 (Restricting a small refutation to low degree)**.** *Let F' be a CNF on $n' \geq 2$ variables, let $d \geq 2$ and $q \geq 2$, and let Π be a* Res(PC) *refutation of $F' \circ \mathrm{Sel}_q$, with no bound on the degree of its polynomials, of total size T. Then with probability at least $1 - Tq^{-d}$ over the choice of a random selector restriction ρ, every polynomial of $\Pi|_\rho$ has degree less than d, and $\Pi|_\rho$ can be completed to a* $\mathrm{Res}(\mathrm{PC}_{2d})$ *refutation of F' with at most $T(n')^{Cd}$ lines, for an absolute constant C. In particular, if every* $\mathrm{Res}(\mathrm{PC}_{2d})$ *refutation of F' has at least L lines, then*

$$T \geq \min\Big\{q^{\,d},\ \frac{L}{(n')^{Cd}}\Big\}. \tag{43}$$

Proof. Step 1: the restriction lowers every degree. Consider one monomial occurrence M in Π (powers replaced by first powers). If M uses two bits of one block, its image is zero for every ρ, because either one of the bits is unchosen or the image contains the factor $y_i(1+y_i)$, which multilinearizes to zero. Otherwise M uses one bit in each of j blocks, each survives with probability q^{-1} independently, and the image is a product of j literals in distinct variables, of degree j. So the probability that M survives with degree at least d is at most q^{-d}. There are at most T monomial occurrences in Π, and a polynomial and its complement differ only in the constant monomial, so by a union bound, with probability at least $1 - Tq^{-d}$, every polynomial u in Π has $\deg \mathrm{ML}(u|_\rho) < d$. Fix such a ρ, and write $z_1, \dots, z_{n'}$ for the surviving variables y_i, identified with the variables of F'.

Step 2: the restricted proof is a proof, after repairing the multiplication steps. A restricted input clause is a clause of F' (possibly with additional false equations $0 = 1$, which are deleted), or contains the equation $1 = 1$ and follows from that axiom by weakening, or contains both equations $(z_i = 1)$ and $(1 + z_i = 1)$ and follows from the default axiom by weakening; see the remarks after Definition 6.2. Restriction and multilinearization commute with addition and with adding the constant one, so each linear-combination step remains a linear-combination step, and weakening, contraction and deletion of false constants remain valid after simplifying constants. The one rule that does not survive is multiplication,

$$A \vee (p = 1) \;\longrightarrow\; A \vee \big(1 + (1+p)v = 1\big),$$

because the multilinearization of a product is not the product of the multilinearizations. Let

$$a = \mathrm{ML}\big((1+p)|_\rho\big), \qquad b = \mathrm{ML}\big(((1+p)v)|_\rho\big)$$

be the restricted complements of the premise and conclusion polynomials; both have degree less than d. As Boolean functions $ab = b$: where $a = 0$ the original product, hence b, is zero, and where $a = 1$ there is nothing to prove. (This uses that the field is $\mathbb{F}_2$.) Hence $b + ab$ has zero multilinearization and formal degree less than $2d$, so it lies in the ideal generated by the Boolean complements $z_i^2 + z_i$: reducing one power at a time writes it as a sum of monomial multiples of these, with $(n')^{O(d)}$ terms. Each $z_i^2 + z_i$ is derivable in two lines from the axiom $(z_i = 1) \vee (1 + z_i = 1)$ (multiply the two complements by z_i and $1 + z_i$ and contract), so the assertion $1 + b + ab = 1$ has a derivation of $(n')^{O(d)}$ lines in degree less than $2d$. The multiplication step is then simulated by

$$A' \vee (1 + a = 1) \;\longrightarrow\; A' \vee (1 + ab = 1) \;\longrightarrow\; A' \vee (1 + b = 1),$$

multiplying the complement a by b and then adding the derived unit assertion, with A' the restricted context. All polynomials in the completed proof have degree at most $2d$, nothing about the original multiplier v was used, and since Π has $S \le T$ lines the completed proof has at most $S(n')^{Cd} \le T(n')^{Cd}$ lines.

The lower bound. If $T < q^d$, Step 1 gives a restriction ρ, and Step 2 a $\mathrm{Res}(\mathrm{PC}_{2d})$ refutation of F' with at most $T(n')^{Cd}$ lines; so if every such refutation has at least L lines then $T \ge L/(n')^{Cd}$. □

6.2 Lower bounds for Res(PC) without a degree restriction

We now apply Theorem 6.3, with $F' = G$, to the line lower bound of Corollary 4.21. Throughout, F is a constant-width CNF on n variables with no PC refutation of degree $D = \lfloor cn \rfloor$, and $\ell = \Theta(\log n)$ is chosen as in Corollary 4.21, so that $G = F \circ \mathrm{Maj}_\ell$ has polynomial size. The minimum in (43) will be attained by the first term: the line lower bound L is $\exp(n^{\Omega(1)})$, while the overheads $(n')^{Cd}$ and q^d are quasipolynomial. The resulting bound is quasipolynomial in the number of variables and superpolynomial in the size of the formula.

Theorem 1.5 (restated)**.** *Let F be a constant-width CNF on n variables with no PC refutation of degree cn, let $G = F \circ \mathrm{Maj}_\ell$ with $\ell = \Theta(\log n)$ chosen as in Corollary 4.21, and let $q = N$. Then $G \circ \mathrm{Sel}_q$ is an unsatisfiable CNF of size polynomial in n, and every* Res(PC) *refutation of it, with no bound on the degree of its polynomials, has total size $T \geq \exp\big(\Omega((\log n)^{5/2})\big)$.*

Proof. Take $d = \lfloor (\log n)^{3/2}/2 \rfloor$, so that $2d$ is within the range of Corollary 4.21, which gives $\log L \geq n^{\delta}$. Here $n' = N = \Theta(n \log n)$, so $Cd \log n' = O((\log n)^{5/2})$ is negligible against $\log L$, and (43) gives $T \geq q^d = \exp(\Omega((\log n)^{3/2} \log n))$. For the size, G has polynomial size, and the selector substitution keeps every clause a single clause and adds $N + N\binom{2N}{2}$ block clauses; the formula has $2N^2 = \Theta(n^2 \log^2 n)$ variables. □

A The one-factor gain for an arbitrary admissible target space

This appendix proves Lemma 3.10: for a suitable affine factor $q = 1 + \sum_i \epsilon_i u_i$,

$$c_{Pq}(K) \leq c_P(K) - \tfrac{1}{2}\, c_P(K - r), \tag{44}$$

so that appending one factor to the approximator removes at least half of the error conditions counted at degree $K - r$. Combined with the degree comparison of Lemma 3.11, which says that the conditions at degree $K - r$ are at least an $M_n(K - r)/M_n(K)$ fraction of all of them, this is the constant-factor decrease $c_{Pq}(K) \leq (1 - \eta) c_P(K)$ that drives Pang's bound.

We prove it in a form that does not mention the formula F or the space C_D. The argument uses only three properties of the target space: it is an $\mathbb{F}_2$-subspace of the multilinear polynomials, it is closed under multiplication within a degree budget, and its remainder map is defined over $\mathbb{F}_2$. The third is automatic for a subspace over $\mathbb{F}_2$, so we state the result for any subspace with the first two, prove it once, and read off Lemma 3.10 as the case $C = C_D$. This generality has two uses. It makes the toy model of Subsection A.1 a special case of the theorem rather than an analogy for it, and it isolates exactly what Appendix B has to re-examine when the target space is $\widehat{C}_D$, which satisfies the closure property only on a subspace.

Definition A.1 (Admissible target space)**.**
Let $V = V_n$ be the space of multilinear polynomials over $\mathbb{F}_2$ in $y_1, \ldots, y_n$, and let $D \in \mathbb{N} \cup \{\infty\}$. An $\mathbb{F}_2$-subspace $C \subseteq V$ is *D-admissible* if

$$u \in C, \qquad \deg u + \deg v \leq D \quad \Longrightarrow \quad uv \in C, \tag{45}$$

with products multilinearized. As in Definition 3.9, fix a basis of C in reduced row echelon form with columns indexed by monomials, complete it to a basis of V by the nonpivot monomials, and let R be the linear map sending a polynomial p to the coefficients of its component along the nonpivot monomials (so $R(p)$ computes the coefficients of the remainder of p mod C.) Then $R(p) = 0$ if and only if $p \in C$.

The two instances we need are the following.

- $C = C_D$, the consequence space of the formula F of Section 3. It is D-admissible by Fact 2.5, and $R = \mathrm{rem}_D$ is the remainder map of Definition 3.9.
- $C = \{0\}$, with $D = \infty$. Admissibility is trivial, since $0 \cdot v = 0$. Here R is injective, and written in the basis of point evaluations it sends a polynomial to its table of values on $\{0,1\}^n$; so $p \in C$ says that p vanishes at every point. This is the toy model of Subsection A.1.

Note that $\{0\}$ is not the consequence space of any CNF, since every C_D contains the clause complements. The toy model is therefore not an instance of Lemma 3.10 itself; it is an instance of Theorem A.2 below.

Now fix a D-admissible C, a line $L = \bigvee_{i=1}^{s}(u_i = 1)$ with $\deg u_i \le r$, and a current approximator $P = \prod_\nu q_\nu$, each q_ν an affine combination $1 + \sum_i a_{\nu i} u_i$. Products are multilinearized, so $u_i^2 = u_i$ and, for every affine combination q of the u_i, $q^2 = q$ and $q(1+q) = 0$. For $0 \le j \le K$ put

$$Z_P(j) = \{f \in V_j : \ fPu_i \in C \text{ for every } i\},$$
$$c_P(j) = \dim V_j - \dim Z_P(j) = \operatorname{rank}(f \mapsto (R(fPu_i))_i, \ f \in V_j).$$

Thus $Z_P(j)$ is the space of multipliers that are *good for* P up to degree j, and $c_P(j)$ the number of independent conditions imposed on them. For $C = C_D$ these are the $Z_P(j)$ and $c_P(j)$ of Definition 3.9.

Theorem A.2 (One-factor gain, general form)**.** *Let C be D-admissible, and suppose*

$$K \ge r, \qquad \deg P + K + 2r \le D. \tag{46}$$

Then there is $\epsilon \in \mathbb{F}_2^s$ such that, with $q = 1 + \sum_i \epsilon_i u_i$ and $P' = Pq$,

$$c_{P'}(K) \ \le \ c_P(K) - \tfrac{1}{2}\, c_P(K - r);$$

equivalently, $\dim Z_{P'}(K) \ge \dim Z_P(K) + \frac{1}{2} c_P(K-r)$.

Lemma 3.10 is the case $C = C_D$, in which (46) is implied by the degree budget (4). For $C = \{0\}$ the second condition in (46) is vacuous. Throughout the appendix, (46) is what permits every multiplication through (45).

A.1 Roadmap, and the proof in the toy model

Everything in this appendix is an algebraic version of a two-line probabilistic argument, which we give first in the case $C = \{0\}$, where the algebra is absent.

The toy model. Take $C = \{0\}$. Then f is good for P exactly when fPu_i is the zero function for every i, i.e. when f vanishes on the error set

$$\operatorname{Err}(P) = \{a \in \{0,1\}^n : \ P(a) = 1 \text{ and } L \text{ is true at } a\} = \bigcup_i \{a : \ P(a)u_i(a) = 1\}$$

of Section 3.1. So $Z_P(K)$ is the space of $f \in V_K$ vanishing on $\operatorname{Err}(P)$, and $c_P(j)$ is the rank of the evaluation map $V_j \to \mathbb{F}_2^{\operatorname{Err}(P)}$: the number of points of $\operatorname{Err}(P)$ that polynomials of degree at most j can "see" independently.

In this case the idea behind Theorem A.2 is as follows. We want to show $Z_{Pq}(K)$ (the space of f's good for Pq) is bigger than $Z_P(K)$, so we need to exhibit elements f of V_K that are good for Pq and new, i.e. not already good for P. Whatever factor $q = 1 + \sum_i \epsilon_i u_i$ we end up choosing, every f of the form $(1+q)v$ with $\deg v \le K - r$ is good for Pq: it lies in V_K since $\deg(1+q) \le r$, and $q(1+q) = 0$ makes $(1+q)vPqu_i$ vanish identically. The question is which of these are new. Say that a point a in the error set of P is *corrected* by q if a is no longer in the error set of Pq, i.e.

$q(a) = 0$. Now $(1+q)v$ is new exactly when it is not in $Z_P(K)$, i.e. when it is nonzero at some point of $\mathrm{Err}(P)$. Split $\mathrm{Err}(P)$ into the uncorrected points, where $q = 1$, and the corrected points, where $q = 0$. At uncorrected points $1 + q = 0$, so $(1+q)v$ vanishes there whatever v is; at corrected points $1 + q = 1$, so $(1+q)v = v$. Hence $(1+q)v$ is new exactly when v is nonzero at some corrected point. So we will choose, in step (i), error points a_k together with polynomials $v_k \in V_{K-r}$ with $v_k(a_k) = 1$ and $v_k(a_{k'}) = 0$ for $k' \neq k$; then for each corrected a_k the polynomial $(1+q)v_k$ is 1 at a_k and 0 at the other chosen points, making the new solutions independent. The degree restriction $\deg v_k \leq K - r$ is what makes the gain be measured at degree $K - r$ rather than K.

In this case Theorem A.2 is proved as follows.

(i) Choose $c = c_P(K - r)$ points $a_1, \ldots, a_c \in \mathrm{Err}(P)$ on which evaluation of V_{K-r} is independent: for each j there is $v_j \in V_{K-r}$ with $v_j(a_j) = 1$ and $v_j(a_{j'}) = 0$ for $j' \neq j$. (By duality, this is equivalent to saying that there are c linearly independent rows in "error" matrix where each row of the matrix corresponds to an error point a_i, each column corresponds to a degree $\leq K - r$ monomial, so the row corresponding to a_i is the vector of all monomial values on input a_i.)

(ii) Choose $\epsilon \in \mathbb{F}_2^s$ uniformly at random and put $q = 1 + \sum_i \epsilon_i u_i$. At every point $a \in \mathrm{Err}(P)$ the line L is true, so the vector $(u_i(a))_i$ is nonzero, and the bit $1 + q(a) = \sum_i \epsilon_i u_i(a)$ is uniformly distributed. The new approximator Pq has error set $\mathrm{Err}(P) \cap \{q = 1\}$: each error point is corrected with probability exactly $1/2$. (This is the only place where "L is true on $\mathrm{Err}(P)$" is used; it is what makes every error point correctable.) The events for different points are not independent, and need not be: by linearity of expectation, $c/2$ of the chosen points a_j are corrected on average, so some ϵ corrects at least $c/2$ of them.

(iii) For each corrected a_j, the polynomial $f_j = (1+q)v_j \in V_K$ is good for Pq, since it vanishes wherever $q = 1$. By the choice of the v_j, f_j is 1 at a_j and 0 at the other chosen points; so the f_j for different corrected j are not in $Z_P(K)$ and are independent modulo it. So the solution space grows (since $Z_P(K) \subseteq Z_{Pq}(K)$) by at least the number of corrected chosen points, which is at least $c/2$.

That is the whole intuition: *a random affine factor corrects half of the error points, and each corrected point that a degree-$(K - r)$ polynomial can see independently buys one new independent multiplier.*

What breaks, and the roadmap. For a general admissible C the conditions $fPu_i \in C$ are still linear conditions on the coefficients of f, but the coordinates of R are the nonpivot monomials of an echelon basis of C, not points. There is no longer a set of "error points" to be corrected one at a time, and no meaning to "a random half" of them. Pang's proof reconstructs each step of the toy argument using only admissibility:

- **Step (iii) survives verbatim** (Subsection A.2). The candidates $(1+q)v$ are good for Pq because $q(1+q) = 0$, and $(1+q)v$ was already good for P exactly when $P(1+q)v \in C$. So the gain is the rank of one linear map, $v \mapsto R(P(1+q)v)$ on V_{K-r} (Lemma A.4).
- **Step (i) becomes linear algebra** (Subsection A.3). The conditions coming from equation i are recorded in a binary matrix M_i, and the map just mentioned is the *binary combination* $\sum_i \epsilon_i M_i$. In the toy model the rows of the M_i are evaluations at points, and a point a contributes its row to exactly those M_i with $u_i(a) = 1$; a combination then either keeps or corrects each row. In general the rows of different M_i can interact, and a combination can

lose rank for a reason that has nothing to do with correcting an error point.

- **Step (ii) is replaced by Pang's matrix lemma** (Theorem A.5), proved in two halves. First, with the coefficients ϵ_i replaced by formal variables λ_i, the combination loses *no* rank (Subsection A.4). This is where the structure of our matrices enters: the idempotence $u_i^2 = u_i$, together with one multiplication permitted by admissibility, gives a "kernel condition" (Lemma A.6) which plays the role of "rows are only rescaled, never mixed" from the toy model, and a Frobenius-descent argument (Lemma A.7) turns it into the statement that the formal kernel is spanned by binary vectors. Second, specializing the formal variables to random bits loses at most half the rank (Subsection A.5); this is the method of multiplicities of Dvir–Kopparty–Saraf–Sudan [DKSS13], in the form used by Guruswami–Kopparty [GK16] and Forbes–Guruswami [FG15], and it is exactly the algebraic counterpart of "each point is corrected with probability 1/2".

Subsection A.6 assembles the pieces. Pang's own presentation [Pan26, Section 3] credits the same sources.

A.2 Step (iii): corrected points give new solutions

Let $q = 1 + \sum_i \epsilon_i u_i$ and $P' = Pq$. The first lemma says that nothing is lost and that every $(1+q)v$ is gained; the second says exactly which of these gains are genuinely new.

Lemma A.3 (Old solutions persist, candidates appear)**.** *Under* (46),

$$Z_{P'}(K) \;\supseteq\; Z_P(K) \;+\; (1+q)\,V_{K-r}.$$

Proof. If $fPu_i \in C$ then $fP'u_i = (fPu_i)q \in C$ by (45), since $\deg(fPu_i) + \deg q \le K + \deg P + 2r \le D$. If $f = (1+q)v$ with $v \in V_{K-r}$, then $f \in V_K$ and $fP'u_i = Pu_i v \cdot q(1+q) = 0 \in C$. □

In the toy model, $(1+q)v$ was already good for P exactly when v vanishes on the corrected error points, i.e. when $P(1+q)v = 0$. The statement for general C is the same with C in place of $\{0\}$, and it identifies the gain as a rank.

Lemma A.4 (The gain is a rank)**.** *Under* (46)*, for* $v \in V_{K-r}$*,*

$$(1+q)v \in Z_P(K) \quad\Longleftrightarrow\quad R(P(1+q)v) = 0,$$

and consequently

$$\dim Z_{P'}(K) - \dim Z_P(K) \;\ge\; \operatorname{rank}\big(v \mapsto R(P(1+q)v),\; v \in V_{K-r}\big). \tag{47}$$

Proof. If $Pu_i(1+q)v \in C$ for all i, then taking the combination $\sum_i \epsilon_i$ gives $P(1+q)^2 v = P(1+q)v \in C$. Conversely, if $P(1+q)v \in C$, multiplying by each u_i (allowed: $\deg P + K + r \le D$) gives $Pu_i(1+q)v \in C$. This proves the equivalence.

For (47), let $\Phi(v) = R(P(1+q)v)$ on V_{K-r}. The map $v \mapsto (1+q)v$ sends $\ker\Phi$ onto $(1+q)V_{K-r} \cap Z_P(K)$ by the equivalence, and its own kernel is contained in $\ker\Phi$; hence $\dim\big((1+q)V_{K-r}\big) - \dim\big((1+q)V_{K-r} \cap Z_P(K)\big) = \dim V_{K-r} - \dim\ker\Phi = \operatorname{rank}\Phi$, and this is the dimension that $(1+q)V_{K-r}$ adds to $Z_P(K)$. Apply Lemma A.3. □

A.3 Step (i): the conditions as a family of matrices

For each equation $u_i = 1$ of L define the binary matrix

$$M_i : V_{K-r} \to \text{remainder coordinates}, \qquad M_i v = R(P u_i v).$$

Its kernel is the set of v that satisfy the error condition of equation i for P; the common kernel of all M_i is $Z_P(K-r)$, so the stacked map $v \mapsto (M_1 v, \dots, M_s v)$ has rank

$$c := c_P(K-r),$$

the toy model's "number of error points that degree $K-r$ can see". Since $R(P(1+q)v) = \sum_i \epsilon_i R(P u_i v)$, the map whose rank is the gain in (47) is the binary combination

$$M(\epsilon) = \sum_i \epsilon_i M_i.$$

Lemma A.4 has thus reduced (44) to a statement about s matrices:

> *find* $\epsilon \in \mathbb{F}_2^s$ *with* $\operatorname{rank} M(\epsilon) \geq c/2$.

Why this is not automatic, and what makes it true here. In the toy model the rows of every M_i are evaluations at points of $\mathrm{Err}(P)$, and the row of a point a appears in M_i exactly when $u_i(a) = 1$. A combination $\sum_i \epsilon_i M_i$ therefore has, at row a, the row ev_a multiplied by the scalar $\sum_i \epsilon_i u_i(a)$: rows are kept or dropped, never mixed with one another, and a random ϵ keeps each with probability $1/2$. For general matrices nothing of the sort holds. For $M_1 = (1\ 0)$ and $M_2 = (0\ 1)$ the stacked rank is 2 while every combination has rank at most 1; the two rows interfere.

The property of our matrices that excludes such interference is inherited from the idempotence $u_i^2 = u_i$, and it is cleanest to state with the coefficients left formal. Let $E = \mathbb{F}_2(\lambda_1, \dots, \lambda_s)$ be the field of rational functions in formal variables λ_i, and write

$$M(\lambda) = \sum_i \lambda_i M_i, \qquad M(\lambda^2) = \sum_i \lambda_i^2 M_i,$$

both regarded as matrices over E.

Theorem A.5 (Pang's matrix lemma [Pan26, Lemma 3.3])**.** *Let $M_1, \dots, M_s$ be matrices over $\mathbb{F}_2$ of the same shape, let c be the rank of the stacked map $v \mapsto (M_1 v, \dots, M_s v)$, and suppose*

$$\ker_E M(\lambda) \subseteq \ker_E M(\lambda^2). \tag{48}$$

Then $\operatorname{rank}_E M(\lambda) = c$, *and* $\mathbb{E}_{\epsilon \in \mathbb{F}_2^s} \operatorname{rank}_{\mathbb{F}_2} M(\epsilon) \geq c/2$. *In particular some* $\epsilon \in \mathbb{F}_2^s$ *has* $\operatorname{rank} M(\epsilon) \geq c/2$.

In the toy model ($C = \{0\}$) the kernel condition is obvious: $M(\lambda)$ is the evaluation matrix with row a scaled by the linear form $\ell_a(\lambda) = \sum_{i : u_i(a) = 1} \lambda_i \neq 0$, and $M(\lambda^2)$ is the same with $\ell_a(\lambda^2)$, so both have the same kernel, namely the polynomials vanishing on $\mathrm{Err}(P)$. The theorem says that this condition alone, with no points in sight, suffices for both conclusions. We prove the two conclusions in Subsections A.4 and A.5; first we verify the hypothesis for our matrices.

A.3.1 The kernel condition, from admissibility

Write $C(E)$ for the E-span of C inside the polynomials in y with coefficients in E. Because R is a binary matrix, its kernel over E is the E-span of its kernel over $\mathbb{F}_2$:

$$R(u) = 0 \quad \Longleftrightarrow \quad u \in C(E).$$

Admissibility (45) also holds over E: if $u \in C(E)$ has y-degree at most j, then u is an E-combination of elements of $C \cap V_j$ (the binary linear system "$R(u) = 0$ and all coefficients of degree $> j$ vanish" has binary kernel $C \cap V_j$, and its kernel over E is the E-span of that), so for a of y-degree at most $D - j$, expanding a into monomials and applying (45) to each element of $C \cap V_j$ gives $ua \in C(E)$. Throughout, degree means degree in y; the λ_i are scalars.

Lemma A.6 (The kernel condition holds)**.** *Under* (46)*, the matrices M_i of Subsection A.3 satisfy* (48)*.*

Proof. Put $q_\lambda = 1 + \sum_i \lambda_i u_i$. For $v \in E^{V_{K-r}}$ (a polynomial of degree at most $K - r$ with coefficients in E), $M(\lambda)v = R(P(1 + q_\lambda)v)$, so $M(\lambda)v = 0$ means $P(1 + q_\lambda)v \in C(E)$. Multiply once more by $1+q_\lambda$, which has y-degree at most r; this is allowed since $\deg P+(K-r)+r+r \le D$. Multilinearizing in y,

$$(1 + q_\lambda)^2 = \Big(\sum_i \lambda_i u_i\Big)^2 = \sum_i \lambda_i^2 u_i^2 = \sum_i \lambda_i^2 u_i,$$

because cross terms cancel in characteristic two and $u_i^2 = u_i$. Thus $P(\sum_i \lambda_i^2 u_i)v \in C(E)$, i.e. $M(\lambda^2)v = 0$. □

The identity $(1 + q_\lambda)^2 = \sum_i \lambda_i^2 u_i$ is the whole mechanism: squaring an affine combination of idempotents is the same as applying the Frobenius map $\lambda \mapsto \lambda^2$ to its coefficients. (We stress that $(1 + q_\lambda)^2 \neq 1 + q_\lambda$ as a polynomial with formal coefficients; the idempotence is in y only.) In the toy model this is literally the statement that row a is scaled by $\ell_a(\lambda)$ in one matrix and by $\ell_a(\lambda)^2 = \ell_a(\lambda^2)$ in the other.

A.4 Step (ii), first half: no rank is lost formally

We prove the first assertion of Theorem A.5: with formal coefficients the combination $M(\lambda)$ has the full stacked rank c. What could go wrong is that over E the unknown vector v may have coefficients depending on λ, and such coefficients can produce spurious solutions: $\lambda_1 a + \lambda_2 b = 0$ has the solution $(a, b) = (\lambda_2, \lambda_1)$ although its only binary solution is $(0, 0)$. The kernel condition rules this out: it forces the kernel of $M(\lambda)$ to be spanned by *binary* vectors, and a binary vector killed by $\sum_i \lambda_i M_i$ is killed by every M_i separately, because the λ_i are independent formal variables. The tool is the following descent lemma, in which $v \mapsto v^{[2]}$ denotes coordinatewise squaring of a vector over E; it says that a subspace stable under the Frobenius map is "defined over $\mathbb{F}_2$".

Lemma A.7 (Frobenius-stable subspaces are defined over $\mathbb{F}_2$)**.** *Let $W \subseteq E^m$ be an E-subspace with $v^{[2]} \in W$ for every $v \in W$. Then W is the E-span of $W \cap \mathbb{F}_2^m$.*

Proof. Bring a basis of W into reduced echelon form: there are pivot coordinates $p_1 < \cdots < p_d$ and basis vectors $w^{(1)}, \ldots, w^{(d)}$ with $w^{(j)}_{p_k} = \delta_{jk}$. Each $w^{(j)}$ is the unique element of W with these pivot values. Now $(w^{(j)})^{[2]} \in W$ has the same pivot values, since $0^2 = 0$ and $1^2 = 1$, so $(w^{(j)})^{[2]} = w^{(j)}$. Every coordinate a of $w^{(j)}$ therefore satisfies $a^2 = a$, i.e. $a(a + 1) = 0$ in the field E, so $a \in \{0, 1\}$. Thus the $w^{(j)}$ are binary and span W. □

Proof of Theorem A.5, first assertion. Let $W = \ker_E M(\lambda)$ and let $W_0 = \bigcap_i \ker_{\mathbb{F}_2} M_i$ be the common binary kernel, of dimension $m - c$. Clearly $W \supseteq (W_0)_E$, the E-span; we show equality, in three steps.

Step 1: $W = \ker_E M(\lambda^2)$. The inclusion $\subseteq$ is the hypothesis (48). For equality it suffices that the two matrices have the same rank over E. Every minor of $M(\lambda)$ is a polynomial $\Delta(\lambda)$ with coefficients in $\mathbb{F}_2$, and the corresponding minor of $M(\lambda^2)$ is $\Delta(\lambda^2) = \Delta(\lambda)^2$, since squaring is additive in characteristic two and fixes $\mathbb{F}_2$. A minor is nonzero exactly when its square is, so the ranks agree.

Step 2: W is Frobenius-stable. For $v \in W$, squaring every coordinate of the identity $\sum_i \lambda_i M_i v = 0$ and using that the entries of the M_i are in $\mathbb{F}_2$ gives $\sum_i \lambda_i^2 M_i v^{[2]} = 0$, i.e. $v^{[2]} \in \ker_E M(\lambda^2) = W$.

Step 3: descent. By Lemma A.7, W is spanned by binary vectors $v \in W \cap \mathbb{F}_2^m$. For such v the vectors $M_i v$ are binary, and $\sum_i \lambda_i (M_i v) = 0$ with $\lambda_1, \dots, \lambda_s$ linearly independent over $\mathbb{F}_2$ forces $M_i v = 0$ for every i. Hence $W \cap \mathbb{F}_2^m = W_0$ and $W = (W_0)_E$.

Therefore $\dim_E W = m - c$ and $\operatorname{rank}_E M(\lambda) = c$. □

A.5 Step (ii), second half: random bits keep half the rank

It remains to pass from the formal rank c of $M(\lambda)$ to a binary point ϵ with $\operatorname{rank} M(\epsilon) \geq c/2$. This is the algebraic form of "each point is corrected with probability $1/2$". It knows nothing about the target space C; it is the argument by which Guruswami and Kopparty [GK16] control the rank of a matrix of polynomials at the points of a grid, further developed by Forbes and Guruswami [FG15, Section 6]. The idea: a $c \times c$ minor $\Delta(\lambda)$ of $M(\lambda)$ is a nonzero polynomial of degree c; wherever the rank drops by d, that polynomial vanishes to order d; and a nonzero polynomial of degree c cannot vanish to high order at many points of the cube $\{0,1\}^s$, the total order being at most $c \cdot 2^{s-1}$, i.e. $c/2$ per point on average.

Recall that the *multiplicity* $\operatorname{mult}(\Delta, \epsilon)$ of a polynomial Δ at a point ϵ is the least degree of a monomial appearing in $\Delta(\epsilon + \mu)$ as a polynomial in the shift μ (and ∞ if $\Delta = 0$).

Lemma A.8 (Rank drop forces multiplicity [GK16, Claims 15–16])**.** *Let $M(\lambda) = \sum_i \lambda_i M_i$ with M_i over a field $\mathbb{K}$, let $\Delta(\lambda)$ be the determinant of a $c \times c$ submatrix of $M(\lambda)$, and let $\epsilon \in \mathbb{K}^s$. If $\operatorname{rank} M(\epsilon) \leq c - d$, then $\operatorname{mult}(\Delta, \epsilon) \geq d$.*

Proof. The submatrix $N(\lambda)$ of $M(\lambda)$ is itself of the form $\sum_i \lambda_i N_i$, and $\operatorname{rank} N(\epsilon) \leq c - d$. Choose invertible matrices A, B over $\mathbb{K}$ with $AN(\epsilon)B = \operatorname{diag}(I_{c-d}, 0)$. Since $N(\epsilon + \mu) = N(\epsilon) + N(\mu)$,

$$\Delta(\epsilon + \mu) = \det(A)^{-1} \det(B)^{-1} \det\big(\operatorname{diag}(I_{c-d}, 0) + AN(\mu)B\big),$$

where $AN(\mu)B$ has entries that are linear forms in μ with no constant term. The last d rows of the matrix inside the determinant consist entirely of such linear forms, so in the Leibniz expansion every term contains at least d factors of degree one in μ. Hence every monomial of $\Delta(\epsilon + \mu)$ has degree at least d. □

Theorem A.9 (Multiplicity Schwartz–Zippel [DKSS13, Lemma 8])**.** *Let $\Delta \neq 0$ be a polynomial of degree at most c in s variables over a field $\mathbb{K}$, and let $S \subseteq \mathbb{K}$ be finite. Then*

$$\sum_{\epsilon \in S^s} \operatorname{mult}(\Delta, \epsilon) \leq c\,|S|^{s-1}.$$

Lemma A.10 (Rank retention under specialization)**.** *Let $M(\lambda) = \sum_i \lambda_i M_i$ with M_i over a field $\mathbb{K}$, let $S \subseteq \mathbb{K}$ be finite, and let $c = \operatorname{rank}_{\mathbb{K}(\lambda)} M(\lambda)$. Then*

$$\mathbb{E}_{\epsilon \in S^s} \operatorname{rank} M(\epsilon) \;\geq\; c\Big(1 - \frac{1}{|S|}\Big).$$

In particular, for $\mathbb{K} = S = \mathbb{F}_2$, the expected rank is at least $c/2$.

Proof. Since $M(\lambda)$ has rank c over $\mathbb{K}(\lambda)$, some $c \times c$ minor $\Delta(\lambda)$ is a nonzero polynomial; its entries are linear forms in λ, so $\deg \Delta \leq c$. By Lemma A.8, at every ϵ the rank loss $c - \operatorname{rank} M(\epsilon)$ is at most $\operatorname{mult}(\Delta, \epsilon)$. Summing over S^s and applying Theorem A.9,

$$\sum_{\epsilon \in S^s} \big(c - \operatorname{rank} M(\epsilon)\big) \leq \sum_{\epsilon \in S^s} \operatorname{mult}(\Delta, \epsilon) \leq c\,|S|^{s-1},$$

and dividing by $|S|^s$ gives the claim. □

This completes Theorem A.5: its first assertion supplies the formal rank c, and Lemma A.10 with $\mathbb{K} = S = \mathbb{F}_2$ supplies the binary point. The factor $1/2$ is the price of working over the two-element field, and it is tight: for $M_1 = M_2 = M \neq 0$ the hypothesis holds, the stacked rank is $\operatorname{rank} M$, and $M(\epsilon)$ is M or 0 with equal probability. This matches the toy model, where a single error point seen by two equations is corrected exactly when $\epsilon_1 + \epsilon_2 = 1$.

A.6 Conclusion

Proof of Theorem A.2. Let M_i be the matrices of Subsection A.3, with stacked rank $c = c_P(K - r)$. By Lemma A.6 they satisfy the hypothesis of Theorem A.5, which provides $\epsilon \in \mathbb{F}_2^s$ with $\operatorname{rank} M(\epsilon) \geq c/2$. Put $q = 1 + \sum_i \epsilon_i u_i$. Then $M(\epsilon)$ is the map $v \mapsto R(P(1+q)v)$ on V_{K-r}, so Lemma A.4 gives $\dim Z_{Pq}(K) - \dim Z_P(K) \geq c_P(K - r)/2$. □

Proof of Lemma 3.10. Apply Theorem A.2 with $C = C_D$, which is D-admissible by Fact 2.5. For an intermediate approximator with at most $t - 1$ factors, $\deg P \leq (t-1)r$, so the degree budget (4) gives $\deg P + K + 2r \leq K + rt + r \leq D$, which is (46). □

Remark (Why the decrease is by a constant factor). The gain (44) is measured at degree $K - r$ while the conditions are counted at degree K. Lemma 3.11 bridges the two: the monomials that carry a condition (those that are not leading monomials of solutions) are closed under taking divisors, and such a down-set has nonincreasing density in each degree, so the degree-$(K - r)$ conditions are at least their proportional share $M_n(K - r)/M_n(K)$ of the total. Hence each factor removes at least the fraction $\eta = \frac{1}{2} M_n(K - r)/M_n(K)$ of all conditions. For Pang's choice $K \approx D/4 \leq n/4$ the ratio $M_n(K - r)/M_n(K)$ is about $(K/n)^r$, which is a constant for constant r and is the source of the $\log n$ ceiling on r; Section 4 is devoted to moving K close to half the number of variables, where the ratio becomes $e^{-O(r/\sqrt{\ell})}$.

Remark (What the argument used). Beyond the admissibility of C, which is used for the two multiplications in Lemmas A.3 and A.4 and for the one in Lemma A.6, the argument used only that the polynomials u_i are idempotent, $u_i^2 = u_i$, which gives $(1 + q_\lambda)^2 = \sum_i \lambda_i^2 u_i$, and that the matrices M_i are binary, which makes the Frobenius argument available. The choice of the field $\mathbb{F}_2$ appears only in the constant $1/2$ of Lemma A.10. The lifted target space $\widehat{C}_D$ of Section 4 is not admissible in the sense of Definition A.1 for either of its two degrees, so Theorem A.2 does not apply to it directly; Appendix B explains why and repeats the argument with the extra bookkeeping this requires.

B Rank estimates with the coefficient-degree restriction

We prove Lemmas 4.16 and 4.15, and then Proposition 4.17. Lemma 4.16 is the lifted version of the degree comparison, Lemma 3.11, and Lemma 4.15 is the lifted version of the one-factor gain, Lemma 3.10; we prove each by repeating the corresponding argument (that of Lemma 3.11, respectively that of Appendix A) over $\widehat{C}_D$ instead of C_D, with one piece of bookkeeping added. Each of those arguments multiplies a known member of the consequence space by a polynomial of small degree; over $\widehat{C}_D$, Lemma 4.10 justifies such a multiplication only when the known member has coefficient degree at most Q. So before every multiplication we restrict the relevant space of multipliers to W_Q, which by (22) costs at most E_j dimensions at degree j, and we carry these losses through the two estimates. Nothing else changes.

Why $\widehat{C}_D$ is not covered by Theorem A.2. Appendix A proves the one-factor gain for any *admissible* target space, i.e. any subspace C closed under multiplication whenever $\deg u + \deg v \le D$ for a single degree function deg (Definition A.1). The space $\widehat{C}_D$ fails this for both of its degrees. With $\deg = \deg_x$ the hypothesis $\deg_x u + \deg_x v \le D$ is never met in the regime of Section 4, where the multipliers have ordinary degree near $N/2 \gg D$. With $\deg = \beta$ the closure rule we actually have, Lemma 4.10, is the mixed statement $p \in \widehat{C}_D,\ \beta(p) + \deg_x v \le D \Rightarrow pv \in \widehat{C}_D$, and the corresponding statement with $\beta(v)$ in place of $\deg_x v$ is false because β is not submultiplicative (the example after Lemma 4.10). So every multiplication in the argument is permitted only when the known member of $\widehat{C}_D$ has coefficient degree at most Q, which is a condition on the multiplier f or v being multiplied, not on $\widehat{C}_D$. This is exactly the restriction to W_Q described above, and the two lemmas below are Theorem A.2 and Lemma 3.11 with that restriction carried through.

For $0 \le j \le K$, let E_j count the monomials through degree j excluded by W_Q. Let $T_{P,j}$ be the corresponding columns of the error matrix in Definition 4.14, and write

$$Z_P(j) = \{f \in V_j(x) : fPu_i \in \widehat{C}_D \text{ for every } i\}, \qquad c_P(j) = \operatorname{rank} T_{P,j}.$$

Thus $Z_P(j) = Z_P(K) \cap V_j(x)$ and $E_j \le \varepsilon M_N(j)$, with ε as in Definition 4.13.

B.1 Comparing multiplier degrees

Proof of Lemma 4.16. Recall the argument of Lemma 3.11. With monomials ordered by degree and then lexicographically, $c_P(j)$ counts the monomials of degree at most j that are not leading monomials of elements of $Z_P(K)$; call these the *counted* monomials, and let b_j be the number of them of degree exactly j, so that $c_P(j) = \sum_{a \le j} b_a$. The counted monomials are closed under taking divisors, because a solution whose leading monomial is m can be multiplied by a variable absent from m to give a solution with leading monomial that extension. This closure forces the density $b_j / \binom{N}{j}$ to be nonincreasing in j.

Over $\widehat{C}_D$ the multiplication by a variable is justified by Lemma 4.10 only for solutions of coefficient degree at most Q (the hypothesis $Q + \deg P + r + 1 \le D$ is exactly what it needs). We therefore first pass from $Z_P(j)$ to $Z_P(j) \cap W_Q$, which by (22) loses at most E_j leading monomials; every leading monomial of the smaller space has a witness in W_Q, and that witness can be multiplied by a variable as before. Consequently each degree-j divisor of a counted degree-$(j+1)$ monomial is either counted or one of the at most E_j lost leading monomials. Counting the pairs (counted

degree-$(j+1)$ monomial, degree-j divisor) in two ways, as in Lemma 3.11, gives

$$(j+1)b_{j+1} \le (N-j)(b_j + E_j), \qquad \text{i.e.} \qquad \frac{b_{j+1}}{\binom{N}{j+1}} \le \frac{b_j}{\binom{N}{j}} + \varepsilon(j+1),$$

where we used $E_j \le \varepsilon M_N(j) \le \varepsilon(j+1)\binom{N}{j}$, valid since $j \le K < N/2$. So the density is no longer nonincreasing, but it can increase by at most $\varepsilon(j+1)$ per step, hence by at most $\varepsilon \sum_{j<K}(j+1) \le N^2\varepsilon$ in total between any two degrees up to K.

Write $\delta_j = b_j/\binom{N}{j}$. For every $j > K-r$ and every $i \le K-r$ we have $\delta_j \le \delta_i + N^2\varepsilon$, hence δ_j is at most the minimum of $\delta_0, \dots, \delta_{K-r}$ plus $N^2\varepsilon$, and in particular at most their binomially weighted average $c_P(K-r)/M_N(K-r)$ plus $N^2\varepsilon$. The weighted average $c_P(K)/M_N(K)$ of $\delta_0, \dots, \delta_K$ is therefore at most $c_P(K-r)/M_N(K-r) + N^2\varepsilon$. Multiplying by $M_N(K-r)$ and rearranging gives (25). □

B.2 The gain from one factor

We use Pang's matrix lemma in the form of Theorem A.5: if binary matrices $A_1, \dots, A_s$ satisfy the kernel condition $\ker_E A(\lambda) \subseteq \ker_E A(\lambda^2)$ over $E = \mathbb{F}_2(\lambda_1, \dots, \lambda_s)$, where $A(\lambda) = \sum_i \lambda_i A_i$ and $A(\lambda^2) = \sum_i \lambda_i^2 A_i$, then some binary combination $\sum_i a_i A_i$ has rank at least half the rank of the stacked map $v \mapsto (A_1 v, \dots, A_s v)$. As in Appendix A, the kernel condition is verified by one further multiplication, now in $\widehat{C}_D$; the only new point is that the multiplication must be justified in majority coordinates.

Allowing formal scalar coefficients. For an $\mathbb{F}_2$-subspace V, let V_E be the space of E-linear combinations of its elements. As shown in Subsubsection A.3.1, the bounded-degree multiplication rule for any admissible space, in particular for C_D, remains valid over E. Applying that argument to each majority coefficient polynomial shows that Lemma 4.10 also holds over E, with the same bounds. All remainder maps extend E-linearly.

Proof of Lemma 4.15. Write $P' = Pq$. We follow Appendix A, with the space of candidate multipliers v restricted to $X = V_{K-r}(x) \cap W_Q$ so that every multiplication is permitted. Using a fixed basis of X, define the binary matrices

$$A_i(v) = T_{P,K}(u_i v) \qquad (v \in X),$$

so that $A_i(v) = 0$ exactly when $u_i v$ satisfies all error conditions for P. Their common kernel is $X \cap Z_P(K-r)$: the common kernel requires $vPu_\alpha u_i \in \widehat{C}_D$ for every α, i; these conditions follow from $vPu_\alpha \in \widehat{C}_D$ by permitted multiplication, and conversely taking $i = \alpha$ recovers $vPu_\alpha \in \widehat{C}_D$ because $u_\alpha^2 = u_\alpha$. The stacked map $v \mapsto (A_1 v, \dots, A_s v)$ on X therefore has rank $\dim X - \dim(X \cap Z_P(K-r)) \ge \dim X - \dim Z_P(K-r) = \big(M_N(K-r) - E_{K-r}\big) - \dim Z_P(K-r) = c_P(K-r) - E_{K-r}$.

To check the kernel condition, put $q_\lambda = 1 + \sum_i \lambda_i u_i$. For $v \in X_E$, the equation $A(\lambda)v = 0$ says that $vPu_\alpha(1+q_\lambda)$ belongs to $(\widehat{C}_D)_E$ for every α. Multiply once more by $1+q_\lambda$. This is allowed by $\beta(v) + \deg P + 3r \le Q + \deg P + 3r \le D$. Since $(1+q_\lambda)^2 = \sum_i \lambda_i^2 u_i$, the resulting conditions are exactly $A(\lambda^2)v = 0$.

Theorem A.5 now supplies $q = 1 + \sum_i a_i u_i$ such that $v \mapsto T_{P,K}((1+q)v)$, $v \in X$, has rank at least $\frac{1}{2}(c_P(K-r) - E_{K-r})$. As in Lemma A.4, this rank is the number of dimensions that $(1+q)X$ adds

to $Z_P(K)$, and a fortiori the number it adds to the smaller space $Z_P(K) \cap W_Q$, whose dimension is at least $\dim Z_P(K) - E_K$ by (22).

Both $Z_P(K) \cap W_Q$ and $(1+q)X$ lie in $Z_{P'}(K)$: for the first, multiply $fPu_i \in \widehat{C}_D$ by q, permitted since $\beta(f) \le Q$; for the second, $(1+q)vP'u_i = Pu_iv \cdot q(1+q) = 0$. Hence

$$\dim Z_{P'}(K) \ge (\dim Z_P(K) - E_K) + \tfrac{1}{2}(c_P(K-r) - E_{K-r}),$$

and subtracting from $M_N(K)$ gives (24). □

B.3 Iteration

Proof of Proposition 4.17. Before the last factor is appended, P is a product of at most $t-1$ factors, so $\deg P \le (t-1)r$. Thus $Q + (t+2)r \le D$ implies the degree conditions of Lemmas 4.16 and 4.15. Combining (25) and (24) with $E_j \le \varepsilon M_N(j)$ gives the one-factor inequality (26) for $P' = Pq$. Initially, $B \cap W_Q \subseteq Z_1(K)$, so $c_1(K) \le h + E_K \le h + \varepsilon M_N(K)$. After t factors,

$$\begin{aligned} c_{P_L}(K) &\le (1-\eta)^t\big(h + \varepsilon M_N(K)\big) + 2N^2 \varepsilon M_N(K) \sum_{\nu=0}^{t-1} (1-\eta)^\nu \\ &\le e^{-\eta t}\big(h + \varepsilon M_N(K)\big) + \frac{2N^2 \varepsilon M_N(K)}{\eta}. \end{aligned}$$

This is (27). Corollary 4.18 then imposes the final W_Q restrictions and chooses the common multiplier outside $\widehat{C}_D$. □

C Counting estimates

We prove the two estimates quoted in Subsection 4.4: the bound (33) on the fraction of monomials excluded by W_Q, which rests on a tail bound for the number of selected bits in a union of blocks, and the lower bound (35) on the ratio $M_N(K-r)/M_N(K)$ near $K \approx N/2$, which rests on the flatness of the binomial coefficients there.

A tail bound for sampling without replacement. Choose a uniform j-subset of an N-element set. For a fixed set of $L \ge 1$ coordinates, let X count how many are selected, and put $p = j/N$. For every $u \ge 0$,

$$\Pr[X - Lp \ge u] \le \exp(-2u^2/L). \tag{49}$$

Indeed, if $X_1, \dots, X_L$ are the selection indicators, then for any set A of s of these coordinates,

$$\mathbb{E} \prod_{i \in A} X_i = \frac{j(j-1)\cdots(j-s+1)}{N(N-1)\cdots(N-s+1)} \le p^s,$$

where the expectation is zero for $s > j$. For $\lambda \ge 0$, expand the product to obtain

$$\mathbb{E}e^{\lambda X} = \mathbb{E} \prod_{i=1}^{L} (1 + (e^\lambda - 1)X_i) \le (1 - p + pe^\lambda)^L \le \exp(Lp\lambda + L\lambda^2/8).$$

The last inequality follows because $\log(1 - p + pe^\lambda) - p\lambda$ has value and derivative zero at zero, and second derivative at most $1/4$. Markov's inequality with $\lambda = 4u/L$ now proves (49).

The excluded fraction in each degree. Recall $Q = \lfloor cn/16 \rfloor$, $K = \lfloor N/2 - an\sqrt{\ell} \rfloor$, and $N = n\ell$. Fix $0 \le j \le K$, and choose a monomial x^T of degree j uniformly. Fix any $Q+1$ blocks. Their union has $L = (Q+1)\ell$ coordinates. If T selects more than half the bits in every one of these blocks, the number selected in the union is at least $(Q+1)\ell/2$, whereas its mean is at most

$$(Q+1)\ell\frac{K}{N} \le (Q+1)\ell/2 - (Q+1)a\sqrt{\ell}.$$

Equation (49), with $u = (Q+1)a\sqrt{\ell}$, bounds the probability by $e^{-2(Q+1)a^2}$. A union bound over the sets of $Q+1$ blocks gives, in every degree $j \le K$,

$$\Pr[|J(T)| > Q] \le \binom{n}{Q+1} e^{-2(Q+1)a^2} \le \exp[-(ca^2/8 - \log 2)n],$$

using $Q + 1 > cn/16$ and $\binom{n}{Q+1} \le 2^n$. This proves (33). The main-text bound $h \ge 2^{-n} M_N(K)$ therefore gives

$$\frac{\varepsilon M_N(K)}{h} \le \exp[-(ca^2/8 - 2\log 2)n] \le e^{-12n}$$

when $ca^2/8 \ge 2\log 2 + 12$, proving (34).

Lemma C.1 (Cumulative monomial counts near half the variables)**.** *Let $n \ge 2$, $a \ge 1$, $N = n\ell$, and $\sqrt{\ell} \ge 128(a+2)$. For $K = \lfloor N/2 - an\sqrt{\ell} \rfloor$ and $1 \le r \le n$,*

$$\frac{1}{2}\frac{M_N(K-r)}{M_N(K)} \ge \frac{1}{2}\exp\left(-64(a+2)\frac{r}{\sqrt{\ell}}\right).$$

Proof. The idea is that $M_N(j)$ is dominated by its top $\Theta(\sqrt{\ell})$ binomial coefficients, each of which is at least half the largest one, so that removing the top coefficient changes $M_N(j)$ by a factor $1 - O(1/\sqrt{\ell})$. Put $u = (a+2)/\sqrt{\ell}$ and $L = \lfloor 1/(16u) \rfloor$. For $K - r + 1 \le j \le K$, each of the indices $j, j-1, \dots, j-L$ is at least

$$N/2 - (a+2)n\sqrt{\ell} > 0,$$

because the floor in K, the shift by at most r, and the further shift by at most L together cost at most $1 + r + L \le 2n\sqrt{\ell}$. For $0 \le \nu < L$, it follows that

$$\frac{\binom{N}{j-\nu-1}}{\binom{N}{j-\nu}} = \frac{j-\nu}{N-(j-\nu)+1} \ge \frac{1-2u}{1+3u} \ge 1 - 5u,$$

where $2/N \le u$. Since $u \le 1/128$ and $uL \le 1/16$, $(1-5u)^L \ge e^{-10uL} > 1/2$. Thus every one of those $L+1$ binomial coefficients is at least $\frac{1}{2}\binom{N}{j}$, and

$$M_N(j) \ge \frac{L+1}{2}\binom{N}{j} \ge \frac{1}{32u}\binom{N}{j}, \qquad \frac{M_N(j-1)}{M_N(j)} = 1 - \frac{\binom{N}{j}}{M_N(j)} \ge 1 - 32u.$$

Multiply the last inequality for $j = K - r + 1, \dots, K$. Since $32u \le 1/4$, $\log(1-32u) \ge -64u$, giving the claimed bound. □

D Depth-3.5 $\mathrm{AC}^0[2]$-Frege: the rules and the simulation

This appendix gives the precise rules of the system of Section 5 and proves Proposition 5.2, the preservation of the rules under the two-level approximation. The nested approximation follows Pang's discussion [Pan26, Section 8].

Definition D.1 (Depth-3.5 $\mathrm{AC}^0[2]$-Frege)**.**
Fix $b \geq 1$. Every equation $u = 1$ occurring in a line has u a multilinear polynomial over $\mathbb{F}_2$ in the lifted variables, with $\deg u \leq b$. A term is $\mathcal{T} = \bigwedge_{j \in I_\mathcal{T}} (u_{\mathcal{T},j} = 1)$; a line is a disjunction of terms. A singleton equation is a term, and the empty term is true. Ordering and repetition of equations within a term are immaterial. In addition to weakening and contraction of disjunctions, the rules are

$$\frac{A \vee \mathcal{T}}{A \vee \mathcal{U}}, \qquad \frac{A \vee \mathcal{T} \qquad B \vee \mathcal{U}}{A \vee B \vee (\mathcal{T} \wedge \mathcal{U})},$$

$$\frac{A \vee \mathcal{T} \qquad B \vee \bigvee_{j \in I_\mathcal{T}} (1 + u_{\mathcal{T},j} = 1)}{A \vee B}.$$

Here A, B are disjunctions of terms. The first rule requires that every equation of $\mathcal{U}$ occurs in $\mathcal{T}$. The polynomial rules are those of Definition 2.6, with these disjunctive contexts. Multiplication may conclude $1 + \mathrm{ML}((1+u)v) = 1$ from $u = 1$ whenever the polynomial $1 + \mathrm{ML}((1+u)v)$ has degree at most b; there is no separate degree bound on the multiplier v. A singleton false constant equation may be deleted. The default axioms are $(1 = 1)$, the empty true term, and $(u = 1) \vee (1 + u = 1)$ for every allowed polynomial u. Input clauses are disjunctions of singleton terms. The total size T counts all term occurrences as well as all monomial occurrences, so it bounds both the number of terms and the number of proof lines.

Equivalence with the system of Buss, Kołodziejczyk and Zdanowski. The normal-form theorem of [BKZ15] is stated for the Tait-style system $\mathrm{PCK}^1_{\mathbb{F}_2}$ [BKZ15, Definitions 11–13], which we recall. A line is a *cedent*, a finite set of formulas read as their disjunction. The formulas are multilinear polynomials f over $\mathbb{F}_2$, read as the assertions $f = 0$, together with conjunctions $\bigwedge \Phi$ and disjunctions $\bigvee \Phi$ of finite sets Φ of such polynomials; there is no negation connective, and the negation of a polynomial f is the polynomial $\bar{f} = 1 + f$. The axioms are the cedents $f, \bar{f}$ for every polynomial f. The rules are the two rules of polynomial calculus with side formulas,

$$\frac{\Gamma, f}{\Gamma, fg} \quad \text{(product)}, \qquad \frac{\Gamma, f \qquad \Delta, g}{\Gamma, \Delta, f + g} \quad \text{(sum)},$$

together with weakening, cut on an arbitrary formula, the $\bigwedge$-rule (from Γ, φ for every $\varphi \in \Phi$ infer $\Gamma, \bigwedge \Phi$) and the $\bigvee$-rule (from Γ, φ for some $\varphi \in \Phi$ infer $\Gamma, \bigvee \Phi$). The degree of a proof is the maximum degree of a polynomial in it, and its size is the number of symbols. Theorem 43 of [BKZ15] converts a quasipolynomial-size depth-d $\mathrm{AC}^0[2]$-Frege refutation of a CNF into a quasipolynomial-size $\mathrm{PCK}^1_{\mathbb{F}_2}$ refutation of degree $(\log n)^{O(d)}$.

Lemma D.2 (Equivalence with $\mathrm{PCK}^1_{\mathbb{F}_2}$)**.** *For every b, the depth-3.5 $\mathrm{AC}^0[2]$-Frege system of Definition D.1 with bottom degree b and $\mathrm{PCK}^1_{\mathbb{F}_2}$ restricted to polynomials of degree at most b simulate each other with polynomial overhead in size.*

Proof sketch. The two systems have the same lines. A $\mathrm{PCK}^1_{\mathbb{F}_2}$ cedent is an OR of polynomial assertions, conjunctions of them, and disjunctions of them; complementing every polynomial, $f = 0$ becoming $\bar{f} = 1$, translates the assertions into our equations and changes neither degrees nor sizes by more than a constant factor, and a disjunction $\bigvee \Phi$ inside a cedent is the same as its members listed in the cedent, i.e. as singleton terms. Under this translation the rules correspond one to one. The sum rule is our linear combination, with the side formulas as contexts, and the product rule is our multiplication, the multiplier g being unrestricted in both systems while the displayed polynomials have degree at most b. The axiom $f, \bar{f}$ is our default axiom $(u = 1) \vee (\bar{u} = 1)$. A cut

on a polynomial is a cut on a singleton term; a cut on a conjunction $\bigwedge \Phi$ is against its negation, which in $\mathrm{PCK}^1_{\mathbb{F}_2}$ is the disjunction $\bigvee \bar{\Phi}$ of the complemented polynomials and in our system is the disjunction of the complements of the equations of the term, which is our cut rule; and a cut on a disjunction is the same inference with the premises exchanged. The $\bigwedge$-rule with $|\Phi|$ premises is $|\Phi| - 1$ binary conjunction introductions followed by contractions, the $\bigvee$-rule is weakening, and weakening is weakening. In the other direction, projection from a term $\bigwedge \Phi$ to a subterm $\bigwedge \Phi'$, $\Phi' \subseteq \Phi$, is a cut on the formula $\bigwedge \Phi$, whose negation is $\bigvee \bar{\Phi}$. For each $f \in \Phi'$ the axiom $\bar{f}, f$ gives the cedent $\bigvee \bar{\Phi}, f$ by the $\bigvee$-rule, since $\bar{f} \in \bar{\Phi}$; the $\bigwedge$-rule over Φ' then gives $\bigvee \bar{\Phi}, \bigwedge \Phi'$, and a cut against the premise $\Gamma, \bigwedge \Phi$ yields $\Gamma, \bigwedge \Phi'$. Binary conjunction introduction is the $\bigwedge$-rule with two premises. Every step is local; each original inference is simulated by a constant number of lines, except projection, which uses one line per equation of the subterm; and the degree of every polynomial, including the multilinearized conclusions of the product rule, stays within b. □

Together with Theorem 43 of [BKZ15], the lemma shows that a quasipolynomial-size depth-d $\mathrm{AC}^0[2]$-Frege refutation of a CNF yields a quasipolynomial-size refutation in our system with $b = (\log n)^{O(d)}$, which is the sense in which Theorem 1.4 is one parameter away from a lower bound for depth-d $\mathrm{AC}^0[2]$-Frege.

Proof of Proposition 5.2. All identities are multilinearized. We use induction on the proof and Lemma 4.10 to multiply known members of $\widehat{C}_D$. The ordinary degree of an appended factor bounds the increase in coefficient degree. Expanding the products defines polynomials $V_{\mathcal{T},j}$ and $U_{L,\mathcal{T}}$ with

$$1 + A_{\mathcal{T}} = \sum_j V_{\mathcal{T},j}(1 + u_{\mathcal{T},j}), \qquad \deg V_{\mathcal{T},j} \le R - b, \tag{50}$$

$$1 + P_L = \sum_{\mathcal{T} \in L} U_{L,\mathcal{T}} A_{\mathcal{T}}, \qquad \deg U_{L,\mathcal{T}} \le R t_{\mathrm{out}} - R. \tag{51}$$

Relations between terms. The inner error conditions say that on $\operatorname{supp} f$ each $A_{\mathcal{T}}$ is exact. The following two relations are the consequences we need: on $\operatorname{supp} f$, a larger term's approximant implies a smaller term's, and the approximant of a conjunction is the product of the approximants of its parts. If every equation of $\mathcal{U}$ occurs in $\mathcal{T}$, the inner error conditions and (50) give

$$f A_{\mathcal{T}}(1 + A_{\mathcal{U}}) \in \widehat{C}_D \qquad \text{in coefficient degree at most } Q + 2R. \tag{52}$$

If $\mathcal{V} = \mathcal{T} \wedge \mathcal{U}$, each equation of $\mathcal{V}$ occurs in one of its two parts. Multiplying that part's inner error by the other approximant, then using (50) for $\mathcal{V}$, gives

$$f A_{\mathcal{T}} A_{\mathcal{U}}(1 + A_{\mathcal{V}}) \in \widehat{C}_D \qquad \text{in coefficient degree at most } Q + 3R. \tag{53}$$

Also, (52) for the two parts and $1 + A_{\mathcal{T}} A_{\mathcal{U}} = (1 + A_{\mathcal{T}}) + A_{\mathcal{T}}(1 + A_{\mathcal{U}})$ give $f A_{\mathcal{V}}(1 + A_{\mathcal{T}} A_{\mathcal{U}}) \in \widehat{C}_D$. Adding this to (53) yields

$$f(A_{\mathcal{V}} + A_{\mathcal{T}} A_{\mathcal{U}}) \in \widehat{C}_D. \tag{54}$$

The polynomial in (54) has coefficient degree at most $Q + 2R$; this is the bound used when it is multiplied further. Thus the relation needed for conjunction follows from the inner error conditions for the separately chosen approximants.

Removing a premise term. Suppose a premise is $L_0 = A \vee \mathcal{T}$, and the conclusion L retains every term of A. The calculation of Lemma 3.6 (the surviving disjunct is true), with the equations replaced by terms, their indicators u_i by the approximants $A_{\mathcal{T}}$, and (5) by (51), gives

$$(1 + A_{\mathcal{T}}) f P_L \in \widehat{C}_D. \tag{55}$$

Its derivation has coefficient degree at most $Q + 2Rt_{\text{out}} + R$, and its final polynomial has coefficient degree at most $Q + Rt_{\text{out}} + R$.

Conjunction introduction and projection. For introduction, apply (55) to both premises, and use the identity

$$\begin{aligned} f P_L = (1 + A_{\mathcal{T}}) f P_L + A_{\mathcal{T}}(1 + A_{\mathcal{U}}) f P_L \\ + (A_{\mathcal{V}} + A_{\mathcal{T}} A_{\mathcal{U}}) f P_L + A_{\mathcal{V}} f P_L. \end{aligned}$$

The first two terms are (55) for the two premises, multiplied by 1 and by $A_{\mathcal{T}}$; the third follows from (54); the last is the conclusion's outer error condition for $\mathcal{V}$. For projection use

$$f P_L = (1 + A_{\mathcal{T}}) f P_L + A_{\mathcal{T}}(1 + A_{\mathcal{U}}) f P_L + A_{\mathcal{T}} A_{\mathcal{U}} f P_L,$$

where the second term follows from (52) and the last from the outer error for $\mathcal{U}$. All these multiplications obey (41): the largest new coefficient degree is $Q + Rt_{\text{out}} + 2R$, in addition to the derivations already given.

Conjunction cut. Write the premises as $L_1 = A \vee \mathcal{T}$ and $L_2 = B \vee \bigvee_j (1 + u_{\mathcal{T},j} = 1)$, with conclusion $L = A \vee B$. The exact approximant of the singleton $(1 + u_{\mathcal{T},j} = 1)$ is $1 + u_{\mathcal{T},j}$. Telescoping for L_2 therefore gives

$$\begin{aligned} f P_L A_{\mathcal{T}} = A_{\mathcal{T}} P_L (f P_{L_2}) + \sum_{\mathcal{U} \in B} A_{\mathcal{T}} U_{L_2,\mathcal{U}} (f P_L A_{\mathcal{U}}) \\ + \sum_j P_L U_{L_2,j} (f A_{\mathcal{T}} (1 + u_{\mathcal{T},j})) \in \widehat{C}_D. \end{aligned}$$

Here $U_{L_2,j}$ is the coefficient for that singleton in (51). The three kinds of terms follow from the simulated second premise, outer error conditions, and inner error conditions, respectively. Their coefficient degrees are at most $Q + 2Rt_{\text{out}} + R$. Adding (55) for the first premise gives $f P_L \in \widehat{C}_D$.

The remaining rules and inputs. The polynomial and structural calculations of Subsection 3.1 apply with these disjunctive contexts: singleton approximants are exact, and every retained term has the same approximant in premise and conclusion. Their coefficient degrees are at most $Q + 2Rt_{\text{out}} + 2R$. This includes multiplication with an unrestricted multiplier, since the calculation uses only the conclusion of degree at most b. A default axiom contains an inner approximant equal to one, or the two exact inner approximants $1 + u, u$; their outer error conditions give the required membership. For an input clause encoded by $1 + a = 1$, Lemma 4.11 gives $f a \in \widehat{C}_D$, and Lemma 3.7 gives $f P_L \in \widehat{C}_D$ within coefficient degree $Q + Rt_{\text{out}} + w(G)$. These uses satisfy the other two allowances in (41). At the empty line $P_L = 1$, completing the induction. □

References

[AGIPRT26] Yaroslav Alekseev, Svyatosav Gryaznov, Dmitry Itsykson, Toniann Pitassi, Robert Robere and Iddo Tzameret. Hard Instances for Strong Semantic Proof Systems: CNFs for Depth-Four PC and Deep Instances for $\mathsf{AC}^0[2]$-Frege, Manuscript, November 2026.

[Ajt88] Miklós Ajtai. The complexity of the pigeonhole principle. In *Proceedings of the 29th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 346–355, 1988. Journal version: *Combinatorica*, 14(4):417–433, 1994. https://doi.org/10.1109/SFCS.1988.21951.

[BIK+92] Paul Beame, Russell Impagliazzo, Jan Krajíček, Toniann Pitassi, Pavel Pudlák, and Alan Woods. Exponential lower bounds for the pigeonhole principle. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing (STOC)*, pages 200–220, 1992. https://doi.org/10.1145/129712.129733.

[BIK+96] 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(1):1–26, 1996. Preliminary version in FOCS 1994. https://homes.cs.washington.edu/~beame/papers/nsatz.pdf.

[BKZ15] S. R. Buss, L. A. Kołodziejczyk, and K. Zdanowski, Collapsing modular counting in bounded arithmetic and constant depth propositional proofs, *Transactions of the American Mathematical Society*, 367(11):7517–7563, 2015.

[Bra26] Kamil Braun. An exponential lower bound for the bit pigeonhole principle in resolution over parities. *arXiv:2609.23015v1*, 2026. https://arxiv.org/abs/2609.23015.

[CEI96] Matthew Clegg, Jeffery Edmonds, and Russell Impagliazzo. Using the Gröbner basis algorithm to find proofs of unsatisfiability. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC)*, pages 174–183, 1996. https://doi.org/10.1145/237814.237860.

[CHL+26] Eshan Chattopadhyay, Pooya Hatami, Chin Ho Lee, Shachar Lovett, Avishay Tal, and Emanuele Viola. Exponential correlation bounds for polynomials. *arXiv:2609.28839v2*, 2026. https://arxiv.org/abs/2609.28839.

[Con26] Jonas Conneryd. Random $(\log n)$-CNF formulas are hard for depth-2 Frege with modular gates. *arXiv:2610.40397v1*, 2026. https://arxiv.org/abs/2610.40397.

[DKSS13] Zeev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to Kakeya sets and mergers. *SIAM Journal on Computing*, 42(6):2305–2328, 2013. https://doi.org/10.1137/100783704.

[FG15] Michael A. Forbes and Venkatesan Guruswami. Dimension expanders via rank condensers. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 800–814. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2015. https://doi.org/10.4230/LIPIcs.APPROX-RANDOM.2015.800. Full version: https://arxiv.org/abs/1411.7455.

[FSS84] Merrick Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17:13–27, 1984. Preliminary version in FOCS 1981. https://doi.org/10.1007/BF01744431.

[GK16] Venkatesan Guruswami and Swastik Kopparty. Explicit subspace designs. *Combinatorica*, 36(2):161–185, 2016. https://doi.org/10.1007/s00493-014-3169-1.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing (STOC)*, pages 6–20, 1986. https://doi.org/10.1145/12130.12132.

[KPW95] Jan Krajíček, Pavel Pudlák, and Alan Woods. An exponential lower bound to the size of bounded depth Frege proofs of the pigeonhole principle. *Random Structures & Algorithms*, 7(1):15–39, 1995. https://doi.org/10.1002/rsa.3240070103.

[Pan26] Shuo Pang. A degree–size relation for resolution over polynomials. *arXiv:2610.00837v1*, 2026. https://arxiv.org/abs/2610.00837.

[PBI93] Toniann Pitassi, Paul Beame, and Russell Impagliazzo. Exponential lower bounds for the pigeonhole principle. *Computational Complexity*, 3(2):97–140, 1993. https://doi.org/10.1007/BF01200117.

[Pit97] Toniann Pitassi. Algebraic propositional proof systems. In Neil Immerman and Phokion G. Kolaitis, editors, *Descriptive Complexity and Finite Models*, volume 31 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 215–244. American Mathematical Society, 1997. Author manuscript dated July 11, 1996; see Section 6.2. https://www.cs.toronto.edu/~toni/Papers/pitassi-algebraic.pdf.

[Raz87] Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes*, 41(4):333–338, 1987. Russian original: *Matematicheskie Zametki*, 41(4):598–607. https://doi.org/10.1007/BF01137685.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC)*, pages 77–82, 1987. https://doi.org/10.1145/28395.28404.