Quasi-polynomial Frege Simulation of IPS beyond Noncommutativity

Abhranil Chatterjee* Prerona Chatterjee† Utsab Ghosal‡ Partha Mukhopadhyay§

February 8, 2026

Abstract

In a landmark result, Grochow and Pitassi [GP18] established a close connection between propositional proof systems (such as Frege or Extended Frege) and the algebraic proof system IPS. In particular, they showed that the ability of propositional systems to reason about polynomial identities—and, more specifically, the expressibility power of polynomial identity witnesses within propositional proof systems—plays a central role. The strongest unconditional result in this direction is due to Li, Tzameret, and Wang [LTW18], who proved that *Frege* and the *noncommutative* IPS are quasi-polynomially equivalent, by giving a Frege simulation of noncommutative polynomial identity witnesses.

In this paper, we extend this line of work by showing such a quasi-polynomial equivalence holds even beyond the noncommutative setting. Specifically, our result applies when the commuting graph of the variables (a graph with one vertex for each variable, and an edge between two vertices if and only if the corresponding variables do not commute) is a disjoint union of two cliques of unbounded size and $O(1)$ isolated vertices, a structure motivated by the classical work of Cartier and Foata [CF69]. This is a step towards bridging the gap between our understanding of noncommutative IPS and commutative IPS.

To extend the result beyond the noncommutative setting, we must overcome several additional technical obstacles. Beyond the techniques of [LTW18, GP18], our approach relies on tools from skew field theory [Coh95a] and on algorithms for computing noncommutative rank [FR04, IKQS15, IQS18]. This may be of independent interest in algebraic proof complexity.

*Department of Computer Science and Engineering, IIT Kanpur. Email: abhneil@gmail.com

†School of Computer Sciences, NISER (OCC of HBNI), Bhubaneswar, India. Email: prerona.ch@gmail.com

‡Chennai Mathematical Institute, Chennai, India. Email: ghosal@cmi.ac.in.

§Chennai Mathematical Institute, Chennai, India. Email: partham@cmi.ac.in.

Contents

1 Introduction **3**

1.1 Our Contribution . . . 5

1.2 Proof Overview . . . 7

1.2.1 Proof Overview for the Frege Simulation . . . 7

1.2.2 Proof Overview for the Construction of *Witness Identities* . . . 8

1.3 Further Remarks . . . 11

2 Preliminaries **12**

2.1 Computational Models . . . 12

2.2 The Ideal Proof Systems . . . 13

2.3 The Frege Proof System . . . 14

3 Partially Commutative Proof Systems **15**

3.1 Soundness and Completeness . . . 16

3.2 Verifiability . . . 18

3.3 $\mathsf{PC}_{\mathsf{p,q}}$- IPS simulates the *Frege* proof system . . . 18

4 Frege Quasi-Polynomially Simulates $\mathsf{PC}_{\mathsf{k,2}}$- IPS for any constant k **19**

4.1 Formal Statement of our Main Theorem . . . 20

4.2 The Reflection Principle . . . 21

4.3 Homogenization of Partially Commutative Formulas . . . 21

4.4 Translation Between Formulas and ABPs . . . 22

4.5 Refined Translation Between Formulas and ABPs . . . 24

4.5.1 Refined ABP Construction . . . 24

4.5.2 Refined ABP Tracking . . . 25

4.5.3 Size of the degree-refined induced formulas. . . . 27

4.5.4 Local ABP identities in A'_F. . . . 28

4.6 Identity Witnesses . . . 29

4.6.1 ABP identity witnesses over $\mathbb{F}[\overline{y}]$ . . . 29

4.7 The Frege Simulation . . . 30

5 Existence of a Partially Commutative ABP Identity Witness **34**

5.1 Linear Matrix Factorization . . . 35

5.2 Matrix Polynomial Factorization . . . 36

5.2.1 Higman's Linearization for a Matrix of Noncommutative ABPs . . . 36

5.2.2 Proof of Factorization Lemma . . . 40

5.3 Construction of witness matrices . . . 45

A Remaining details from Section 3 **54**
A.1 Boolean Solutions Exist if and only if Matrix Solutions Exist 54
A.2 $\mathsf{PC}_{p,q}$-IPS simulating *Frege* 55

B Remaining Details from Section 4 **57**
B.1 Frege Proof for Homogenisation of Partially Commutative Formulas 57
B.2 Definition of the Translation Map between ABP and Formula 59
B.3 D′ as defined in Subsection 4.5.2 is Well-Defined 60

C Remaining details of Section Section 5 **61**
C.1 Proof of linear matrix factorization 61
C.2 Toy Example for the factorisation 65

1 Introduction

The Ideal Proof System (IPS), introduced by Grochow and Pitassi [GP18], reformulates propositional unsatisfiability as an algebraic ideal membership problem: a refutation is a single algebraic circuit certifying that the constant polynomial 1 belongs to the ideal generated by the polynomial translation of a CNF together with the Boolean axioms. This formulation enables a direct connection between propositional proof complexity and algebraic circuit complexity, since lower bounds on IPS refutations imply corresponding lower bounds for algebraic circuits.

A central obstacle in relating IPS to a standard propositional proof systems is verification. For the original *commutative* IPS, correctness of a certificate is known only via *Polynomial Identity Testing (PIT)*, and hence is inherently randomized. Grochow and Pitassi isolated a small collection of propositional *PIT axioms* expressing the correctness of a Boolean circuit for PIT and showed that, if a propositional proof system admits short proofs of these axioms, then it can simulate commutative IPS. Subsequent work shows that this dependence is unavoidable: commutative IPS can be p-simulated by a Cook–Reckhow proof system [CR79] if and only if PIT lies in **NP** [Gro23]. Consequently, understanding *which PIT axioms can be efficiently proved inside propositional proof systems* becomes a central question for connecting propositional and algebraic proof systems.

Informally, these PIT axioms can be viewed as directly reflecting the definition of an IPS certificate (Definition 2.3). Indeed, an IPS refutation is a polynomial $P(X, Z)$ satisfying the two identities $P(X, \overline{0}) = 0$ and $P(X, f_1, \ldots, f_m) = 1$. The PIT axioms assert, in propositional form, the soundness of these two instantiations of the same proof polynomial, together with basic closure properties of polynomial identity.

Commutative IPS. We first recall the original (commutative) Ideal Proof System introduced by Grochow and Pitassi [GP18], which will serve as the main algebraic proof system underlying our discussion on the PIT axioms and their provability in propositional proof systems.

> Let $\mathbb{F}$ be a field and let $f_1, \ldots, f_m \in \mathbb{F}[X]$ be a set of unsatisfiable polynomials, where the Boolean axioms $x_i^2 - x_i$ are included among the f_i. A commutative IPS certificate is a polynomial $P(X, Z) \in \mathbb{F}[X, Z]$ such that
>
> $$P(X, \overline{0}) = 0 \qquad \text{and} \qquad P(X, f_1(X), \ldots, f_m(X)) = 1.$$
>
> The size of a refutation is the size of an algebraic circuit[1] computing $P(X, Z)$.

While the commutative IPS provides a uniform algebraic formalization of propositional unsatisfiability, its interaction with propositional proof systems is mediated through the provability of *PIT axioms*, and hence through the ability of propositional systems to reason about polynomial identity testing. A strikingly different situation arises in the noncommutative setting, where deterministic polynomial identity testing is available. This leads to a much tighter and unconditional connection with *Frege* proofs, as shown by Li, Tzameret, and Wang [LTW18].

Noncommutative IPS. The noncommutative IPS (NC-IPS) of Li, Tzameret, and Wang [LTW18] provides a crucial bridge after the above discussion on the role and provability of PIT axioms in propositional proof systems.

> Let $\mathbb{F}$ be a field and let $f_1, \ldots, f_m \in \mathbb{F}\langle x_1, \ldots, x_n \rangle$ be noncommutative unsatisfiable polynomials which include the boolean axioms ($x_i(1 - x_i)$ for all $i \in [n]$) and the commutator axioms ($x_i x_j - x_j x_i$ for all $i < j$).
>
> An NC-IPS certificate is a noncommutative polynomial $P(X, Z) \in \mathbb{F}\langle X, Z \rangle$ such that
>
> $$P(x_1, \ldots, x_n, \overline{0}) = 0 \qquad \text{and} \qquad P(X, f_1, \ldots, f_m) = 1.$$
>
> The size of an NC-IPS refutation is the size of a noncommutative formula computing the polynomial $P(X, Z)$.

By working with noncommutative formulas and commutator axioms, and by exploiting the deterministic PIT algorithm of Raz and Shpilka [RS05] for noncommutative formulas, Li, Tzameret, and Wang [LTW18] obtain an essentially tight, unconditional characterization of *Frege*: *Frege* and noncommutative IPS are quasi-polynomially equivalent.

From this perspective, it follows that strong propositional proof systems can reason about a nontrivial form of PIT without additional assumptions. This motivates a systematic comparison between the PIT axioms required for simulating commutative IPS and the deterministic PIT principles available in the noncommutative setting.

[1] or sometimes algebraic formula. See Definition 2.1 for the definitions of algebraic circuits and formulas.

In particular, understanding which PIT axioms can be proved inside propositional proof systems provides a natural route to connect commutative and noncommutative algebraic proof systems. A natural quest is therefore to search for intermediate proof systems that act as a bridge between NC-IPS and commutative IPS, for which suitable PIT axioms can be efficiently proved within propositional proof systems.

1.1 Our Contribution

Partially Commutative IPS. Inspired by the previous discussions, we define a family of proof systems, called *Partially Commutative* IPS, $\{\mathsf{PC_{p,q}\text{-}IPS}\}_{q\leq p}$. It bridges noncommutative IPS and Commutative IPS in the following sense: $\mathsf{PC_{1,1}\text{-}IPS} \equiv \mathsf{NC\text{-}IPS}$ and $\mathsf{PC_{n,0}\text{-}IPS} \equiv \mathsf{IPS}$.

We begin by looking at the notion of a *partially commutative polynomial*.

> Let X_1, X_2 be two sets of variables such that $X = \{x_{1,1}, \ldots, x_{1,n}\}$ and $X_2 = \{x_{2,1}, \ldots, x_{2,n}\}$ and let $\mathbb{F}$ be a field.
>
> A partially commutative monomial $m \in \mathbb{F}\langle X_1 \sqcup X_2\rangle$ is uniquely determined by its restrictions to X_1 and to X_2: $m|_{X_1} = m_{X_1} \in \mathbb{F}\langle X_1\rangle$ and $m|_{X_2} = m_{X_2} \in \mathbb{F}\langle X_2\rangle$, where m_{X_1} and m_{X_2} are noncommutative monomials. All variables in X_1 commute with all variables in X_2. Hence, for $m, m' \in \mathbb{F}\langle X_1 \sqcup X_2\rangle$, we have $m = m'$ if and only if $m|_{X_1} = m'|_{X_1}$ and $m|_{X_2} = m'|_{X_2}$. This induces an equivalence relation $\sim$ on the set of monomials.[2]
>
> A partially commutative polynomial $f \in \mathbb{F}\langle X_1 \sqcup X_2\rangle$ is an $\mathbb{F}$-linear combination of the partially commutative monomials over X_1 and X_2.[3] A partially commutative polynomial over p buckets, say $f \in \mathbb{F}\langle X_1 \sqcup \cdots \sqcup X_p\rangle$, can be defined analogously.

This is a very well-studied model with connections to the problem of multiplicity equivalence testing of multi-tape automata [HK91, Wor13]. The study of one-way multi-tape finite automata was initiated in the seminal paper of Rabin and Scott [RS59]. The multiplicity equivalence testing problem for multi-tape automata is reduced to the identity testing problem over the partially commutative ring and a deterministic polynomial-time algorithm is designed only very recently (over $\mathbb{Q}$) for the number of tapes $O(1)$ [ACM24].

In particular, the bucketing structure of the variables defines a *commuting graph*, a structure motivated by the classical work of Cartier and Foata [CF69].

> The commuting graph has one vertex for each variable in the polynomial and an edge exists between two vertices if and only if the corresponding variables do not commute.

[2]For example, consider two monomials $m_1 = x_{1,1}x_{2,1}x_{1,2}x_{2,2}$ and $m_2 = x_{2,1}x_{1,1}x_{2,2}x_{1,2}$. These two monomials are the same with $m_1|_{X_1} = m_2|_{X_1} = x_{1,1}x_{1,2}$ and $m_1|_{X_2} = m_2|_{X_2} = x_{2,1}x_{2,2}$.

[3]It is sometimes useful to think of polynomials in $\mathbb{F}\langle X_1 \sqcup X_2\rangle$ as a noncommutative polynomial in $\mathbb{F}\langle X_2\rangle\langle X_1\rangle$. That is, f can be uniquely written as $f = \sum_{m_i \in \mathbb{F}\langle X_1\rangle} m_i \cdot f_i$ where $f_i \in \mathbb{F}\langle X_2\rangle$.

Note that a commutative polynomial $f \in \mathbb{F}(X)$ can be thought of as a partially commutative polynomial in $\mathbb{F}\langle X_1 \sqcup \cdots \sqcup X_n \rangle$ where $|X_i| = 1$. Thus a natural generalization of both IPS and NC-IPS is the proof system $\mathsf{PC_{p,q}}$- IPS, which we define in such a way that $\mathsf{PC_{1,1}}$- IPS $\equiv$ NC-IPS and $\mathsf{PC_{n,0}}$- IPS $\equiv$ IPS if the total number of variables is n.

In particular, the input set of equations to a $\mathsf{PC_{p,q}}$- IPS proof are partially commutative polynomials in $\mathbb{F}\left\langle X_1 \sqcup \cdots \sqcup X_p \right\rangle$ where $\left|X_{q+1}\right| = \cdots = \left|X_p\right| = 1$. A formal definition of $\mathsf{PC_{p,q}}$- IPS is given below.

Definition 1.1. (Partially commutative IPS ($\mathsf{PC_{p,q}}$- IPS))) *Let $\mathbb{F}$ be a field and $X = X_1 \sqcup \cdots \sqcup X_p$ be a set of partially commutative variables*[4] *such that $\left|X_{q+1}\right| = \left|X_{q+2}\right| = \cdots = \left|X_p\right| = 1$. Further, for every $i \leq q$, let $X_i = \left\{x_{i,j} \ : \ j \in [n]\right\}$ without loss of generality and, for every $q < i \leq p$, let $X_i = \left\{y_{i-q}\right\}$.*

Assume that $f_1 = f_2 = \cdots = f_m = 0$ is a set of partially commutative polynomial equations from $\mathbb{F}\langle X \rangle$ and suppose that the following set of equations (axioms) are included in f_is.

- ***Boolean axioms**: $x_{i,j}(x_{i,j} - 1)$ for every $i \in [q]$, $j \in [n]$; $y_i^2 - y_i$ for every $i \in [p - q]$.*
- ***Commutator axioms**: $x_{i,j}x_{i,j'} - x_{i,j'}x_{i,j}$ for every $j \neq j' \in [n]$ and $i \in [q]$.*

A partially commutative IPS *proof* ($\mathsf{PC_{p,q}}$- IPS) *of unsatisfiability of the system $\{f_i\}$ is a partially commutative polynomial $P(X, z_1, \ldots, z_m) \in \mathbb{F}\langle X, Z \rangle$ such that,*

1. *The Z variables are noncommuting with every X variable.*
2. *$P(X, \bar{0}) = 0$.*
3. *$P(X, f_1, \ldots, f_m) = 1$.*
4. *The Z variables are the place holder variables. Once they are substituted by partially commutative polynomials f_i, the multiplication respects the partial commutativity.* ◊

Our Result We show that this proof system is sound and complete in Subsection 3.1. We also discuss its verifiability in Subsection 3.2 and the fact that it can efficiently simulate the Frege proof system, in particular Schoenfield's system (Definition 2.3), in Subsection 3.3.

Moreover, the converse is also true up to a quasi-polynomial blow-up in a certain restricted setting. In particular, it extends the work of Li, Tzameret and Wang [LTW18], which showed a quasi-polynomial Frege simulation for $\mathsf{PC_{1,1}}$- IPS.

Theorem 1.2. *Any size-s* $\mathsf{PC_{k,2}}$- IPS *refutation over* $\mathbf{GF}(2)$ *of an unsatisfiable CNF can be simulated by a* Frege *proof of size $s^{(k \log s)^{O(1)}}$.*

Note that the result implies a quasi-polynomial Frege simulation for $\mathsf{PC_{k,2}}$- IPS when $k = \log^{O(1)} s$. Also note that the commuting graph in this setting is the disjoint union of two cliques of size n and $k - 2$ isolated vertices.

[4]For every i, the variables within X_i are noncommuting, but for every $i \neq j$, variables from X_i and X_j commute.

A natural question is whether the result can be extended to the case where the commuting graph is the disjoint union of $O(1)$ cliques of unbounded size. We leave this question open, and we refer to Subsection 1.3 for further remarks along this line.

1.2 Proof Overview

Throughout, the field $\mathbb{F}$ is fixed to be $\mathbf{GF}(2)$. The proof of Theorem 1.2 proceeds in two steps. By the general methodology outlined in [GP18, LTW18], to simulate a $\mathsf{PC}_{\mathsf{k},2}$- IPS refutation of an unsatisfiable CNF by a *Frege* proof, it suffices to give a *Frege* proof of the corresponding partially commutative formula identities over $\mathbf{GF}(2)$, via the reflection principle for $\mathsf{PC}_{\mathsf{k},2}$- IPS (see Subsection 4.2). Given such Frege proofs of identities, the remaining simulation is a direct syntactic extension of the arguments of [LTW18, GP18] to the partially commutative setting.

The novel technical contribution of this work is, therefore, a new witness theorem for partially commutative identities, namely Lemma 4.21, proved in Section 5. Although these results are semantically analogous to the noncommutative witness lemma proven in the work of Li, Tzameret and Wang [LTW18], our proof requires several new ideas since the underlying ring is partially commutative rather than fully noncommutative. Indeed, the definition of IPS itself suggests that any such simulation should have a similar general structure.

We now give a proof sketch of Theorem 1.2 assuming the *witness identities* (that is, the contents of Section 4), and then a proof sketch of our main contribution — a witness lemma for partially commutative identities in certain restricted settings — proved in Section 5.

1.2.1 Proof Overview for the Frege Simulation

Our goal is to prove Theorem 1.2 (see Theorem 4.3 for a more formal statement), namely that every size-s $\mathsf{PC}_{\mathsf{k},2}$- IPS refutation of an unsatisfiable CNF $\Phi := \kappa_1 \wedge \kappa_2 \wedge \ldots \wedge \kappa_m$ can be simulated by a quasi-polynomial size (for $k = O(1)$) *Frege* refutation.

Using the standard arithmetization technique (defined formally in Subsection 4.1), from Φ, we obtain a system of partially commutative polynomials $P_\Phi := \{P_{\Phi,1}, \ldots, P_{\Phi,m}\}$ that are unsatisfiable. Let $F(X, Z)$ be a $\mathsf{PC}_{\mathsf{k},2}$- IPS refutation of the partially commutative system P_Φ over $\mathbb{F}$. Here $X = X_1 \sqcup X_2 \sqcup \cdots \sqcup X_k$ with $|X_1| = |X_2| = n$, $|X_i| = 1$ for $3 \le i \le k$ and Z is a set of placeholder variables. By the definition of $\mathsf{PC}_{\mathsf{k},2}$- IPS, F satisfies the algebraic identities $F(X, 0) = 0$ and $F(X, \overline{P_\Phi}) = 1$. Applying the standard Booleanization map (Definition 4.1), these identities become the Boolean tautologies $\neg\widetilde{F}(X, 0)$ and $\widetilde{F}(X, \widetilde{P}_\Phi)$. By the reflection principle (Lemma 4.4, from [LTW18]), it suffices to give polynomial (or quasi-polynomial) size *Frege* proofs of $\neg\widetilde{F}(X, 0)$ and $\widetilde{F}(X, \widetilde{P}_\Phi)$ in order to derive a *Frege* refutation of Φ.

Since both of these formulas are Booleanizations of the identities of partially commutative formulas that compute the zero polynomial, it is enough to show that whenever a partially commutative formula G computes the identically zero polynomial, the Boolean formula $\neg\widetilde{G}$ has a quasi-polynomial size *Frege* proof. We now briefly explain how this is shown.

We first homogenize the given partially commutative formula F and reduce the task of proving $\neg\widetilde{F}$ in *Frege* to proving certain Booleanized identities for a layered partially commutative ABP A'_F computing the same polynomial (by Lemma 4.6 and the construction of A'_F in Subsection 4.5). Using the refined tracking map and the refined operator D', every partial ABP computation $A'_F[v,t]$ is represented by a degree–refined induced formula F^*_v of controlled size (by Lemma 4.15 and Lemma 4.16). The computation of A'_F can be reasoned about locally inside *Frege* via the layer-wise identities

$$F^*_v = \sum_{(v,u)\in E(A'_F)} A'_F[v,u]\cdot F^*_u$$

provided by Lemma 4.18 and Claim 4.28. These identities allow *Frege* to propagate zeroness from one layer of the ABP to the next. We now discuss how this is done.

Conceptually, we want to show that Frege can certify that the ABP A'_F computes the zero polynomial by propagating local identities layer by layer. This is achieved by using the partially commutative ABP identity witnesses (Lemma 4.21), which yield certain relations (Equation 4.24 and Equation 4.25) consisting matrices in $\mathbb{F}[y_1,\dots y_{k-2}]\langle X_1 \sqcup X_2\rangle^{w\times w}$ [5] such that the degree of $\overline{y}$-variables in each entry of these matrices are polynomially bounded. We call these matrices *witness matrices*. These relations form a backward-propagating chain of identities in the partially commutative ring, each linear in the vector of partial-ABP polynomials (with partially commutative coefficients), and together they certify that the polynomial computed by A'_F is identically zero.

We call the relations obtained in Equation 4.25 as *transition identities*. These reside in the partially commutative algebra $\mathbb{F}[\overline{y}]\langle X_1 \sqcup X_2\rangle$, and so we can reduce each such identity to a polynomially bounded collection of fully noncommutative identities over $\mathbb{F}[\overline{y}]\langle X_2\rangle$ by extracting coefficients with respect to the leftmost X_1-variables (Claim 4.28). Each resulting noncommutative identity over X_2 is simulated in *Frege* using the [LTW18, Lemma 4.9], and the entry-wise identities are combined to obtain the Booleanisation of Equation 4.25 (Equation 4.27) inside *Frege*.

Finally, *Frege* derives the polynomial identity, layer by layer, using the Booleanization of the *vanishing identities* (Equation 4.24), namely Equation 4.26, as the base case and the local equivalences, namely Equation 4.27, as propagation rules. This allows us to prove $\neg\widetilde{F}$ within *Frege*.

The overall proof size is quasi-polynomial ($s^{(k\log s)^{O(1)}}$), since the number of layers and witness identities is polynomial in $|A'_F|$, each induced formula F^*_v has size at most $s^{O(\log^3 s)}$ by Lemma 4.16 and entries of the *witness matrices* has formula complexity $s^{(k\log s)^{O(1)}}$ by Lemma 4.21.

1.2.2 Proof Overview for the Construction of *Witness Identities*

We now give an overview of that fact that we can construct the witness matrices with entries in $\mathbb{F}(y_1,\dots y_{k-2})\langle X_1 \sqcup X_2\rangle$[6] (Lemma 5.1). Additionally, the degree of $\overline{y}$-variables in both the numer-

[5] Recall that $X_i = \{y_{i-2}\}$ for every $i \in [3,k]$.

[6] $\mathbf{GF}(2)(y_1,y_2,\dots,y_{k-2})$ is the (commutative) function field over the polynomial ring $\mathbf{GF}(2)[y_1,y_2,\dots,y_{k-2}]$.

ator and the denominator of each entry is polynomially bounded. Note that "clearing the denominators" from the entries of these matrices shows that we can construct the witness matrices with entries in $\mathbb{F}[y_1, \dots y_{k-2}] \langle X_1 \sqcup X_2 \rangle$ having the required property (Lemma 4.21). For the remainder of this section, we denote $\mathbb{F}(y_1, \dots, y_{k-2})$ by $\mathbb{F}'$.

In order to explain the technicalities of our proof, we first give a brief overview of the proof of the witness lemma shown in [LTW18].

Let A be a noncommutative ABP of depth d computing the identically zero polynomial and let the width of layer i be w_i, with nodes $\{u_{i,1}, \dots, u_{i,w_i}\}$. Also, let s, t be the source and sink nodes in A. For $i \in [d]$, define $\overline{A}_i = \left[A[u_{d-i,1}, t], \dots, A[u_{d-i,w_{d-i}}, t]\right]^\top$. Assuming A is homogeneous, each entry of $\overline{A}_i$ computes a homogeneous polynomial of degree i. We use $A[\ell_1, \ell_2]$ to denote the polynomials computed between the nodes ℓ_1, ℓ_2.

In this setting, [LTW18] constructs matrices $\{\lambda_1, \dots, \lambda_d \mid \lambda_i \in \mathbb{F}^{w_i \times w_i}\}$ and $\{T_1, \dots, T_d \mid T_i \in \mathbb{F}\langle X \rangle^{w_i \times w_{i+1}}\}$ witnessing the zeroness of the polynomial computed by A. They do so using an inductive argument based on Gaussian elimination over $\mathbb{F}$.

Formally, let M be the transition matrix from layer i to $i+1$. Note that it can be uniquely expressed as $M = \sum_{j=1}^{n} x_j M_j$. From the inductive hypothesis, $\lambda_i \cdot \overline{A}_{d-i} = 0$ and $\overline{A}_{d-i} = M \cdot \overline{A}_{d-i-1}$. So, $\sum_{j=1}^{n} x_j (\lambda_i M_j) \cdot \overline{A}_{d-i-1} = 0$, and hence, by homogeneity and noncommutativity, $(\lambda_i M_j) \cdot \overline{A}_{d-i-1} = 0$ for $j \in [n]$.

Let $V = \mathrm{Span}_{\mathbb{F}}\{\mathrm{Rows}(\lambda_i M_j) \mid j \in [n]\} \subseteq \mathbb{F}^{w_{i+1}}$. Every vector in V is orthogonal to $\overline{A}_{d-i-1}$. Choosing a basis matrix of V, say B, by Gaussian elimination on the stacked matrix $(\lambda_i M_1; \dots; \lambda_i M_n)$, we obtain matrices T_j such that $\lambda_i M_j = T_j B$ for all $j \in [n]$. Now if we define $T_{i+1} := \sum_{j=1}^{n} x_j T_j$, the required properties are satisfied. Since all matrices λ_{i+1} and $T_1, \dots, T_n$ have entries in $\mathbb{F}$, this is sufficient for [LTW18] for the expressiveness in *Frege*.

Now we return to our situation. If we apply the same idea, the first important point is the λ_i matrices are not just defined over the field $\mathbb{F}$, rather they are matrices in $\mathbb{F}'\langle X_2 \rangle^{w_i \times w_i}$. Similarly the coefficient matrices $M_j \in \mathbb{F}'\langle X_2 \rangle^{w_i \times w_{i+1}}$. The vector of polynomials $\{\mathrm{Rows}(\lambda_i M_j)\}_{j \in [n]}$ span a left module over the free skew field $\mathbb{F}'\langle\!\langle X_2 \rangle\!\rangle$[7]. So, if we want to construct the T_j matrices using Gaussian elimination, the entries of T_j will be inside the $\mathbb{F}'\langle\!\langle X_2 \rangle\!\rangle$, potentially involving nested inverses. An expression involving nested inverses need not have a canonical expression as a ratio of two noncommutative polynomials. A standard example is given by the following expression: $(z + xy^{-1}x)^{-1} - z^{-1}$. We do not know how to explicitly express such matrices T_j in *Frege*.

We now explain the key idea presented in Section 5. The proof proceeds by induction on $r = 0, 1, \dots, d-1$ and constructs matrices λ'_r and T'_{r+1} satisfying

$$\lambda'_{d-i} \cdot \overline{A}_i = \overline{0} \;\; \forall \;\; i = 1, 2 \dots, d, \quad \lambda'_i \in \mathbb{F}'\langle X_2 \rangle^{w_i \times w_i}. \tag{1.3}$$

[7] $\mathbb{F}'\langle\!\langle X_2 \rangle\!\rangle$ is the universal skew field containing the noncommutative polynomial ring $\mathbb{F}'\langle X_2 \rangle$.

$$\lambda'_{d-i} \cdot \overline{A}_i = T'_{d-i+1} \cdot \lambda'_{d-i+1} \cdot \overline{A}_{i-1} \ \forall \ i = 2, 3, \ldots, d \quad T_i \in \mathbb{F}' \langle X_2 \sqcup X_1 \rangle^{w_i \times w_{i+1}}. \tag{1.4}$$

The base case $r = 0$ is trivial since $\overline{A}_d = 0$.

For the inductive step, assume $\lambda'_r \cdot \overline{A}_{d-r} = 0$. Since the ABP is homogenized over the X_1 variables, applying the procedure described above for the noncommutative case, we obtain that $(\lambda'_r M_j) \cdot \overline{A}_{d-r-1} = 0$ for every $j \in [n]$. As before, we construct the stacked matrix $T = \begin{pmatrix} \lambda'_r M_1 \\ \vdots \\ \lambda'_r M_n \end{pmatrix}$.

Notice that, the next-layer vector $\overline{A}_{d-r-1}$ is in the right kernel of T. At this stage, instead of using Gaussian elimination; we apply the *matrix factorization theory* developed by Cohn [Coh95b] to recover a factorization of T as $C \cdot B$, such that the ABP complexity of the entries of C and B are small. However, Cohn's theory is not fully constructive for the purpose of proof complexity, and we need a more explicit version of it. This is achieved in Lemma 5.8 — our main technical contribution.

The first idea is to *linearize* the matrix T in a co-rank preserving manner. This is usually achieved by a technique called Higman's linearization [Coh95b]. We give a computationally efficient version of it for the matrix of noncommutative ABPs (Lemma 5.14). Let the resulting matrix be L. One of the key steps is to use concepts from *matrix decomposability theorem* of Fortin-Reutenauer [FR04]. In particular, up to a invertible linear transformation we can express L in the following form

$$\left[\begin{array}{c|c} L_{11} & 0 \\ \hline L_{21} & L_{22} \end{array}\right].$$

The rest of the factor extraction procedure follows from a case analysis on the structure of L and it involves Cohn's factorization theorem (Lemma 5.7). The matrix C has the following form

$$C := \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} \quad \text{such that} \quad \lambda'_r M_j = C_j B \text{ for all } j.$$

We set $\lambda'_{r+1} := B$ and $T'_{r+1} := \sum_{j=1}^{n} x_{1,j} C_j$. By construction this gives the transition identity

$$\lambda'_r \cdot \overline{A}_{d-r} = T'_{r+1} \cdot \lambda'_{r+1} \cdot \overline{A}_{d-r-1}.$$

To derive the next vanishing condition, we distinguish two cases. If $\mathrm{ncrank}(T) < w_{r+1}$, then the stacked coefficient matrix $C = (C_1^\top, \ldots, C_n^\top)^\top$ has full noncommutative column rank, hence a left inverse over the free skew field, which implies $\lambda'_{r+1} \cdot \overline{A}_{d-r-1} = 0$. If $\mathrm{ncrank}(T) = w_{r+1}$, then T has full column rank and the identity $T \cdot \overline{A}_{d-r-1} = 0$, together with the X_1-homogeneity of $\overline{A}_{d-r-1}$, forces $\overline{A}_{d-r-1} = 0$, and hence also $\lambda'_{r+1} \cdot \overline{A}_{d-r-1} = 0$.

A crucial technical aspect is the constructive version of Cohn's factorization which is done in

Lemma 5.7. The proof involves computing a shrunk subspace of the linear matrix L, which is given by a basis. Note that for a linear matrix $L = \sum_{i=1}^{n} L_i x_i$ a subspace V is shrunk if

$$\dim\left(\mathrm{Span}\left\langle \bigcup_{i=1}^{n} L_i V \right\rangle\right) < \dim(V).$$

Here, the linear matrix is defined over the variables x_i, while the coefficients lie in the function field $\mathbb{F}'$.

The main algebraic tool is the computation of the second Wong sequences, which were originally used in the context of the noncommutative rank of linear matrices [IKQS15, IQS17]. We use these results to argue that the degrees of $y_1, \ldots, y_{k-2}$ in the rational expressions of the basis elements are polynomially bounded. As a result, the matrix factorization can be made explicit in Frege.

1.3 Further Remarks

The main contribution of this paper is a quasi-polynomial simulation of $\mathsf{PC}_{k,2}$-IPS by the Frege proof system for $k = (\log s)^{O(1)}$. This advances the program of relating commutative and noncommutative IPS with respect to Frege simulability and sheds further light on the strength of Frege. Our proof proceeds by giving a Frege simulation of identity witnesses for partially commutative formulas whose commuting graph is a disjoint union of two cliques of unbounded size and $(k-2)$ singleton vertices.

A natural question is whether this simulation can be extended to commuting graphs that are disjoint unions of $O(1)$ cliques of unbounded size. It would suffice to simulate, in Frege, the identity witness for formulas respecting such a bucketing structure. Arvind, Chatterjee, and Mukhopadhyay [ACM24] give a deterministic polynomial-time algorithm for testing whether a formula (and more generally, an ABP) over $X_1 \sqcup \cdots \sqcup X_k$ is identically zero, for constant k and over the field $\mathbb{Q}$. However, their result is based on solving the more general singularity testing problem for linear matrices over $\mathbb{Q}\langle X_1 \sqcup \cdots \sqcup X_k \rangle$, and it is currently unclear how to adapt this approach to a proof-complexity setting.

A further obstacle is the construction of skew fields containing partially commutative rings. The algorithm of [ACM24] relies on a skew-field construction over $\mathbb{Q}\langle X_1 \sqcup \cdots \sqcup X_k \rangle$, which is presently known only in characteristic zero for $k \geq 3$ [KVV20]. Since our approach also relies on rank arguments, extending our result to $k \geq 3$ buckets would require a skew field containing $\mathbb{F}\langle X_1, \ldots, X_k \rangle$ over fields of positive characteristic, in particular over $\mathbf{GF}(2)$. The existence of such a construction remains open.

By contrast, for two buckets a skew field containing $\mathbb{F}\langle X_1, X_2 \rangle$ exists over arbitrary field ([Coh97]). Nevertheless, this does not currently yield a simulation for $\mathsf{PC}_{k,3}$-IPS, as the matrix factorization results developed in Section 5 do not appear to extend to the ring $\mathbb{F}\langle X_1, X_2 \rangle$.

Organization.

The remainder of the paper is organized as follows. Section 2 introduces computational models and the ideal proof system. Section 3 defines the partially commutative ideal proof system, establishes soundness and completeness, and explains how it simulates Frege. Section 4 presents a detailed proof of the main theorem, assuming Lemma 4.21. Finally, Section 5 develops our main technical contribution, namely the construction of identity witnesses for partially commutative formulas with two unbounded buckets.

2 Preliminaries

Notation. For any $i, j \in \mathbb{N}$ with $i \leq j$, we denote the set $\{i, \ldots, j\}$ by $[i, j]$. In addition, for any $n \in \mathbb{N}$, we denote the set $\{1, \ldots, n\}$ by $[n]$.

2.1 Computational Models

Definition 2.1 (Algebraic Circuits and Formulas)**.** *An algebraic circuit C is a directed acyclic graph with a unique output gate (root) of out-degree 0, and input gates of in-degree 0 (leaves) labeled by variables $x_1, \ldots, x_n$ or constants from $\mathbb{F}$. The internal gates are labeled by $+$ or $\times$.*

Each gate v computes a polynomial f_v defined recursively: if v is an input, then $f_v = \mathsf{label}(v) \in \{x_1, \ldots, x_n\} \cup \mathbb{F}$; if $v = u \ \mathsf{op}\ w$ for $\mathsf{op} \in \{+, \times\}$, then $f_v = f_u \ \mathsf{op}\ f_w$. The polynomial computed at the output gate is the polynomial computed by the circuit.

If the underlying graph is restricted to be a tree, then the model is that of an algebraic formula. The size of an algebraic circuit or formula is the number of wires in it. $\Diamond$

Definition 2.2 (Algebraic Branching Program)**.** *An* algebraic branching program *(ABP) is a layered directed acyclic graph. The vertex set is partitioned into layers $0, 1, \ldots, \ell$, with directed edges only between adjacent layers (i to $i+1$). There is a* source *vertex of in-degree 0 in layer 0, and one out-degree-0* sink *vertex in layer k. Each edge is labeled by an affine $\mathbb{F}$-linear form where $\mathbb{F}$ is the underlying field. The polynomial computed by the ABP is the sum over all source-to-sink directed paths of the ordered product of affine linear forms labeling the path edges.*

The size of an ABP is the number of vertices in it. $\Diamond$

Partially Commutative ABPs Let A be a partially commutative ABP computing a polynomial $\hat{A} \in \mathbb{F}\langle X_1 \sqcup X_2 \rangle$. Moreover, let us assume that A is homogenized with respect to the bucket X_1. That is, every edge of the ABP is labeled by a homogeneous linear form over the X_1 bucket, $\sum_{j=1}^{n} x_{1,j} \cdot f_j$ with coefficients $f_j \in \mathbb{F}\langle X_2 \rangle$. Further, let the number of layers in A be d and the width of A be w. Thus $\deg_{X_1}(\hat{A}) = d$. Finally, let the nodes in layer i be $\{u_{i,1}, u_{i,2}, \ldots, u_{i,w_i}\}$ and s, t be the source and sink nodes respectively.

We then define the following vector of ABPs:

$$\overline{A}_i := [A[u_{d-i,1}, t], A[u_{d-i,2}, t], \cdots, A[u_{d-i,w_{d-i}}, t]]$$

where $A[u_{d-i,j}, t]$ is the partial ABP between $u_{d-i,j}$ and t for every $j \in [w_{d-i}]$.

Observe that each ABP in $\overline{A}_i$ computes a partially commutative polynomial which is homogeneous over X_1 and has X_1-degree i.

Similar notation is used if the ABP A is computing a partially commutative polynomial $\mathbb{F}\langle X_1 \sqcup \cdots \sqcup X_p \rangle$ for an arbitrary number p.

2.2 The Ideal Proof Systems

We start with the definition of the Ideal proof system (IPS) defined in [GP18].

Definition 2.3 (Ideal Proof Systems [FSTW21, GP18, LTW18]). *Let $f_1, \ldots, f_m \in \mathbb{F}[X]$ be a set of polynomials such that $\{f_1, \ldots, f_m, x_1^2 - x_1, \ldots, x_n^2 - x_n\}$ has no common solution*[8]. *A* proof *of the unsatisfiability of this set of polynomial equations, in the* Ideal Proof System (IPS)*, is a polynomial $P(X, y_1, \ldots, y_m, z_1, \ldots, z_n) \in \mathbb{F}[X, Y, Z]$ such that the following holds:*

- $P(X, \overline{0}, \overline{0}) = 0$;
- $P(X, f_1, \ldots, f_m, x_1^2 - x_1, \ldots, x_n^2 - x_n) = 1$.

The size *of an* IPS *proof is the minimal size of a circuit (formula or ABP) computing $P(X, Y, Z)$.*

If $f_1, \ldots, f_m \in \mathbb{F}\langle X \rangle$ are, instead, a set of noncommutative polynomials such that $\{f_1, \ldots, f_m\} \cup \{x_i^2 - x_i : i \in [n]\} \cup \{x_i x_j - x_j x_i : i, j \in [n]\}$ has no common solutions, then a proof *of the unsatisfiability of this set of polynomial equations, in the* noncommutative Ideal Proof System (NC-IPS)*, is a polynomial $P(X, y_1, \ldots, y_m, z_1, \ldots, z_n, w_1, \ldots w_{n^2}) \in \mathbb{F}\langle X, Y, Z, W \rangle$ such that the following holds:*

- $P(X, \overline{0}, \overline{0}) = 0$;
- $P(X, f_1, \ldots, f_m, x_1^2 - x_1, \ldots, x_n^2 - x_n, x_1 x_2 - x_2 x_1, \ldots, x_{n-1} x_n - x_n x_{n-1}) = 1$.

The size *of an* NC-IPS *proof is the minimal size of a noncommutative formula computing $P(X, Y, Z, W)$.* $\diamondsuit$

Remark 2.4. *Suppose we are given a system of equations, say $\{f_1 = 0, \ldots, f_m = 0 : f_i \in \mathbb{F}[x_1, \ldots, x_n]\}$, in the commutative setting. We say that the system is unsatisfiable if there is no $\overline{a} \in \mathbb{F}^n$ such that $f_i(\overline{a}) = 0$ for every $i \in [m]$. However, when we consider the variable set X to be noncommuting, we need to look for a common solution in the matrix algebra over $\mathbb{F}$, denoted by $\mathbb{F}^{d \times d}$ for some $d > 1$. In particular, elements from $\mathbb{F}$ are thought of as 1×1 dimensional matrices. Hence, given a set of noncommutative equations $\{f_1 = 0, \ldots, f_m = 0 : f_i \in \mathbb{F}\langle x_1, \ldots, x_n \rangle\}$, we say that they are unsatisfiable if there does not exist any d, $\{\overline{A} = (A_1, \ldots, A_n) : A_i \in \mathbb{F}^{d \times d}\}$ such that $f_i(\overline{A}) = 0_{d \times d}$ for every $i \in [m]$.* $\diamondsuit$

[8]That is, there does not exist $\bar{x} \in \{0,1\}^n$ such that for every $i \in [m]$, $f_i(\bar{x}) = 0$.

Example 2.5 (Evaluation in Matrix algebra)**.** *$xy - yx + 2$ evaluates to*

$$\begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \cdot \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} - \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \cdot \end{pmatrix} \cdot \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} + 2I_{2\times 2}$$

when we substitute $x = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}, y = \begin{pmatrix} b_1 & b_2 \\ b_3 & b_4 \end{pmatrix} \in \mathbb{F}^{2\times 2}$. ♢

2.3 The Frege Proof System

Definition 2.6 (Boolean Formula)**.** *A* Boolean formula Φ *over variables $x_1, \ldots, x_n$ is a finite rooted tree with a fan-in of at most* 2 *in which:*

- *Each leaf is labeled by either a variable x_i or a Boolean constant $\{0, 1\}$.*
- *Each internal node is labeled by one of the Boolean connectives $\{\vee, \wedge, \neg\}$.*
- *A node labeled by $\neg$ has exactly one child (unary connective).*
- *Nodes labeled by $\vee$ or $\wedge$ have exactly two children (binary connectives).*

The formula computes the Boolean function obtained by evaluating the tree in the natural way: each gate applies its connective to the values computed by its child/ children. The size of the formula is the number of nodes in it, denoted as $|\Phi|$. ♢

We now define the *Frege* proof system.

Definition 2.7. (*Frege* derivation rule, [LTW18, Definition 2.2], [CR79]) *Let $\overline{x}$ be a set of boolean variables. A* Frege *rule is a sequence of propositional formulas $A_0(\overline{x}), \ldots, A_k(\overline{x})$, for $k \geq 0$, written as $\frac{A_1(\overline{x}),\ldots,A_k(\overline{x})}{A_0(\overline{x})}$. When $k = 0$, the Frege rule is called axiom scheme.*

A formula F_0 is said to be derived by the rules from $F_1, \ldots, F_k$ if there are formulas $B_1, \ldots, B_n$ such that for every $i \in [0, 1, , \ldots, k]$, $A_i(B_1/x_1, \ldots, B_n/x_n) = F_i$[9]*. The rule is said to be sound if any assignment satisfying $A_1(\overline{x}), \ldots, A_k(\overline{x})$ also satisfies $A_0(\overline{x})$.* ♢

A proof system is sound if it admits proofs only of tautologies. It is said to be *implicationally complete* if, for all sets of formulas T, if T semantically imply F, then there is a derivation of F in the proof system from the axioms T.

Definition 2.8. (Frege Proof, [LTW18, Definition 2.3], [CR79]) *Given a set of Frege rules, a Frege proof of a boolean formula A is a sequence of boolean formulas such that every formula is either an axiom or derived by one of the rules from previous formulas and terminates at A. The size of the proof is the sum of the sizes of all formulas in the proof. Given a set of sound Frege rules P, we say P is a Frege proof system if it is* implicationally complete. ♢

[9] we define $A_i(B_1/x_1, \ldots, B_n/x_n)$ by the formula A_i where x_i is substituted by B_i.

It is known from the work of Reckhow [Rec76], that all *Frege* proof systems are polynomially equivalent to each other. We now describe the instantiation of the *Frege* proof system that we will work with in this paper, known as the Schoenfield's system. For propositional formulas A, B, C, we abbreviate $\neg A \vee B$ as $A \to B$.

The system has only three axiom schemes and they are as follows.

1. $A \to (B \to A)$.
2. $(\neg A \vee \neg B) \to ((\neg A \to B) \to A)$.
3. $(A \to (B \to C)) \to ((A \to B) \to (A \to C))$.

3 Partially Commutative Proof Systems

In this section we first define partially commutative polynomials and the partially commutative proof system ($\mathsf{PC_{p,q}}$- IPS). We then prove that it is sound and complete, discuss its verifiability and show that it simulates the *Frege* proof system.

Definition 3.1 (Partially Commutative Polynomial). *Let X_1, X_2 be two sets of variables such that $X = \{x_{1,1}, \ldots, x_{1,n}\}$ and $X_2 = \{x_{2,1}, \ldots, x_{2,n}\}$ and let $\mathbb{F}$ be a field. A partially commutative monomial $m \in \mathbb{F}\langle X_1 \sqcup X_2\rangle$ is uniquely determined by its restrictions to X_1 and to X_2: $m|_{X_1} = m_{X_1} \in \mathbb{F}\langle X_1\rangle$ and $m|_{X_2} = m_{X_2} \in \mathbb{F}\langle X_2\rangle$, where m_{X_1} and m_{X_2} are noncommutative monomials. All variables in X_1 commute with all variables in X_2. Hence, for $m, m' \in \mathbb{F}\langle X_1 \sqcup X_2\rangle$, we have $m = m'$ if and only if $m|_{X_1} = m'|_{X_1}$ and $m|_{X_2} = m'|_{X_2}$. This induces an equivalence relation $\sim$ on the set of monomials.*[10]

A partially commutative polynomial $f \in \mathbb{F}\langle X_1 \sqcup X_2\rangle$ is an $\mathbb{F}$-linear combination of the partially commutative monomials over X_1 and X_2.[11] *A partially commutative polynomial over p buckets, say $f \in \mathbb{F}\langle X_1 \sqcup \cdots \sqcup X_p\rangle$, can be defined analogously.* $\Diamond$

Note that a commutative polynomial $f \in \mathbb{F}(X)$ can be thought of as a partially commutative polynomial in $\mathbb{F}\langle X_1, \ldots, X_n\rangle$ where $|X_i| = 1$. As mentioned in the inroduction, inspired by this, we define a generalization of both IPS and NC-IPS. In particular, we define $\mathsf{PC_{p,q}}$- IPS where $\mathsf{PC_{1,1}}$- IPS $\equiv$ NC-IPS and $\mathsf{PC_{n,0}}$- IPS $\equiv$ IPS if the total number of variables is n.

We recall the definition here.

Definition 1.1. (Partially commutative IPS ($\mathsf{PC_{p,q}}$- IPS))) *Let $\mathbb{F}$ be a field and $X = X_1 \sqcup \cdots \sqcup X_p$ be a set of partially commutative variables*[12] *such that $|X_{q+1}| = |X_{q+2}| = \cdots = |X_p| = 1$. Further, for every $i \le q$, let $X_i = \{x_{i,j} \,:\, j \in [n]\}$ without loss of generality and, for every $q < i \le p$, let $X_i = \{y_{i-q}\}$.*

Assume that $f_1 = f_2 = \cdots = f_m = 0$ is a set of partially commutative polynomial equations from $\mathbb{F}\langle X\rangle$ and suppose that the following set of equations (axioms) are included in f_is.

[10] For example, consider two monomials $m_1 = x_{1,1}x_{2,1}x_{1,2}x_{2,2}$ and $m_2 = x_{2,1}x_{1,1}x_{2,2}x_{1,2}$. These two monomials are the same with $m_1|_{X_1} = m_2|_{X_1} = x_{1,1}x_{1,2}$ and $m_1|_{X_2} = m_2|_{X_2} = x_{2,1}x_{2,2}$.

[11] It is sometimes useful to think of polynomials in $\mathbb{F}\langle X_1 \sqcup X_2\rangle$ as a noncommutative polynomial in $\mathbb{F}\langle X_2\rangle\langle X_1\rangle$. That is, f can be uniquely written as $f = \sum_{m_i \in \mathbb{F}\langle X_1\rangle} m_i \cdot f_i$ where $f_i \in \mathbb{F}\langle X_1\rangle$.

[12] For every i, the variables within X_i are noncommuting, but for every $i \neq j$, variables from X_i and X_j commute.

- ***Boolean axioms***: $x_{i,j}(x_{i,j}-1)$ *for every* $i \in [q]$, $j \in [n]$; $y_i^2 - y_i$ *for every* $i \in [p-q]$.
- ***Commutator axioms***: $x_{i,j}x_{i,j'} - x_{i,j'}x_{i,j}$ *for every* $j \neq j' \in [n]$ *and* $i \in [q]$.

A partially commutative IPS *proof* ($\mathsf{PC_{p,q}}$- IPS) *of unsatisfiability of the system* $\{f_i\}$ *is a partially commutative polynomial* $P(X, z_1, \ldots, z_m) \in \mathbb{F}\langle X, Z\rangle$ *such that,*

1. *The* Z *variables are noncommuting with every* X *variable.*
2. $P(X, \overline{0}) = 0$.
3. $P(X, f_1, \ldots, f_m) = 1$.
4. *The* Z *variables are the place holder variables. Once they are substituted by partially commutative polynomials* f_i*, the multiplication respects the partial commutativity.* $\Diamond$

Remark 3.2. *In Definition 1.1, if* $P\langle X, Z\rangle$ *is computed by a model* $\mathcal{C}$ *(can be a circuit, ABP or formula), we denote the proof system by* $\mathcal{C}$- $\mathsf{PC_{p,q}}$- IPS*. If* $\mathcal{C}$ *is circuits or formulas, then each product gate has designated left and right children; the leaves are labeled either by* X *variables or the set of placeholder variables* Z *that satisfy the conditions* (2) *and* (3) *from Definition 1.1. In the case of ABPs, the order of multiplication is maintained layer-wise. The* size *of a* $\mathcal{C}$- $\mathsf{PC_{p,q}}$- IPS *proof is the minimal size of a* $\mathcal{C}$ *(circuit or ABP or formula) that computes the proof polynomial* $P(X, Z)$. $\Diamond$

Remark 3.3. *Analogous to the noncommutative setting, if we consider the partially commutative ring* $\mathbb{F}\langle X_1 \sqcup X_2\rangle$ *with* $|X_1| = |X_2| = n$ *and the set of equations* $\{f_1 = 0, \ldots, f_m = 0 \,:\, f_i \in \mathbb{F}\langle X_1 \sqcup X_2\rangle\}$*, a common solution is a set of* $d \times d$ *matrices (say* $\overline{A} = \{A_1, \ldots, A_n\}$ *and* $\overline{B} = \{B_1, \ldots, B_n\}$*) for some* $d > 1$*, such that the matrices from* $\overline{A}$ *and* $\overline{B}$ *commute with each other and that* $f_i(\overline{A}, \overline{B}) = 0$.

For the partially commutative ring $\mathbb{F}\left\langle X_1 \sqcup X_2 \sqcup \cdots \sqcup X_p\right\rangle$ *where, for some* $0 \leq q \leq p$, $|X_i| = n$ *for every* $i \in [q]$ *and* $|X_i| = 1$ *for every* $i \in [q+1, p]$*, we think of it as the ring* $\mathbb{F}[X_{q+1}, \ldots, X_p]\left\langle X_1 \sqcup \cdots \sqcup X_q\right\rangle$. *So a set of equations* $\{f_1 = 0, \ldots, f_m = 0\}$ *in this setting have a common solution if there exist* $\left\{a_{q+1}, \ldots, a_p\right\} \in \mathbb{F}$ *and a set of* $d \times d$ *matrices (say* $\overline{A_1} = \{A_{1,1}, \ldots, A_{1,n}\}, \ldots, \overline{A_q} = \left\{A_{q,1}, \ldots, A_{q,n}\right\}$*) for some* $d > 1$*, such that the matrices from* $\overline{A_i}$ *and* $\overline{A_j}$ *commute with each other for every* i, j *and that* $f_i(\overline{A_1}, \ldots, \overline{A_q}, a_{q+1}, \ldots, a_p) = 0$. $\Diamond$

3.1 Soundness and Completeness

Lemma 3.4. *For every* p, q*, the proof system* $\mathsf{PC_{p,q}}$- IPS *is sound and complete for CNFs.*

Proof. Soundness follows from the definition of the system. Assume we have a polynomial $P(X, Z)$ that satisfies the properties $(1) : P(X, \overline{0}) = 0$ and $(2) : P(X, f_1, \ldots, f_m) = 1$. If the set of equations $\{f_i\}$ has a common solution $\overline{A} = \left(A_{1,1}, \ldots, A_{q,n}, a_{q+1}, \ldots a_p\right)$, where each $A_{i,j} \in \mathbb{F}^{d\times d}$ for some $d \geq 1$ when $i \leq q$ and each $a_i \in \mathbb{F}$ for $i > q$, then substituting the variables with $\overline{A}$ we obtain a contradiction since

$$(1) \implies P(\overline{A}, \overline{0}) = 0_{d\times d} \neq I_{d\times d} = P(\overline{A}, \overline{0}) \impliedby (2)$$

To prove *completeness*, we first claim the following.

Claim 3.5. *Let* $\{f_1 = 0, \ldots, f_m = 0 \,:\, f_i \in \mathbb{F}\left\langle X_1 \sqcup \cdots \sqcup X_p \right\rangle\}$ *(with* $|X_j| = n$ *for every* $j \in [q]$ *and* $|X_j| = 1$ *for every* $j \in [q+1, p]$*) be a set of equations that include the* boolean axioms *and the* commutator axioms*. Then the given system has a common solution in* $\mathbb{F}^{(n\times q)+(p-q)}$ *if and only if there exist* $a_{q+1}, \ldots a_p \in \mathbb{F}$ *and* $d \geq 1$*, set of matrices* $\{A_{1,1}, \ldots, A_{q,n}\} \subseteq \mathbb{F}^{d\times d}$ *which satisfy the given system.*

The proof of the above claim can be found in Subsection A.1, but here we use the statement to prove completeness.

Consider the system of unsatisfiable equations $\{f_1 = 0, \ldots, f_m = 0 \mid f_i \in \mathbb{F}\langle X\rangle\}$ that includes the *boolean and commutator* axioms. Here $X = X_1 \sqcup X_2 \sqcup \cdots \sqcup X_p$ where, for some $0 \leq q \leq p$, $|X_i| = n$ for every $i \in [q]$ and $|X_i| = 1$ for every $i \in [q+1, p]$. We want to demonstrate the existence of the proof polynomial $P(X, Z) \in \mathbb{F}\langle X, Z\rangle$.

Moreover, assume that the axioms of the system other than the boolean and commutator axioms are obtained by a suitable arithmetization of an unsatisfiable CNF (as described below).

$$T(x_{ij}) = (1 - x_{i,j}) \quad T(y_i) = (1 - y_i) \qquad T(\neg x_{ij}) = x_{i,j} \quad T(\neg y_i) = y_i \qquad T(A \vee B) = T(A)\cdot T(B).$$

For the partially commutative setting, we fix an order of variables in each bucket and multiply them accordingly.

Now, for every $t \in [q]$, let $\mathcal{I}_t$ be the two sided ideal generated by the commutators $\{x_{t,j}x_{t,j'} - x_{t,j'}x_{t,j} \mid j' \neq j \in [n]\}$ and let $\pi : \mathbb{F}\left\langle X_1 \sqcup \cdots \sqcup X_p\right\rangle \to \mathbb{F}[X_1 \sqcup \cdots \sqcup X_p]$ be the surjective homomorphism that maps $f \in \mathbb{F}\left\langle X_1 \sqcup \cdots \sqcup X_p\right\rangle$ to $f \pmod{\sqcup_{t=1}^{q} \mathcal{I}_t} \in \mathbb{F}[X_1 \sqcup \cdots \sqcup X_p]$. We denote $\pi(f_i)$ by f_i'.

By Claim 3.5, we know that partially commutative unsatisfiability in matrix algebra over $\mathbb{F}$ implies commutative unsatisfiability over $\mathbb{F}$. Since the *boolean axioms* are included in $\{f_i\}$, the only possible solutions in $\mathbb{F}$ must be boolean. However, the system does not have solutions due to unsatisfiability in matrix algebra over $\mathbb{F}$. Thus, we conclude that the system has no solution in $\overline{\mathbb{F}}$ as well. So, by Hilbert's Nullstellensatz, there are polynomials $g_1', \ldots, g_m' \in \mathbb{F}[X_1 \sqcup \cdots \sqcup X_p]$ such that $\sum_{i=1}^{m} g_i' \cdot f_i' = 1$. Let $g_i \in \mathbb{F}\left\langle X_1 \sqcup \cdots \sqcup X_p\right\rangle$ be the polynomial such that $\pi(g_i) = g_i'$. Then,

$$\pi(1 - \sum_i g_i \cdot f_i) = 1 - \sum_i \pi(g_i)\cdot\pi(f_i) = 1 - \sum_i g_i' \cdot f_i' = 0$$

But $1 - \sum_i g_i \cdot f_i \in \sqcup_{t=1}^{q} \mathcal{I}_t \implies 1 - \sum_i g_i \cdot f_i = \sum_{t=1}^{k} \sum_{i,j} h_{t,i,j}(x_{t,i}x_{t,j} - x_{t,j}x_{t,i})\hat{h}_{t,i,j}$ for some $\{h_{t,i,j}, \hat{h}_{t,i,j}\}$. So, for $z_i, z_{t,i,j} \in Z$, we can define

$$P(X, Z) := \sum_i g_i \cdot z_i + \sum_{t=1}^{k}\sum_{i,j} h_{t,i,j}\cdot z_{t,i,j}\cdot \hat{h}_{t,i,j},$$

to be the proof polynomial since it clearly satisfies condition (2) and (3) from Definition 1.1. □

3.2 Verifiability

From the definition of IPS, verifying a $\mathcal{C}$-IPS refutation amounts to solving the $\mathcal{C}$-PIT problem. For $\mathsf{PC_{p,q}}$- IPS, refutations can be verified in randomized polynomial time when $q = O(1)$, and in randomized quasi-polynomial time when $q = (\log s)^{O(1)}$, where s is the refutation size. This follows from the randomized identity testing algorithm for formulas (and more generally, ABPs) over $\mathbb{Q}\langle X_1, \ldots, X_q\rangle$ due to Worrell [Wor13].

Arvind, Chatterjee, and Mukhopadhyay [ACM24] recently gave a deterministic identity testing algorithm for formulas (and ABPs) over $\mathbb{Q}\langle X_1, \ldots, X_q\rangle$. This implies the following: when p is constant, for any $q \leq p$, $\mathsf{PC_{p,q}}$- IPS is a sound and complete refutation system for unsatisfiable propositional formulas (represented as partially commutative formulas), and refutations can be verified in deterministic polynomial time over $\mathbb{F} = \mathbb{Q}$.

Note that, the algorithm given in [ACM24] does not work over finite field in general as discussed in Subsection 1.3. Nevertheless, when $p = O(1)$ and $q = 2$, the algorithm of [ACM24] can be applied over an arbitrary field $\mathbb{F}$[13]. Hence $\mathsf{PC_{p,2}}$- IPS refutations admit deterministic polynomial-time verification in this case.

3.3 $\mathsf{PC_{p,q}}$- IPS simulates the *Frege* proof system

The commutative IPS and the noncommutative IPS simulate the Frege proof system [GP18, LTW18] and since $\mathsf{PC_{p,q}}$- IPS is a generalization of both, it should simulate Frege. We prove that this is indeed the case, for the sake of completeness.

We start with the standard translation between propositional formulas and algebraic formulas.

Definition 3.6. *Given a propositional formula T defined over $\wedge, \vee, \neg$, and variables $X = \{x_1, \ldots, x_N\}$, we define*

$$\mathrm{Tr}'(x_i) := x_i, \qquad \mathrm{Tr}'(\mathit{false}) := 1, \qquad \mathrm{Tr}'(\mathit{true}) := 0$$

and by induction on the size of the formula,

$$\mathrm{Tr}'(\neg T_i) := 1 - \mathrm{Tr}'(T_i); \mathrm{Tr}'(T_1 \vee T_2) := \mathrm{Tr}'(T_1)\,\mathrm{Tr}'(T_2); \mathrm{Tr}'(T_1 \wedge T_2) := 1 - (1 - \mathrm{Tr}'(T_1))(1 - \mathrm{Tr}'(T_2)).$$

If T is a propositional tautology then the polynomial computed by $\mathrm{Tr}'(T)$ is 0 over every boolean evaluation, i.e. $\widehat{\mathrm{Tr}'(T)}(\overline{a}) = 0$ for every $\overline{a} \in \{0,1\}^N$. ◊

We now prove the following theorem.

Theorem 3.7. *If a propositional tautology T has a size s Frege proof, then for every $p \geq q \geq 0$, $\mathrm{Tr}'(T)$ has a $\mathsf{PC_{p,q}}$- IPS-proof of size $\mathrm{poly}(s)$.*

[13] Personal communication with the authors of [ACM24].

In particular, we show how $\mathsf{PC}_{\mathsf{p,q}}$- IPS simulates Schoenfield's system (Definition 2.3) in polynomial size.

Lemma 3.8. ($\mathsf{PC}_{\mathsf{p,q}}$- IPS simulates Schoenfield-Frege: analogue of [LTW18, Lemma 3.3]) *Let $\mathbb{F}$ be a field. Further, let $X = X_1 \sqcup \cdots \sqcup X_p$ be partially commutative variables where, for some $0 \leq q \leq p$, $|X_i| = n$ for every $i \in [q]$ and $|X_i| = 1$ for every $i \in [q+1, p]$. Here the variables inside each X_i are noncommuting and variables from different buckets commute.*

Let π be a tree-like *Frege (Schoenfield) proof of a propositional formula T, defined over X, from assumptions $\{F_1, \ldots, F_m\}$, and let the proof-lines be $\ell_1, \ell_2, \ldots, \ell_s$. Let $\mathrm{Tr}'(\cdot)$ be the algebraic translation from Boolean formulas to $\mathbb{F}\langle X\rangle$ as in Definition 3.6. For each $i \in [N]$ define the partially commutative algebraic translation $L_i := \mathrm{Tr}'(\ell_i) \in \mathbb{F}\langle X\rangle$ and let $F := \big(\mathrm{Tr}'(F_1), \ldots, \mathrm{Tr}'(F_m)\big)$.*

*Finally, let B be the set of all **Boolean axioms**: $x(x-1)$ for every $x \in X$ and let C be the set of all **partial commutator axioms**: $xy - yx$ for every $x, y \in X_i$, $i \in [p]$. Then for every $i \in [s]$, there exists a partially commutative algebraic formula $\Phi_i(X, Y, Z, W) \in \mathbb{F}\langle X, Y, Z, W\rangle$, where $Y = (y_1, \ldots, y_m)$ and Z, W are placeholder vectors for the Boolean and commutator axioms, such that:*

1. *$\Phi_i(X, \overline{0}, \overline{0}, \overline{0}) = 0$.*
2. *$\Phi_i(X, F, B, C) = L_i$.*
3. *$|\Phi_i| \leq \left(\sum_{\ell \in A_i} |L_\ell|\right)^4$, where $A_i \subseteq [s]$ is the set of Frege proof-lines involved in deriving ℓ_i in the tree-like proof π (i.e., the indices of the sub-tree rooted at ℓ_i).*

In particular, Φ_s is a $\mathsf{PC}_{\mathsf{p,q}}$- IPS proof of $\mathrm{Tr}'(T)$ from assumptions $\{\mathrm{Tr}'(F_1), \ldots, \mathrm{Tr}'(F_m)\}$, and its size is $\mathrm{poly}(|\pi|)$.

Proof. We follow the inductive construction of Lemma 3.3 in [LTW18]. The only modification is that whenever [LTW18] uses commutator axioms to swap adjacent variables, in the partially commutative setting it suffices to use commutators *within the same bucket*, since variables from different buckets commute by definition. A complete proof is given in Subsection A.2. ☐

4 Frege Quasi-Polynomially Simulates $\mathsf{PC}_{\mathsf{k,2}}$- IPS for any constant k

In this section, we prove our main theorem (Theorem 1.2). In particular, we show that the existence of a size s $\mathsf{PC}_{\mathsf{k,2}}$- IPS refutation over $\mathbf{GF}(2)$ of an unsatisfiable CNF $\Phi = \kappa_1 \wedge \kappa_2 \wedge \ldots \wedge \kappa_m$ implies the existence of a *Frege* refutation of size $s^{(k \log s)^{O(1)}}$ for Φ.

We begin by giving a more formal version of our main theorem (Subsection 4.1) and then reduce the problem, using the Reflection Principle, to simulating the proof of certain polynomial identities in Frege given that the polynomials are being computed by arithmetic formulas (Subsection 4.2). In Subsection 4.3 we show that it is enough to consider the formulas computing these polynomials to be homogeneous and over Subsection 4.4, Subsection 4.5.2 we show that we can instead work with homogeneous ABPs. It might be useful to notice that these steps do not require $q = 2$ to quasi-polynomially simulate $\mathsf{PC}_{\mathsf{p,q}}$- IPS via Frege for any $p = \log^{O(1)} s)$.

Finally in Subsection 4.7, we give the required Frege simulation assuming the existence of certain *witness identities* (described in Subsection 4.6). We can construct witness identities that allow a quasi-polynomial Frege simulation only when $q = 2$.

4.1 Formal Statement of our Main Theorem

Let k be an arbitrarily fixed constant. Given a boolean formula Ψ over boolean variables $X = \{x_1, \ldots, x_N\}$ with internal gates $\wedge, \vee, \neg$, we define the corresponding partially commutative formula $\mathrm{Tr}(\Psi)$ in the following inductive way:

1. Partition the variables into $X = X_1 \sqcup \cdots \sqcup X_k$ where $|X_1| = \lceil (N-k+2)/2 \rceil$, $|X_2| = \lfloor (N-k+2)/2 \rfloor$ and $|X_i| = 1$ for every $i \in \{3, \ldots, k\}$. We treat X as algebraic variables now.
2. $\mathrm{Tr}(x_i) := 1 - x_i$ and $\mathrm{Tr}(\neg x_i) := x_i$ for every $x_i \in X$.
3. $\mathrm{Tr}(\mathrm{False}) := 0, \mathrm{Tr}(\mathrm{True}) := 1,\ \mathrm{Tr}(\psi_1 \vee \psi_2 \vee \ldots \vee \psi_t) := \prod_i \mathrm{Tr}(\psi_i)$. The product maintains the partial commutativity of the variables across the sets X_1, X_2.
4. Given a CNF $\Phi = \kappa_1 \wedge \kappa_2 \wedge \ldots \wedge \kappa_m$, we denote the set of partially commutative polynomials $\{\mathrm{Tr}(\kappa_1), \ldots, \mathrm{Tr}(\kappa_m)\}$ by P_Φ with $\mathrm{Tr}(\kappa_j) := P_{\Phi,j}$.

Let F be a $\mathsf{PC}_{\mathsf{k,2}}$- IPS proof over $\mathbf{GF}(2)$ that witnesses the unsatisfiability of $P_\Phi = \{P_{\Phi,1}, \ldots, P_{\Phi,m}\}$ and let $\overline{P_\Phi}$ be the set that includes the *boolean and commutator* axioms to P_Φ. Then, by Definition 1.1, F satisfies the conditions: $F(X, 0) = 0$ and $F(X, \overline{P_\Phi}) = 1$.

Note. For the remainder of this paper, in order to increase readability, we assume without loss of generality that $|X_1| = |X_2| = n$.

Definition 4.1. *Let $X = X_1 \sqcup \cdots \sqcup X_k$ be partially commutative variables where $|X_1| = n = |X_2|$ and $|X_i| = 1$ for every $i \in \{3, \ldots, k\}$. Let F be a partially commutative formula computing a polynomial in $\mathbb{F}\langle X \rangle$ over $\mathbf{GF}(2)$. We define the boolean formula $\widetilde{F}$ from F by changing the $+$ gate to $\oplus$, the $\times$ gate to $\wedge$ and treating the algebraic variables X as boolean variables.* $\Diamond$

Observe that over the boolean cube $\{0,1\}^N$, both formulas F and $\widetilde{F}$ evaluate the same value. Now define the set $\widetilde{P}_\Phi := \left\{\widetilde{P}_{\Phi,1}, \ldots, \widetilde{P}_{\Phi,m}\right\}$. From the $\mathsf{PC}_{\mathsf{k,2}}$- IPS proof F, we obtain the following two tautologies using Definition 4.1:

$$\neg\widetilde{F}(X, 0) \qquad \widetilde{F}(X, \widetilde{P}_\Phi) \tag{4.2}$$

The formal statement of Theorem 1.2 is then the following.

Theorem 4.3. *Let $\Phi = \kappa_1 \wedge \cdots \wedge \kappa_m$ be an unsatisfiable CNF and $P_\Phi = \{\mathrm{Tr}(\kappa_1), \ldots, \mathrm{Tr}(\kappa_m)\}$ be the corresponding partially commutative system that has no common boolean root. If there is a $\mathsf{PC}_{\mathsf{k,2}}$- IPS refutation for P_Φ of size s over $\mathbf{GF}(2)$ then there is a Frege proof for $\neg\Phi$ of size $s^{O(k \log s)^{O(1)}}$.*

4.2 The Reflection Principle

In order to prove Theorem 4.3, we use the reflection principle.

Lemma 4.4. ([LTW18, Lemma 4.3]) *There is a polynomial size* Frege *proof of* $\neg\Phi$ *assuming* $\neg\widetilde{F}(X,0)$ *and* $\widetilde{F}(X,\widetilde{P}_\Phi)$ *has a polynomial size* Frege *proof.*

Even though the work of Li, Tzameret and Wang [LTW18] prove this when P_Φ is a noncommutative formula, the same proof applies here as well, since there is no distinction between the two settings once we have written a proof of $\neg\widetilde{F}(X,0)$ and $\widetilde{F}(X,\widetilde{P}_\Phi)$ within Frege. Therefore, our goal is to obtain a *Frege* Proof of Equation 4.2.

In particular, using Lemma 4.4, Theorem 4.3 is a direct corollary of the following theorem.

Theorem 4.5. *Let* $X = X_1 \sqcup \cdots \sqcup X_k$ *be partially commutative variables where* $|X_1| = n = |X_2|$ *and* $|X_i| = 1$ *for every* $i \in \{3,\dots,k\}$*. Further, let* F *be a partially commutative formula of size* s *over* $\mathbf{GF}(2)$*, with variable set* X*, that computes the* 0 *polynomial. For any* k*, there exists a constant* $c > 0$ *such that the boolean formula* $\neg\widetilde{F}(X)$ *has a* Frege *proof of size* $s^{(k\log s)^c}$.

The remainder of the section is dedicated to the proof of Theorem 4.5.

Following the notation of [LTW18], for Boolean formulas $\widetilde{F}$ and $\widetilde{G}$, we write $\widetilde{F} \vdash \widetilde{G}$ to denote that there exists a Frege proof of $\widetilde{G}$ from the axiom $\widetilde{F}$ of size polynomial in $|\widetilde{G}|$. That is, $\widetilde{G}$ can be derived from $\widetilde{F}$ by purely syntactic manipulations using standard Frege inference rules and Boolean equivalences, including associativity and distributivity, as well as identities such as False $\oplus\, \widetilde{F} \equiv \widetilde{F}$, True $\vee\, \widetilde{F} \equiv$ True, True $\oplus\, \widetilde{F} \equiv \neg\widetilde{F}$, False $\wedge\, \widetilde{F} \equiv$ False, True $\wedge\, \widetilde{F} \equiv \widetilde{F}$, and similar Boolean equivalences.

Similarly, for two vectors of formulas $\overline{F}$ and $\overline{G}$, we write $\overline{F} \vdash \overline{G}$ to denote that, for each index, the corresponding entry of $\overline{G}$ can be derived from the corresponding entry of $\overline{F}$ by a Frege proof of size polynomial in the size of that entry.

4.3 Homogenization of Partially Commutative Formulas

Let F be a partially commutative formula over $\mathbf{GF}(2)$ with the variable set $X = X_1 \sqcup \ldots \sqcup X_k$.

Given a partially commutative monomial $m = \prod_{i\in[k]} m_i$ with m_i being noncommutative in X_i for every $i \in [k]$, define the degree signature of m to be $d(m) := (d_1,\dots,d_k)$ if the degree of m_i is d_i for every $i \in [k]$. A partially commutative polynomial f is homogeneous if every monomial of f has the same degree signature. Thus any partially commutative polynomial computed by a formula F, say $\hat{F}$, can be uniquely written as

$$\hat{F} = \sum_{\substack{d_3,\dots,d_k \le 1 \\ \sum d_i \le d}} \hat{F}_{(d_1,\dots,d_k)}$$

where the degree of $\hat{F}$ is d and $\hat{F}_{(d_1,\dots,d_k)}$ is the homogeneous component of $\hat{F}$ with degree signature $(d_1,\dots,d_k)$.

Lemma 4.6. *Let F be a partially commutative formula of size s and depth $O(\log s)$ computing a polynomial in $\mathbb{F}\langle X_1 \sqcup \ldots \sqcup X_k\rangle$ where $\mathbb{F} = \mathbf{GF}(2)$. Moreover, let $F_{(d_1,\dots,d_k)}$ be the homogeneous component of the polynomial computed by F with degree signature $(d_1,\dots,d_k)$. Then there exists a $s^{O(\log s)}$ size Frege proof of*

$$\widetilde{F}(X) \longleftrightarrow \bigoplus_{(d_1,\dots,d_k)} \widetilde{F}_{(d_1,\dots,d_k)}. \tag{4.7}$$

This is analogous to [LTW18, Lemma 4.6] and the proof follows along the same lines. A full proof is present in Subsection B.1, when $k = 2$, for the sake of completeness.

To show within Frege that the homogeneous partially commutative formula F is identically 0, we first convert F into a layered homogeneous ABP A_F that computes the same polynomial. Following [LTW18], we establish equations that witness A_F computes the zero polynomial.

Since Frege operates on formulas rather than ABPs, we translate this argument back to formulas by associating each partial ABP in A_F with a unique induced formula in F that computes the same polynomial. Let $\widetilde{F}$ be the Boolean formula obtained from the homogeneous partially commutative formula F over $\mathbf{GF}(2)$ via Definition 4.1.

4.4 Translation Between Formulas and ABPs

We first eliminate Boolean constants from $\widetilde{F}$ using the following lemma from [LTW18]. That is, we transform $\widetilde{F}$ into a constant-free formula by applying standard Boolean equivalences. Although the lemma is stated for noncommutative formulas in [LTW18], it applies verbatim to partially commutative formulas.

Lemma 4.8. ([LTW18, Lemma 4.10]) *Let F be a non-constant partially commutative homogeneous formula over $\mathbf{GF}(2)$ computing the zero polynomial. Then there exists a constant-free partially commutative formula F' of size $\mathrm{poly}(|F|)$ computing the same zero polynomial, such that $\widetilde{F} \vdash \widetilde{F}'$.*

Next, given a partially commutative homogeneous constant-free formula F, we construct the corresponding homogeneous partially commutative ABP A_F following the construction of [Nis91]. The construction proceeds inductively as follows.

- If $u = v + w$ is a node of F and the ABPs A_v and A_w have already been constructed, then A_u is obtained by taking the parallel composition of A_v and A_w, merging their source nodes into a single source and their sink nodes into a single sink. The resulting ABP computes the sum of the polynomials computed by A_v and A_w.
- If $u = v \times w$, then A_u is obtained by sequentially composing A_v and A_w: the sink of A_v is identified with the source of A_w, and the source of A_v and the sink of A_w serve as the source and sink of A_u, respectively.

Tracking computation in F corresponding to computation done A_F

The arguments are the same as that of [LTW18], but we recall them here as well for completeness. Firstly, we need the following two notions defined in their work.

Definition 4.9 (Induced formula [LTW18])**.** *Given a formula F, let F' be a sub-formula of F. Assume $\{u_1, \ldots, u_s\}$ are nodes in F' and $\{\alpha_1, \ldots, \alpha_s\}$ are elements of $\mathbb{F}$. Then $F'(u_1 = \alpha_1, \ldots, u_k = \alpha_s)$ is called the induced formula from F.* ♢

Definition 4.10 (v-part [LTW18])**.** *Let F be a formula and A_F be the corresponding ABP for F with source and sink nodes s, t, respectively. For any node $v \in A_F$, the v-part of F is an induced formula computing the polynomial $A_F[v, t]$. Although a formula may admit multiple v-parts in general, for homogeneous formulas and their corresponding homogeneous ABPs, we define a canonical v-part for every node $v \in A_F$, denoted by F^*_v.* ♢

We now consider a map $g : \{\text{nodes of } A_F\} \to \{\text{nodes of } F\} \cup \{\emptyset\}$ where $g(u) = v$ if the polynomial computed by F at v is the polynomial computed between u and t in A_F.

The definition is by induction, and can be found in Subsection B.2.

Definition of the function D. We then define a function D that takes as input a sub-formula F_u and a node $v \in F$, and outputs an induced sub-formula of F. Although D is defined for all nodes of F, we only apply it to nodes in the image of g. Note that constants introduced in D do not appear in the original constant-free formula F.

For every node $u \in F$, define $D(F_u, u) := F_u$. If u is not a node of F, then $D(F, u)$ is undefined. The function D is defined inductively from the leaves upwards.

Let G, H be homogeneous sub-formulas of F. For a $+$-node,

$$D(G + H, u) = \begin{cases} D(G, u) + 0, & u \in G, \\ 0 + D(H, u), & u \in H. \end{cases} \tag{4.11}$$

For a $\times$-node,

$$D(G \times H, u) = \begin{cases} D(G, u) \times H, & u \in G, \\ 1 \times D(H, u), & u \in H. \end{cases} \tag{4.12}$$

Definition of the v-part. Finally, for a node $v \in A_F$, define $F^*_v := D(F, g(v))$. Let r be the root of F. Since $g(s) = r$, where s is the source node of A_F, we have $F^*_s = D(F, r) = F$, which computes the same polynomial $\hat{F}$ as the original homogeneous formula F.

Following the notation that has been defined, [LTW18] shows,

$$F^*_v \vdash \sum_{u\colon u \text{ has an incoming edge from } v} A_F[v, u] \cdot F^*_u.$$

That is, we can prove the following tautology within Frege:

$$F_v^* \quad \leftrightarrow \sum_{u\colon u \text{ has an incoming edge from } v} A_F[v,u] \cdot F_u^*.$$

Thus it is enough to argue about the partially commutative ABP A_F, which is what we do now.

4.5 Refined Translation Between Formulas and ABPs

To handle partial commutativity, we represent polynomials as sums of monomials over a fixed bucket, with coefficients given by polynomials over the remaining buckets. We need this viewpoint to be incorporated directly into the underlying computational model (ABPs or formulas). So, we first transform the homogeneous partially commutative ABP A_F over the variable set $X_{[k]}$ into an equivalent ABP A'_F that computes the same polynomial such that each edge label of A'_F is a homogeneous linear form in the X_1 variables with noncommutative polynomials in the remaining variables as coefficients. It is not hard to see that each of these coefficients are computable by a noncommutative ABP of polynomial size.

Lemma 4.13. [ACM24, Lemma 12] *Let $f \in \mathbb{F}\langle X_1 \sqcup \ldots \sqcup X_k \rangle$ be a partially commutative polynomial of degree d computed by an ABP of size s. Then for any $1 \le i \le k$, we can efficiently homogenize the ABP over the variable set X_i, and the coefficients $\in \mathbb{F}\langle X_1 \sqcup \ldots X_{i-1} \sqcup X_{i+1} \sqcup \cdots \sqcup X_k \rangle$ can also be computed by partially commutative ABPs of size $O(sd)$.*

We now construct, in Frege, these ABPs that compute the coefficients and show that the computation can be tracked in the original formula. We then give a quasi-polynomial upper bound on the size of these induced formulas, in order to ensure that we have a similar bound on the final Frege simulation. Finally, we observe that these induced formulas satisfy certain identities. These will later help us to construct identity witnesses which will allow us to give the Frege simulation.

4.5.1 Refined ABP Construction

Recall that F is a partially commutative homogeneous constant-free formula over the variable set $X = X_1 \sqcup \ldots \sqcup X_k$, and A_F is the corresponding homogeneous partially commutative ABP of size s and depth d computing the same polynomial Thus every edge label in A_F from $u_{i,j}$ to $u_{i+1,\ell}$ is of the form

$$L_{j,\ell} = \sum_{a=1}^{k} L_{j,\ell}^{(a)}(X_a),$$

where each $L_{j,\ell}^{(i)}$ is a homogeneous noncommutative linear forms. Without loss of generality, let us assume that we want to homogenize with respect to X_1.

Construction of A'_F. We construct the ABP A'_F exactly as described earlier. For each node $u_{i,j}$ of A_F, we introduce nodes

$$u_{i,j}^{(0)}, u_{i,j}^{(1)}, \dots, u_{i,j}^{(d)}$$

in A'_F, where the superscript tracks the accumulated X_1-degree.

Base Case: Let us assume that $i = 1$. If $s \xrightarrow{L_i} u_{1,i}$ in A_F with $L_i = \sum_{a=1}^{k} L_i^{(a)}(X_a)$, then A'_F contains

$$s \xrightarrow{L_i^{(1)}} u_{1,i}^{(1)}; \qquad s \xrightarrow{L_i^{(a)}} u_{1,i}^{(0)} \text{ for every } a \in \{2, \dots, k\}.$$

Inductive Step: If $u_{i,j} \xrightarrow{L_{j,\ell}} u_{i+1,\ell}$ in A_F, then for all $p \in \{0, \dots, d\}$ the ABP A'_F contains

$$u_{i,j}^{(p)} \xrightarrow{L_{j,\ell}^{(1)}} u_{i+1,\ell}^{(p+1)} \quad (p < d); \qquad\qquad u_{i,j}^{(p)} \xrightarrow{L_{j,\ell}^{(a)}} u_{i+1,\ell}^{(p)} \text{ for every } a \in \{2, \dots, k\}.$$

For a node $u_{i,j}$ of A_F, we write $(A_F[s, u_{i,j}])^{(p)}$ to denote the X_1-degree-p homogeneous component of the polynomial computed by $A_F[s, u_{i,j}]$. The following claim is now easy to verify.

Claim 4.14. *For every node $u_{i,j}$ of A_F and every $p \in \{0, \dots, d\}$,*

$$A'_F[s, u_{i,j}^{(p)}] \;=\; \big(A_F[s, u_{i,j}]\big)^{(p)},$$

where $(\cdot)^{(p)}$ denotes the X_1-degree-p homogeneous component.

By defining the sink nodes of A'_F to be $t^{(0)}, \dots, t^{(d)}$, we have that A'_F computes the same polynomial as A_F and has size $O(sd)$.

4.5.2 Refined ABP Tracking

We again assume without loss of generality that homogenization has been done with respect to X_1 in the ABP A'_F. Recall the map $g : \{\text{nodes of } A_F\} \to \{\text{nodes of } F\} \cup \{\emptyset\}$. We define a refined map

$$g' : \{\text{nodes of } A'_F\} \to \{\text{nodes of } F\} \cup \{\emptyset\}$$

by $g'(u_{i,j}^{(p)}) := g(u_{i,j})$. That is, g' ignores the superscript (p) and records only the corresponding location in the formula F.

Definition of the refined operator D'. Let g' be the tracking map from the nodes of A'_F to the nodes of F. For a node u of F and a parameter $p \in \{0, \dots, d\}$, we define an induced formula

$$F_u^{*(p)} := D'(F, u, p),$$

where D' extracts the X_1-degree-p component of the induced polynomial corresponding to u.

For a node $v = u_{i,j}^{(p)} \in A'_F$, the induced formula associated with v is $F_v^* := F_{g'(v)}^{*(p)}$. The operator D' is defined inductively on the structure of the formula F and the degree parameter p.

Leaves. If $F_u = x$, then

$$D'(x,u,p) = \begin{cases} x, & x \in X_1 \text{ and } p = 1, \\ x, & x \in X_{[k]} \setminus X_1 \text{ and } p = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Addition. If $F_u = G + H$, then

$$D'(G+H,u,p) = \begin{cases} D'(G,u,p) + 0, & u \in G, \\ 0 + D'(H,u,p), & u \in H. \end{cases}$$

Multiplication. If $F_u = G \cdot H$, then

$$D'(G \cdot H,u,p) = \begin{cases} \sum_{q=0}^{p} D'(G,u,q) \cdot F_{\text{root}(H)}^{*(p-q)}, & u \in G, \\ D'(H,u,p), & u \in H. \end{cases}$$

Here $F_{\text{root}(H)}^{*(p-q)}$ denotes the induced formula obtained by applying D' to the root of H with parameter $p-q$.

The above recursion is well-founded since it is defined on strict sub-formulas of F and proceeds lexicographically on the formula depth and the degree parameter. Further, for every node $v = u^{(p)} \in A'_F$, the induced formula F_v^* computes exactly the polynomial computed by the partial ABP $A'_F[v,t]$, which is the X_1-degree-p homogeneous component of the polynomial computed by $A_F[g'(v),t]$. We formalise this below.

Lemma 4.15. (Well-definedness of the refined operator D') *Let F be a partially commutative homogeneous constant-free formula over the variable set $X = X_1 \sqcup \ldots \sqcup X_k$, and let d be the number of layers of the ABP A_F (hence the maximum possible X_1-degree in the constructions). Let D' be defined inductively on sub-formulas of F and on a parameter $p \in \{0,1,\ldots,d\}$ as in the construction of Subsection 4.5.2.*

Then for every sub-formula F_u of F, every node u of F_u, and every $p \in \{0,\ldots,d\}$, the expression $D'(F_u,u,p)$ is well-defined. Moreover, for every triple (F_u,u,p) for which $D'(F_u,u,p)$ is defined, the output is an induced formula of F_u (obtained from F_u by substituting some sub-formulas by constants $0,1$ and by replacing some sub-formulas by their degree-refined induced parts).

The proof is by induction and can be found in Subsection B.3

4.5.3 Size of the degree-refined induced formulas.

We now give a quasi-polynomial upper bound on the size of these induced formulas. This will help us ensure that we have a similar bound on the final Frege simulation.

Lemma 4.16. *Let F be a partially commutative homogeneous constant-free formula obtained from an original size-s formula by balancing and homogenization, such that*

$$|F| \;=\; s^{O(\log s)} \qquad \textit{and} \qquad \mathsf{depth}(F) = O(\log |F|).$$

Let A_F be the homogeneous partially commutative ABP corresponding to F, and let A'_F be the refined ABP constructed above, with depth (number of layers) d. Then:

1. $|A'_F| \;=\; s^{O(\log s)}$.
2. *For every node u of F and every $p \in \{0, \ldots, d\}$, the induced formula $F_u^{*(p)}$ has size at most $s^{O(\log^3 s)}$. Consequently, for every node $v \in A'_F$, the induced formula F_v^* also has size at most $|F_v^*| \;\leq\; s^{O(\log^3 s)}$.*

Proof. **Size of A'_F.** By the standard construction due to Nisan [Nis91], the ABP A_F has size $|A_F| = O(|F|)$ and depth $d \leq |A_F| \leq O(|F|)$. The refinement from A_F to A'_F replaces each node by $(d+1)$ copies and replaces each edge by at most $2(d+1)$ edges, hence

$$|A'_F| \;=\; O(|A_F| \cdot d) \;\leq\; O(|F| \cdot |F|) \;=\; O(|F|^2).$$

Since $|F| = s^{O(\log s)}$, we obtain $|A'_F| = s^{O(\log s)}$.

Size of $F_u^{*(p)}$. Let $\text{prod-depth}(F_u)$ denote the number of $\times$-gates on a longest root-to-leaf path in the subformula F_u. By the definition of the refined operator D', the only source of blow-up is at a multiplication gate, where for fixed p we introduce a sum of at most $(p+1) \leq (d+1)$ product terms. A standard structural induction on F_u (as in the proof of Lemma 4.15) yields the bound

$$|F_u^{*(p)}| \;\leq\; (d+1)^{\text{prod-depth}(F_u)} \cdot |F_u|. \tag{4.17}$$

Since $\text{prod-depth}(F_u) \leq \mathsf{depth}(F)$ for every u, we have

$$|F_u^{*(p)}| \;\leq\; (d+1)^{\mathsf{depth}(F)} \cdot |F| \;\leq\; (|F|+1)^{O(\log |F|)} \cdot |F| \;=\; |F|^{O(\log |F|)}.$$

Now $|F| = s^{O(\log s)}$ implies $\log |F| = O(\log^2 s)$, and therefore

$$|F|^{O(\log |F|)} = \left(s^{O(\log s)}\right)^{O(\log^2 s)} = s^{O(\log^3 s)}.$$

This proves $|F_u^{*(p)}| \leq s^{O(\log^3 s)}$ for every u and p.

Finally, for a node $v = u_{i,j}^{(p)} \in A'_F$, by definition $F_v^* = F_{g'(v)}^{*(p)}$, and hence the same bound holds for $|F_v^*|$. ∎

4.5.4 Local ABP identities in A'_F.

Recall that for the original ABP A_F, the induced formulas satisfy the identity

$$F_v^* \;=\; \sum_{u\colon (v,u)\text{ is an edge in }A_F} A_F[v,u]\cdot F_u^*.$$

We now show that an analogous identity holds for the refined ABP A'_F. These will help us to construct identity witnesses in the next sub-section.

We fix a sink node t of A'_F (e.g. $t = t^{(d)}$), and for every node v we write $A'_F[v,t]$ for the polynomial computed by the partial ABP from v to t.

Lemma 4.18 (Local identity for A'_F). *For every node $v \in A'_F$, the induced formula F_v^* satisfies*

$$F_v^* \;=\; \sum_{u\colon (v,u)\textit{ is an edge in }A'_F} A'_F[v,u]\cdot F_u^*.$$

Proof. By correctness, the induced formula F_v^* computes exactly the polynomial computed by the partial ABP $A'_F[v,t]$. Every path from v to t in A'_F begins by traversing a unique *outgoing* edge (v,u), and the contribution of all such paths is given by $A'_F[v,u]\cdot A'_F[u,t]$. Summing over all outgoing edges yields

$$A'_F[v,t] = \sum_{u\colon (v,u)\text{ is an edge in }A'_F} A'_F[v,u]\cdot A'_F[u,t].$$

Replacing $A'_F[u,t]$ by the polynomial computed by F_u^* gives the desired identity. ☐

Explicit form of the identity. If $v = u_{i,j}^{(p)}$ is a node of A'_F, then its outgoing edges are of the form

$$u_{i,j}^{(p)} \xrightarrow{L_{j,\ell}^{(1)}} u_{i+1,\ell}^{(p+1)} \quad (p<d); \qquad\qquad u_{i,j}^{(p)} \xrightarrow{L_{j,\ell}^{(a)}} u_{i+1,\ell}^{(p)} \text{ for every } a\in\{2,\ldots,k\}\,.$$

Accordingly, the identity of Lemma 4.18 can be written as

$$F^*_{u_{i,j}^{(p)}} = \sum_{\ell}\left(L_{j,\ell}^{(1)}\cdot F^*_{u_{i+1,\ell}^{(p+1)}} + \sum_{a=2}^{k} L_{j,\ell}^{(a)}\cdot F^*_{u_{i+1,\ell}^{(p)}}\right).$$

This identity is the degree-refined analogue of the corresponding identity for A_F, and it follows directly from the construction of A'_F and the definition of the refined operator D'.

Remark (Role of the local identities). Similar to [LTW18], for any node $v \in A'_F$, we can prove the following within Frege,

$$F_v^* \quad \vdash \sum_{u:\ u \text{ has an incoming edge from } v} A'_F[v,u] \cdot F_u^*. \tag{4.19}$$

Although the degree-refined induced formulas $F_v^* = D'(F, g'(v), p)$ may incur a quasi-polynomial size blow-up due to the degree-convolution at $\times$-gates, the local identity of Lemma 4.18 allows us to reason about A'_F layer-by-layer: each line

$$F_v^* = \sum_{(v,u)\in E(A'_F)} A'_F[v,u] \cdot F_u^*$$

is a short Frege-derivable equivalence, and therefore the Frege proof length is bounded by a polynomial in $|A'_F|$ and in $\max_v |F_v^*|$. Moreover, balancing the sums in Lemma 4.18 yields $\mathsf{depth}(F_v^*) \le O(\log^2 |F|)$, since each $\times$-gate introduces a sum of at most $(d+1)$ terms (which can be balanced to depth $O(\log(d+1))$) and $\mathsf{prod\text{-}depth}(F) \le \mathsf{depth}(F) = O(\log |F|)$.

4.6 Identity Witnesses

We now define the identity witnesses that will help us give the Frege simulation.

We have been given a partially commutative X_1-homogeneous ABP A that computes a polynomial $\hat{A} \in \mathbb{F}\langle X \rangle$. Here $X = X_1 \sqcup \ldots \sqcup X_k$, where $|X_1| = n = |X_2|$ and $|X_i| = 1$ for every $i \in \{3, \ldots, k\}$. Note that we can assume A is X_1-homogeneous because of our discussion in Subsection 4.5. Further observe that $\mathbb{F}\langle X \rangle$ is contained in $\mathbb{F}[y_1, \ldots, y_{k-2}]\langle X_1 \sqcup X_2 \rangle$ in our setting, if we assume $X_i = \{y_{i-2}\}$ for every $i \in \{3, \ldots, k\}$. So we can assume that A is a partially commutative ABP computing a polynomial in $\mathbb{F}[\overline{y}]\langle X_1 \sqcup X_2 \rangle$.

For the remainder the section, we will be using this view.

4.6.1 ABP identity witnesses over $\mathbb{F}[\overline{y}]$

Let $X = X_1 \sqcup X_2$ be partially commutative variables and $\bar{y} := \{y_1, \ldots, y_{k-2}\}$

Assume that A is a partially commutative ABP of size s computing a polynomial in $\mathbb{F}[\overline{y}]\langle X_1 \sqcup X_2 \rangle$ of depth d, which is homogeneous over X_1 and computes the identically zero polynomial. Let the number of layers in A be d. So $\deg_{X_1}(\hat{A}) \le d$. Further, let the i-th layer of A have w_i nodes, say $\{u_{i,1}, u_{i,2}, \ldots, u_{i,w_i}\}$. Also let s, t be the source and sink nodes, respectively.

Define $w = \max_{i\in[d]} \{w_i\}$ and for $i \in [d]$, define the column vector

$$\overline{A}_i = \left(A[u_{d-i,1}, t],\, A[u_{d-i,2}, t],\, \ldots,\, A[u_{d-i,w_{d-i}}, t]\right)^\top \in (\mathbb{F}[\overline{y}]\langle X_1 \sqcup X_2 \rangle)^{w_{d-i}}.$$

where, for every $j \in [w_{d-i}]$, $A[u_{d-i,j}, t]$ is the partial ABP with the source node $u_{d-i,j}$ in the $(d-i)$-th layer and the sink being t. Observe that each ABP in $\overline{A}_i$ is computing a partially commutative polynomial that is homogeneous over X_1, and the X_1-degree is i.

In particular, $\overline{A}_d$ is the polynomial computed by A and hence $\overline{A}_d = 0$.

Homogeneity and coefficient matrices. Since A is homogeneous over X_1, each layer transition can be written as

$$\overline{A}_{d-r} \;=\; \sum_{j=1}^{n} x_{1,j} \cdot (M_j^{(r)} \cdot \overline{A}_{d-r-1}), \qquad r = 0, 1, \ldots, d-1, \tag{4.20}$$

where $M_j^{(r)} \in \mathbb{F}[\overline{y}] \langle X_2 \rangle^{w_r \times w_{r+1}}$ are the coefficient matrices (depending on r), and multiplication is matrix-vector multiplication over $\mathbb{F}[\overline{y}] \langle X_2 \rangle$.

Lemma 4.21. (ABP identity witnesses over $\mathbb{F}[\overline{y}]$) *Let A, $\overline{A_i}$ for every $i \in [d]$, be as described in Subsection 4.6.1. Further suppose that A computes the zero polynomial. Then there exist matrices $\lambda_0'', \lambda_1'', \ldots, \lambda_{d-1}''$ and $T_1'', T_2'', \ldots, T_{d-1}''$ with*

$$\lambda_r'' \in \mathbb{F}[\bar{y}] \langle X_2 \rangle^{w_r \times w_r}, \qquad T_{r+1}'' \in \mathbb{F}[\bar{y}] \langle X_1 \sqcup X_2 \rangle^{w_r \times w_{r+1}},$$

such that the following identities hold.

$$\lambda_{d-i}'' \cdot \overline{A}_i = \overline{0} \qquad \forall\, i = 1, 2, \ldots, d. \tag{4.22}$$

$$\lambda_{d-i}'' \cdot \overline{A}_i = T_{d-i+1}'' \cdot \lambda_{d-i+1}'' \cdot \overline{A}_{i-1} \qquad \forall\, i = 2, 3, \ldots, d. \tag{4.23}$$

Here, λ_0' is a polynomial of degree $\mathrm{poly}(s, d)$, and the degree of the coefficients of each entry in the witness matrices λ_i'', and T_j'' is at most $\mathrm{poly}(s, d)$. Moreover, the entries of λ_i'', T_j'' are computable by $\mathrm{poly}(s)$-size ABP.

Even though the witness equations look similar to those in the work of Li, Tzameret and Wang [LTW18], there is a major difference between the proofs of their existence. We explain this in detail in Section 5. In this section, we first complete the proof of Theorem 4.5 assuming Lemma 4.21.

4.7 The Frege Simulation

Let us first recall Theorem 4.5.

Theorem 4.5. *Let $X = X_1 \sqcup \cdots \sqcup X_k$ be partially commutative variables where $|X_1| = n = |X_2|$ and $|X_i| = 1$ for every $i \in \{3, \ldots, k\}$. Further, let F be a partially commutative formula of size s over $\mathbf{GF}(2)$, with variable set X, that computes the 0 polynomial. For any k, there exists a constant $c > 0$ such that the boolean formula $\neg \widetilde{F}(X)$ has a* Frege *proof of size $s^{(k \log s)^c}$.*

Proof. Let F be a partially commutative homogeneous formula over X computing the zero polynomial over $\mathbf{GF}(2)$ and let $A := A_F'$ be the corresponding partially commutative ABP that is homogeneous over X_1 and computes the same polynomial. Note that this can be assumed because of our discussions in Subsection 4.4 and Subsection 4.5.

Assume A has depth d and layer widths $w_0, \ldots, w_d$. We also think of the polynomial being computed by A_F to be in $\mathbf{GF}(2)[y_1, \ldots, y_{k-2}] \langle X_1 \sqcup X_2 \rangle$ as discussed in Subsection 4.6. Here we are assuming that $X_i = \{y_{i-2}\}$ for every $i \in \{3, \ldots, k\}$.

For $i \in [d]$, recall the column vector of layer-to-sink polynomials

$$\overline{A}_i \;=\; \left(A[u_{d-i,1}, t],\, A[u_{d-i,2}, t],\, \ldots,\, A[u_{d-i,w_{d-i}}, t]\right)^{\top} \in \left(\mathbf{GF}(2)[\overline{y}] \langle X_1 \sqcup X_2 \rangle\right)^{w_{d-i}}.$$

In particular, $\overline{A}_d$ is a scalar and is equal to the polynomial computed by A, and since $F \equiv 0$, we have $\overline{A}_d = 0$. By Lemma 4.21, there exist witness matrices $\lambda''_0, \lambda''_1, \ldots, \lambda''_{d-1}$ and $T''_1, T''_2, \ldots, T''_{d-1}$ with

$$\lambda''_r \in \mathbf{GF}(2)[\overline{y}] \langle X_2 \rangle^{w_r \times w_r}, \qquad T''_{r+1} \in \mathbf{GF}(2)[\overline{y}] \langle X_1 \sqcup X_2 \rangle^{w_r \times w_{r+1}},$$

such that the following algebraic identities hold.

$$\lambda''_{d-i} \cdot \overline{A}_i = \overline{0} \qquad \forall\, i = 1, 2, \ldots, d, \tag{4.24}$$

$$\lambda''_{d-i} \cdot \overline{A}_i = T''_{d-i+1} \cdot \lambda''_{d-i+1} \cdot \overline{A}_{i-1} \qquad \forall\, i = 2, 3, \ldots, d. \tag{4.25}$$

Further, the $\overline{y}$-degree of each entry in the witness matrices λ''_i and T''_j is at most $\mathrm{poly}(s, d)$.

Booleanization. Following Definition 4.1, for every algebraic object P over $\mathbf{GF}(2)$ we write $\widetilde{P}$ for its Booleanization (i.e. $+$ becomes $\oplus$ and $\cdot$ becomes $\wedge$). In particular, $\widetilde{\overline{A}_i}$ is the vector of Boolean formulas obtained by Booleanizing each entry of $\overline{A}_i$. Over $\mathbf{GF}(2)$, an algebraic identity $P = 0$ translates into the Boolean tautology $\neg\widetilde{P}$.

From Equation 4.24, we obtain that for every $i \in [d]$, the Boolean tautologies

$$\bigwedge_{p \in [w_{d-i}]} \left(\neg \bigoplus_{t \in [w_{d-i}]} \left(\widetilde{\lambda''_{d-i}}[p, t] \;\wedge\; \widetilde{\overline{A}_i}[t] \right) \right). \tag{4.26}$$

Similarly, Equation 4.25 yields, for every $i \in \{2, \ldots, d\}$ and every $p \in [w_{d-i}]$, the Boolean equivalences

$$\bigoplus_{t \in [w_{d-i}]} \left(\widetilde{\lambda''_{d-i}}[p, t] \;\wedge\; \widetilde{\overline{A}_i}[t] \right) \longleftrightarrow \bigoplus_{t \in [w_{d-i+1}]} \left(\widetilde{T''_{d-i+1}}[p, t] \;\wedge\; \left(\bigoplus_{q \in [w_{d-i+1}]} \widetilde{\lambda''_{d-i+1}}[t, q] \;\wedge\; \widetilde{\overline{A}_{i-1}}[q] \right) \right). \tag{4.27}$$

Key Frege reflection step. Although Equation 4.27 is *semantically true* as the Booleanization of the algebraic identity Equation 4.25, we must also argue that it has a *short Frege proof*. The subtlety is that Equation 4.25 is an identity in the *partially commutative* algebra $\mathbf{GF}(2)[\overline{y}]\langle X_1 \sqcup X_2\rangle$, while the Frege simulation in [LTW18] is stated for *noncommutative* ABPs over $\mathbf{GF}(2)$. Note that the Frege simulation of [LTW18] can be extended to work for noncommutative ABPs over $\mathbf{GF}(2)[\overline{y}]$ as long as the $\overline{y}$-degree is polynomially bounded on each edge. This is because the change simply translates to replacing edges by a "small gadget" in the formulas on each line of the Frege proof..

Since we have a bound on the $\overline{y}$-degree of each entry of the witness matrices by Lemma 4.21, it is enough to reduce Equation 4.25 to a *polynomial-sized* collection of noncommutative ABP identities over $\mathbf{GF}(2)[\overline{y}]\langle X_2\rangle$ by extracting coefficients only with respect to the *leftmost linear* X_1-variables (not X_1-monomials). Formally, we have the following claim.

Claim 4.28. (Frege derives the local witness equivalences) *For every $i \in \{2, \ldots, d\}$ and every $p \in [w_{d-i}]$, the equivalence Equation 4.27 has a Frege proof of size quasi-polynomial in $|A|$ (equivalently, quasi-polynomial in $|F|$).*

We first finish the proof of Theorem 4.5 assuming it.

Base case: $i = 1$. We would like to write a Frege proof for the following equation.

$$\bigwedge_{p\in[w_{d-1}]} \left(\neg \bigoplus_{t\in[w_{d-1}]} \left(\widetilde{\lambda''_{d-1}}[p,t] \;\wedge\; \widetilde{\overline{A}_1}[t] \right) \right).$$

Fix $p \in [w_{d-1}]$. Since A is homogeneous over X_1, each entry of $\overline{A}_1$ is a homogeneous linear form in the X_1-variables with coefficients in $\mathbf{GF}(2)[\overline{y}]\langle X_2\rangle$: for each $t \in [w_{d-1}]$,

$$\overline{A}_1[t] \;=\; \sum_{j=1}^{n} x_{1,j} \cdot f_{t,j}, \qquad f_{t,j} \in \mathbf{GF}(2)[\overline{y}]\langle X_2\rangle .$$

Therefore, the identity $\sum_t \lambda''_{d-1}[p,t] \cdot \overline{A}_1[t] = 0$ is equivalent to Equation 4.29, namely:

$$\sum_{t=1}^{w_{d-1}} \lambda''_{d-1}[p,t] \cdot \overline{A}_1[t] = 0 \iff \forall j \in [n],\ \sum_{t=1}^{w_{d-1}} \lambda''_{d-1}[p,t] \cdot f_{t,j} = 0. \tag{4.29}$$

Each polynomial $\sum_t \lambda''_{d-1}[p,t] \cdot f_{t,j}$ is computed by a noncommutative ABP (over $\mathbf{GF}(2)$ with variable set X_2) of size $\mathrm{poly}(|A|)$, and hence by [LTW18] each corresponding Boolean tautology

$$\neg \bigoplus_{t\in[w_{d-1}]} \left(\widetilde{\lambda''_{d-1}}[p,t] \;\wedge\; \widetilde{f}_{t,j} \right)$$

has a quasi-polynomial size Frege proof. Thus Frege derives Equation 4.26 for $i = 1$.

Inductive step: Let us assume that there is a Frege derivation of Equation 4.26 for $i-1$. Fix $p \in [w_{d-i}]$. Then, for every $t \in [w_{d-i+1}]$, we have

$$\neg \bigoplus_{q \in [w_{d-i+1}]} \left(\widetilde{\lambda''_{d-i+1}}[t,q] \wedge \widetilde{\overline{A}_{i-1}}[q] \right),$$

so the right-hand side of Equation 4.27 is false. By Claim 4.28, Frege has a short proof of the equivalence Equation 4.27, and therefore Frege derives that the left-hand side of Equation 4.27 is also false. That is,

$$\neg \bigoplus_{t \in [w_{d-i}]} \left(\widetilde{\lambda''_{d-i}}[p,t] \wedge \widetilde{\overline{A}_i}[t] \right).$$

Since p was arbitrary, this yields Equation 4.26 for i.

Iterating through $i = 2, 3, \ldots, d$, yields Equation 4.26 for $i = d$. Since λ'_0 is not identically 0 over $\mathbf{GF}(2)$ from the construction given in Equation 5.41, and $w_0 = 1$, Equation 4.26 for $i = d$ simplifies to $\neg\widetilde{\overline{A}_d}$, i.e. $\neg\widetilde{A}$. Finally, since $A = A'_F$ computes the same polynomial as F, we obtain a Frege proof of $\neg\widetilde{F}$.

Remark (induced formulas viewpoint). Note that, if $v = u_{d-i,t}$ is the t-th node in layer $(d-i)$, then $\overline{A}_i[t] = A[v,t]$. By the definition of the induced-formula tracking in A'_F, the formula F^*_v computes the same polynomial as $\overline{A}_i[t]$. Thus, the Frege reasoning can be phrased either in terms of $\overline{A}_i[t]$ or in terms of F^*_v.

Frege proof size bound via induced formulas. We now bound the size of the resulting Frege derivation in terms of the induced formulas F^*_v tracked in A'_F and the complexity of the entries of witness matrices.

Recall that for every node $v \in A'_F$, the induced formula F^*_v computes the same polynomial as the sub-ABP polynomial $A'_F[v,t]$, and by Lemma 4.16 we have

$$|F^*_v| \leq s^{O(\log^3 s)} \qquad \text{for all } v \in A'_F, \tag{4.30}$$

where s is the original formula size before balancing/homogenization and $|A'_F| = s^{O(\log s)}$.

Formula complexity of the witness matrices. It remains to account for the formula complexity of the entries of the witness matrices λ''_i and T''_j appearing in (4.24)–(4.25).

By the construction given in the proof of Lemma 4.21 and the denominator–clearing procedure of Equation 5.41, every entry of λ''_i and T''_j is computed by an ABP over $\mathbf{GF}(2)[\bar{y}]$ of size at most $\mathrm{poly}(s')$, where the coefficients appearing in the edge labels are k variate polynomials of degree $\mathrm{poly}(s', d)$. Here k is the number of $\bar{y}$–variables, d is the depth of the underlying ABP and $s' = s^{O(\log s)}$ is the size of the ABP A'_F. Each such coefficient polynomials in $\mathbb{F}[\bar{y}]$ has at most $d'^{O(k)}$

many monomials where d' is the maximum degree. If we take the coefficients' size complexity into the account of the ABP size, then the ABP complexity increases to at most $(s'd')^{O(k)}$. Since $d' = \mathrm{poly}(s')$ the ABP complexity is $s'^{O(k)} = s^{O(k \log s)}$.

By the standard conversion from ABPs to formulas (see, e.g., the classical Brent–Spira transformation), any ABP of size S can be converted into an algebraic formula of size $S^{O(\log S)}$. Applying this to the above ABPs, every entry of λ_i'' and T_j'' admits an algebraic formula of size

$$|\lambda_i''[p,q]|,\ |T_j''[p,q]| \ \leq\ \left((s')^{O(k)}\right)^{O(k \log s')} \ =\ s^{O(k^3 \log^3 s)} \qquad \text{for all entries.} \tag{4.31}$$

In the Booleanized identities (4.26) and (4.27), the only formulas that appear are Booleanizations of: (i) entries of the vectors $\overline{A}_i$ (equivalently, of the induced formulas F_v^*), and (ii) entries of the witness matrices λ_i'' and T_j''.

By Equation 4.30, every induced formula F_v^* has size at most $s^{O(\log^3 s)}$. By Equation 4.31, every entry of λ_i'' and T_j'' has algebraic formula size at most $s^{O(k^3 \log^3 s)}$. After Booleanization, the size of the corresponding propositional formulas increases by at most a polynomial factor.

Therefore, every propositional formula occurring in the Frege derivation has size at most

$$s^{O(\log^3 s)} \cdot s^{O(k^3 \log^3 s)} \ =\ s^{O(k^3 \log^6 s)}.$$

Since the number of Frege proof lines is polynomial in $|A'_F|$ and hence at most $s^{O(\log s)}$, the total Frege proof size is bounded by

$$|\pi_{\mathrm{Frege}}| \ \leq\ s^{O(k^3 \log^6 s)} \sim s^{(k \log s)^{O(1)}}. \qquad \square$$

This completes the proof of our main theorem assuming Lemma 4.21, which we prove in the next section.

5 Existence of a Partially Commutative ABP Identity Witness

Our goal in this section is to prove Lemma 4.21. The first step is to prove the following version of Lemma 4.21, where the coefficients of the witness matrices are defined over the (commutative) function field. Throughout this section, we use $\mathbb{F}(\bar{y})$ to denote the commutative function field containing the polynomial ring $\mathbb{F}[\bar{y}]$. In addition, we define the degree of a rational function in $\mathbb{F}(\bar{y})$ to be the maximum degree of the numerator and denominator.

Lemma 5.1. (ABP identity witnesses over $\mathbb{F}(\overline{y})$) *Let A be a partially commutative ABP computing a polynomial in $\mathbb{F}[\overline{y}]\langle X_1 \sqcup X_2 \rangle$ of depth d, which is homogeneous over X_1. For every $i \in [d]$, let $\overline{A_i}$ be as defined in Subsection 4.6.1. Suppose that A computes the zero polynomial. Then there exist matrices*

$\lambda'_0, \lambda'_1, \ldots, \lambda'_{d-1}$ *and* $T'_1, T'_2, \ldots, T'_{d-1}$, *where* $\lambda'_0 := 1$ *and*

$$\lambda'_r \in \mathbb{F}(\bar{y}) \langle X_2 \rangle^{w_r \times w_r}, \qquad T'_{r+1} \in \mathbb{F}(\bar{y}) \langle X_1 \sqcup X_2 \rangle^{w_r \times w_{r+1}},$$

such that the following identities hold.

$$\lambda'_{d-i} \cdot \overline{A}_i = \overline{0} \qquad \forall\, i = 1, 2, \ldots, d. \tag{5.2}$$

$$\lambda'_{d-i} \cdot \overline{A}_i = T'_{d-i+1} \cdot \lambda'_{d-i+1} \cdot \overline{A}_{i-1} \qquad \forall\, i = 2, 3, \ldots, d. \tag{5.3}$$

Further, the coefficients' degree of each entry in the witness matrices λ'_i *and* T'_i *is at most* $\mathrm{poly}(s, d)$ *and the ABP size complexity of each entry in the witness matrices* λ'_i, T'_j *is at most* $\mathrm{poly}(s)$.

The technical core of this proof is to prove a factorization lemma for a matrix of noncommutative ABPs. For that, we need to use some algebraic tools that we develop in the following subsections. Throughout this section, unless stated otherwise, $X = \{x_1, \ldots, x_n\}$ denotes a set of noncommutative variables.

5.1 Linear Matrix Factorization

For a matrix of noncommutative polynomials, invertibility is defined over the universal skew field of fractions of the noncommutative polynomial ring. The noncommutative rank is the size of a largest invertible submatrix. We use the following characterization of the noncommutative rank due to Cohn [Coh95a].

Theorem 5.4. (Noncommutative rank). *Let* $M \in \mathbb{F}\langle X \rangle^{\ell \times m}$ *be a matrix of noncommutative polynomials of noncommutative rank* $r < \min\{\ell, m\}$. *Then there exist matrices* $P \in \mathbb{F}\langle X \rangle^{\ell \times r}$ *and* $Q \in \mathbb{F}\langle X \rangle^{r \times m}$ *such that* $M = P \cdot Q$.

Indeed, if M is a linear matrix, the following equivalent characterizations are known from the work of [Coh95a, FR04, IQS17, GGdOW20].

Theorem 5.5. (Noncommutative rank of a linear matrix). *Let* $L = \sum_{i=1}^{n} A_i x_i \in \mathbb{F}^{\ell \times m} \langle X \rangle$ *be a linear matrix and* $\ell \geq m$. *Then the following are equivalent:*

- *The noncommutative rank of* L *is* r.
- *There exist linear matrices* $L_1 \in \mathbb{F}\langle X \rangle^{\ell \times r}$ *and* $L_2 \in \mathbb{F}\langle X \rangle^{r \times m}$ *such that* $L = L_1 \cdot L_2$.
- L *has a* $(\ell - r)$-shrunk subspace, *i.e. there exist a subspace* $T \subseteq \mathbb{F}^m$ *and a subspace* $W \subseteq \mathbb{F}^{\ell}$ *such that* $A_i T \subseteq W$ *for every* $i \in [n]$ *and* $\dim(W) < \dim(T)$.
- L *is r-decomposable, i.e. there exist invertible matrices* $U \in \mathbb{F}^{\ell \times \ell}$ *and* $V \in \mathbb{F}^{m \times m}$ *such that*

$$ULV = \left[\begin{array}{c|c} L' & 0 \\ \hline B & C \end{array}\right], \tag{5.6}$$

where the zero block is of size $i \times j$ with $i + j = \ell + m - r$.

We need the following stronger version of Cohn's characterization of noncommutative rank in our proof. To the best of our knowledge, any bound on the complexity of their coefficients has not been previously studied.

Lemma 5.7. (Linear Matrix Factorization). *Suppose $\mathbb{F}' = \mathbb{F}(\bar{y})$ be the field of rational functions in $\bar{y}$ over $\mathbb{F}$ and $X = \{x_1, x_2, \ldots, x_n\}$. Let $L \in \mathbb{F}'\langle X\rangle^{\ell\times m}$ be a noncommutative linear matrix of rank $r < \min\{\ell, m\}$ and d be the maximum degree of the coefficients of the entries of L. Then we can construct linear matrices $L_1 \in \mathbb{F}'\langle X\rangle^{\ell\times r}$ and $L_2 \in \mathbb{F}'\langle X\rangle^{r\times m}$ such that $L = L_1 \cdot L_2$ in deterministic $\text{poly}(\ell, m, n)$ time. Moreover, the coefficients of the entries of L_1 and L_2 have degree at most $\text{poly}(\ell, m, d)$.*

We divide the proof into two parts. Given such a linear matrix L, we first use [IQS18] to construct a nontrivial shrunk subspace. We then exploit this subspace to obtain a zero block [FR04], which in turn yields a factorization of the input matrix. We add the complete details of the proof in Subsection C.1.

5.2 Matrix Polynomial Factorization

We are now ready to prove the factorization lemma for matrix of noncommutative ABPs.

Lemma 5.8. *[Factorization Lemma] Let $M \in \mathbb{F}[\bar{y}]\langle X\rangle^{\ell\times m}$ be a matrix of noncommutative polynomials with $\text{ncrank}(M) = r < \min\{\ell, m\}$. Assume each entry of M is computed by a size s noncommutative ABP. Then we can compute a nontrivial factorization of M, namely $M = G \cdot H$, such that the entries of $G \in \mathbb{F}(\bar{y})\langle X\rangle^{\ell\times r}$ and $H \in \mathbb{F}(\bar{y})\langle X\rangle^{r\times m}$ have $\text{poly}(s)$-size noncommutative ABPs. Moreover, if d is the upper bound on the degree of the coefficients in each entry of M, then the degree of the coefficients in each entry of the factors G, H is $\text{poly}(\ell, m, d)$.*

5.2.1 Higman's Linearization for a Matrix of Noncommutative ABPs

In this section, we bound the ABP complexity of Higman linearization for a rectangular matrix whose entries are computed by noncommutative ABPs. Equivalently (and more conveniently for linearization), we may assume that each entry is given as a product of *linear matrices* over $\mathbb{F}\langle X\rangle$.

ABPs as products of linear matrices. A standard way to view a noncommutative ABP of size S is as an iterated product of linear matrices. Concretely, if a polynomial $f \in \mathbb{F}\langle X\rangle$ is computed by an ABP of size S, then there exist integers $d \le S$, widths $w_1, \ldots, w_{d+1} \le S$, linear matrices $A_\ell \in \mathbb{F}\langle X\rangle^{w_\ell\times w_{\ell+1}}$ and vectors $u \in \mathbb{F}^{w_1}, v \in \mathbb{F}^{w_{d+1}}$ such that

$$f = u^\top A_1 A_2 \cdots A_d\, v. \tag{5.9}$$

(Here "linear" means each entry is an affine linear form in X.)

Thus, for a matrix $M \in \mathbb{F}\langle X\rangle^{s\times s'}$ whose entries have ABP size at most S, after possibly padding widths (by inserting identity matrices), we may assume that

$$M[i,j] \;=\; A_1^{(i,j)} A_2^{(i,j)} \cdots A_d^{(i,j)}, \tag{5.10}$$

where for each $\ell \in [d]$ the factor $A_\ell^{(i,j)}$ is a linear matrix of dimension $w_\ell \times w_{\ell+1}$ with $w_\ell \le S$, and the width profile $(w_1, \dots, w_{d+1})$ is *independent of* (i,j). This is the model we linearize.

Such a result was previously known only when each entry of the matrix is computed by a *noncommutative formula* [GGdOW20, Proposition A.2].

A local (one-entry) Higman gadget

We first isolate the basic gadget that linearizes a single product entry while introducing only upper/lower triangular multipliers with diagonal entries are 1.

Lemma 5.11 (Local Higman linearization for one product block [AJ25])**.** *Let $B \in \mathbb{F}\langle X\rangle^{w_1\times w_{d+1}}$ be a block polynomial of the form*

$$B = A_1 A_2 \cdots A_d,$$

where $A_\ell \in \mathbb{F}\langle X\rangle^{w_\ell \times w_{\ell+1}}$ is a linear matrix*. Let $w := \sum_{\ell=2}^{d} w_\ell$. Then there exist matrices P_B, Q_B, L_B such that*

$$\begin{pmatrix} B & 0 \\ 0 & I_w \end{pmatrix} \;=\; P_B \cdot L_B \cdot Q_B, \tag{5.12}$$

where:

1. *P_B is upper triangular and Q_B is lower triangular,*
2. *all diagonal entries of P_B, Q_B are equal to 1,*
3. *L_B is a* linear matrix *(every entry is either 0, 1, or an affine linear form in X),*
4. *every nonzero entry of P_B, Q_B is either ± 1 or an entry of one of the factors A_ℓ. Hence, if each A_ℓ is computable by* $\mathrm{poly}(S)$*-size ABPs, then so are the entries of P_B, Q_B.*

Moreover, L_B can be taken in the companion form

$$L_B = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & A_1 \\ -A_d & I_{w_d} & 0 & \cdots & 0 & 0 \\ 0 & -A_{d-1} & I_{w_{d-1}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & I_{w_3} & 0 \\ 0 & 0 & 0 & \cdots & -A_2 & I_{w_2} \end{pmatrix}. \tag{5.13}$$

Proof. This is the standard Higman linearization (sometimes called "Higman's trick"). One can prove (5.12) by repeatedly applying the $d = 2$ elimination step: for any compatible linear matrices $U \in \mathbb{F}\langle X\rangle^{a\times b}$ and $V \in \mathbb{F}\langle X\rangle^{b\times c}$, set

$$\begin{pmatrix} UV & 0 \\ 0 & I_b \end{pmatrix} = \underbrace{\begin{pmatrix} I_a & -U \\ 0 & I_b \end{pmatrix}}_{\text{upper triangular}} \cdot \underbrace{\begin{pmatrix} 0 & U \\ -V & I_b \end{pmatrix}}_{\text{linear}} \cdot \underbrace{\begin{pmatrix} I_c & 0 \\ V & I_b \end{pmatrix}}_{\text{lower triangular}} .$$

after a suitable block permutation to match dimensions (equivalently, embed into the bottom-right corner). Iterating this construction for the product $A_1 \cdots A_d$ yields the companion matrix (5.13), and the accumulated left and right multipliers remain upper/lower triangular with diagonal 1. Every nonzero off-diagonal entry introduced is a copy of some A_ℓ or a constant; hence, it is ABP-computable whenever the A_ℓ are. ☐

Global linearization for a rectangular matrix of ABP entries

We now lift the local gadget to a full $s \times s'$ matrix M, linearizing each entry in place while keeping the global triangular structure.

Lemma 5.14 (Higman linearization for a matrix of ABP-entries)**.** *Let $M \in \mathbb{F}\langle X\rangle^{s\times s'}$ be a rectangular matrix such that for every $(i, j) \in [s] \times [s']$*

$$M[i, j] \;=\; A_1^{(i,j)} A_2^{(i,j)} \cdots A_d^{(i,j)}, \tag{5.15}$$

where for each $\ell \in [d]$, the factor $A_\ell^{(i,j)}$ is a linear matrix of dimension $w_\ell \times w_{\ell+1}$ (the width profile $(w_1, \ldots, w_{d+1})$ is independent of (i, j)). Let

$$w := \sum_{\ell=2}^{d} w_\ell, \qquad t := w \cdot ss'.$$

Then there exist matrices P, L, Q over $\mathbb{F}\langle X\rangle$ such that

$$M \oplus I_t \;=\; P \cdot L \cdot Q, \tag{5.16}$$

with the following properties:

1. *P is (block) upper triangular and Q is (block) lower triangular,*
2. *all diagonal entries of P, Q are equal to* 1*,*
3. *L is a* linear matrix *(all its entries are affine linear forms in X),*
4. *every nonzero entry of P, Q is computable by an ABP of size* $\mathrm{poly}(s, s', d, w_1, \ldots, w_{d+1})$*.*

Proof. We linearize the entries of M one-by-one using Lemma 5.11 as a block gadget, and we do so in an order that preserves the triangularity of the accumulated multipliers.

Order of linearization. Fix the lexicographic order so that it scans columns from right to left, and within each column, scans the rows from bottom to top:

$$(s, s'), (s-1, s'), \ldots, (1, s'), (s, s'-1), \ldots, (1, 1).$$

At each step, we linearize exactly one block $M[i, j]$ while keeping previously linearized blocks linear.

One step: linearizing a chosen block. Suppose at some stage we have a matrix $\widetilde{M}$ of size

$$(s \cdot w_1 + w \cdot \tau) \times (s' \cdot w_{d+1} + w \cdot \tau)$$

for some $\tau \in \{0, 1, \ldots, ss'\}$, obtained from M by adjoining τ slack identity blocks I_w and performing some triangular multiplications, such that:

- all already-processed positions contain linearized companion blocks,
- all not-yet-processed positions still contain their original product blocks.

Now pick the next unprocessed block position (i, j). We adjoin one more identity block I_w (increasing τ to $\tau + 1$), and we *embed* Lemma 5.11 into the coordinates corresponding to the block (i, j) together with this new slack.

Formally, consider $\widetilde{M} \oplus I_w$. By multiplying on the left and right by suitable *block permutation matrices* Π_L, Π_R (which are invertible and linear), we may move the block (i, j) and the new slack block to the bottom-right corner. In those coordinates, the bottom-right $(w_1 + w) \times (w_{d+1} + w)$ corner equals

$$\begin{pmatrix} M[i,j] & 0 \\ 0 & I_w \end{pmatrix}.$$

Applying Lemma 5.11 to this corner yields upper/lower triangular matrices $P_{i,j}, Q_{i,j}$ with diagonal 1 and replaces this corner by a linear matrix $L_{i,j}$. Undoing the permutations gives a factorization of the form

$$\widetilde{M} \oplus I_w = \widehat{P}_{i,j} \cdot \widetilde{M}^{\text{new}} \cdot \widehat{Q}_{i,j}, \tag{5.17}$$

where:

- $\widehat{P}_{i,j}$ is block upper triangular with diagonal 1,
- $\widehat{Q}_{i,j}$ is block lower triangular with diagonal 1,
- $\widetilde{M}^{\text{new}}$ is identical to $\widetilde{M}$ except that entry (i, j) has been replaced by a linear companion block, and the new slack coordinates are now "used".

Importantly, the nonzero entries of $\widehat{P}_{i,j}, \widehat{Q}_{i,j}$ are either constants or entries from the linear factors $A_\ell^{(i,j)}$, hence ABP-computable with polynomial size.

Completing all steps. Iterate (5.17) for all ss' positions. Each step consumes one fresh slack block I_w. After ss' steps we obtain a matrix L in which *every original product block has been linearized*, hence L is linear everywhere. Let P (resp. Q) be the product of all left (resp. right) multipliers produced by the steps. Since a product of upper triangular matrices with diagonal 1 is again upper triangular with diagonal 1, and similarly for lower triangular, we conclude that the final P, Q satisfy the required triangularity conditions. Hence, the linear matrix L in the right side of Equation 5.16 preserves the *co-rank*. the The final number of slack dimensions is $t = w \cdot ss'$.

ABP size bound. Each step introduces only constants and entries of the linear matrices $A_\ell^{(i,j)}$ into P and Q. Therefore each entry of P, Q is computed by an ABP of size polynomial in $(s, s', d, w_1, \ldots, w_{d+1})$. This proves (5.16). □

Restating in pure-ABP language. If the entries of M are given directly by ABPs of size $\leq S$, then by the standard conversion (5.9) we can rewrite each entry as a product of $d \leq S$ linear matrices with widths $\leq S$. After padding to equalize width profiles across entries, Lemma 5.14 applies and yields $t = \text{poly}(s, s', S)$.

Corollary 5.18 (Higman linearization for ABP entries)**.** *Let $M \in \mathbb{F}\langle X\rangle^{s\times s'}$ be such that each entry is computed by a noncommutative ABP of size $\leq S$. Then there exists $t = \text{poly}(s, s', S)$ and matrices P, L, Q such that*

$$M \oplus I_t = P \cdot L \cdot Q,$$

where P is upper triangular with diagonal 1, Q is lower triangular with diagonal 1, L is linear, and every entry of P, Q is computable by $\text{poly}(s, s', S)$-size ABPs.

An illustrative example is given in the appendix Subsection C.2 for better exposition.

5.2.2 Proof of Factorization Lemma

Now we are ready to prove the existence of *small* factors witnessing the *low rank*.

Proof of Lemma 5.8. Let $M \in \mathbb{F}[\bar{y}]\langle X\rangle^{\ell\times m}$ with $\text{ncrank}(M) = r < m < \ell$. We first apply Corollary 5.18 to M and obtain

$$\left[\begin{array}{c|c} M & 0 \\ \hline 0 & I_{k'} \end{array}\right] = P \cdot L \cdot D, \tag{5.19}$$

where $L \in \mathbb{F}[\bar{y}]\langle X\rangle^{(\ell+k')\times(m+k')}$ is a linear matrix, and $P \in \mathbb{F}[\bar{y}]\langle X\rangle^{(\ell+k')\times(\ell+k')}, D \in \mathbb{F}[\bar{y}]\langle X\rangle^{(m+k')\times(m+k')}$ are upper triangular and lower triangular full matrices, respectively, with all diagonal entries equal to 1. Moreover, every entry of P, L, D is computable by a size $\text{poly}(s)$ noncommutative ABP.

Our goal is to compute a nontrivial rank factorization $M = G \cdot H$ with inner dimension r, such that the entries of G and H have poly(s)-size ABPs.

Write the block decompositions

$$P = \left[\begin{array}{c|c} P_{1,1} & P_{1,2} \\ \hline 0 & P_{2,2} \end{array}\right], \qquad L = \left[\begin{array}{c|c} L_{1,1} & L_{1,2} \\ \hline L_{2,1} & L_{2,2} \end{array}\right], \qquad D = \left[\begin{array}{c|c} D_{1,1} & 0 \\ \hline D_{2,1} & D_{2,2} \end{array}\right].$$

Since P and D are invertible (they are full triangular with 1's on the diagonal), taking noncommutative rank on both sides of Equation 5.19 gives

$$\text{ncrank}(L) = \text{ncrank}\left(\left[\begin{array}{c|c} M & 0 \\ \hline 0 & I_{k'} \end{array}\right]\right) = \text{ncrank}(M) + k' = r + k'.$$

By Theorem 5.5, there exist invertible matrices $U \in \mathbb{F}(\bar{y})^{(\ell+k')\times(\ell+k')}$ and $V' \in \mathbb{F}(\bar{y})^{(m+k')\times(m+k')}$ such that

$$U \cdot L \cdot V' = \left[\begin{array}{c|c} L'_{1,1} & 0 \\ \hline L'_{1,2} & L'_{2,2} \end{array}\right],$$

where the top-right block is the zero matrix. Let the dimensions of this 0 block be $i \times j$, therefore, $i + j = \ell + k' + m - r$. Efficient construction of U and V' follows from Lemma C.2 and Lemma C.1, as discussed in Subsection C.1.

We rewrite Equation 5.19 by multiplying on the left by $U \cdot P^{-1}$ and on the right by $D^{-1} \cdot V'$:

$$\begin{aligned} UP^{-1}\left[\begin{array}{c|c} M & 0 \\ \hline 0 & I_{k'} \end{array}\right] D^{-1}V' &= \underbrace{\left[\begin{array}{c|c} P'_{1,1} & P'_{1,2} \\ \hline 0 & P'_{2,2} \end{array}\right]}_{U\cdot P^{-1}} \left[\begin{array}{c|c} M & 0 \\ \hline 0 & I_{k'} \end{array}\right] \underbrace{\left[\begin{array}{c|c} D'_{1,1} & 0 \\ \hline D'_{2,1} & D'_{2,2} \end{array}\right]}_{D^{-1}\cdot V'} \\ &= \left[\begin{array}{c|c} L'_{1,1} & 0 \\ \hline L'_{1,2} & L'_{2,2} \end{array}\right] = L' \text{ (say)}. \end{aligned} \tag{5.20}$$

Expanding the left-hand side blockwise, Equation 5.20 implies

$$\left[\begin{array}{c|c} P'_{1,1}MD'_{1,1} + P'_{1,2}D'_{2,1} & P'_{1,2}D'_{2,2} \\ \hline P'_{2,2}D'_{2,1} & P'_{2,2}D'_{2,2} \end{array}\right] = \left[\begin{array}{c|c} L'_{1,1} & 0 \\ \hline L'_{1,2} & L'_{2,2} \end{array}\right].$$

Note that $P'_{1,1}, P'_{2,2}, D'_{1,1}, D'_{2,2}$ are invertible, because P', D' remain block triangular with invertible diagonal blocks.

We now analyze the location of the 0 block (of dimension $i \times j$) in the right-hand side of Equation 5.20.

Case 1 : $i \geq \ell,\ j \geq k'$**.** In this case, the matrix $P'_{1,2}D'_{2,2}$ has dimension $\ell \times k'$, and it seats inside the

0 block on the right-hand side. Hence

$$P'_{1,2}D'_{2,2} = 0 \qquad \Longrightarrow \qquad P'_{1,2} = 0 \qquad (\text{since } D'_{2,2} \text{ is invertible}),$$

and therefore also $P'_{1,2}D'_{2,1} = 0$.

Thus, the $(1,1)$-block identity becomes $P'_{1,1}MD'_{1,1} = L''$, where L'' denotes the $\ell \times m$ top-left block of L'. Since $\text{ncrank}(L') = \text{ncrank}(L) = r + k'$, the matrix L' does not have full row rank. Applying Lemma 5.7, we obtain a factorization $L' = L'_1 \cdot L'_2$ with L'_1 of dimension $(\ell + k') \times (r + k')$ and L'_2 of dimension $(r + k') \times (m + k')$. Taking the top-left $\ell \times m$ block in this product yields a nontrivial factorization $L'' = L''_1 \cdot L''_2$ with L''_1 of dimension $\ell \times r$ and L''_2 of dimension $r \times m$. Therefore,

$$P'_{1,1}MD'_{1,1} = L''_1 \cdot L''_2 \qquad \Longrightarrow \qquad M = \underbrace{(P'^{-1}_{1,1}L''_1)}_{G} \cdot \underbrace{(L''_2 D'^{-1}_{1,1})}_{H}.$$

This gives the required nontrivial factorization of M.

Case 2 : $i > \ell$, $j < k'$**.** Note that, assuming $j < k'$, we obtain $i + j = \ell + k' + m - r \implies i = \ell + k' + m - r - j > \ell + k' + (m - r) - k' > \ell$ where $m - r > 0$. Consider the $(2,2)$-block on the left-hand side, namely $P'_{2,2}D'_{2,2}$. This matrix is invertible and hence; has full noncommutative rank k'. However, when $i > \ell$ and $j < k'$, the placement of the 0 block forces a 0 sub-matrix of dimension $(i - \ell) \times j$ inside the $k' \times k'$ invertible block. Here, $(i - \ell) + j = (\ell + k' + m - r - \ell) = k' + (m - r) > k'$. But an invertible $k' \times k'$ matrix over the free skew field cannot contain a zero sub-matrix of size $p \times q$ with $p + q > k'$, since that would force the rank $< k'$. This is a contradiction, so Case 2 cannot occur.

Case 3 : $i < \ell$**.** In this case, the first i rows of the block $P'_{1,2}D'_{2,2}$ must be zero (since they lie in the 0 block on the right-hand side), and since $D'_{2,2}$ is invertible, the top i rows of $P'_{1,2}$ must be zero.

Moreover, $j = \ell + k' + m - r - i > k' + (m - r) > k' \qquad (\text{since } m > r)$, so the top-right block of dimension $i \times (j - k')$ inside $P'_{1,2}D'_{2,1}$ is also forced to be zero. Consequently, the matrix $P'_{1,1}MD'_{1,1}$ has the following block form:

$$P'_{1,1}MD'_{1,1} = \left[\begin{array}{c|c} M'_{1,1} & 0 \\ \hline M'_{2,1} & M'_{2,2} \end{array}\right]. \tag{5.21}$$

The top-right 0 block in Equation 5.21 is of dimension $i \times (j - k')$. Since $i + j = \ell + k' + m - r$, we have $j - k' = \ell + m - r - i$. Therefore, the number of columns of $M'_{1,1}$ is $m - (j - k') = m - (\ell + m - r - i) = r + i - \ell$, and $M'_{1,1}$ has i rows.

We now factor the right-hand side of Equation 5.21 as

$$\left[\begin{array}{c|c} M'_{1,1} & 0 \\ \hline M'_{2,1} & M'_{2,2} \end{array}\right] = \underbrace{\left[\begin{array}{c|c} M'_{1,1} & 0 \\ \hline M'_{2,1} & I_{k-i} \end{array}\right]}_{G} \cdot \underbrace{\left[\begin{array}{c|c} I_{r+i-k} & 0 \\ \hline 0 & M'_{2,2} \end{array}\right]}_{H}.$$

Note that, the G matrix has $i + (\ell - i) = \ell$ rows and $(r + i - \ell) + (\ell - i) = r$ columns and the H matrix has $(r + i - \ell) + (\ell - i) = r$ columns and $(r + i - \ell) + (j - k') = (r + i - \ell) + (\ell + m - r - i) = m$ columns. Finally, multiplying by the invertible matrices $P'^{-1}_{1,1}$ and $D'^{-1}_{1,1}$ on the left and right (respectively) yields a nontrivial factorization of M.

ABP size and the degree of the coefficients. All matrices produced above are obtained from M using a constant number of operations: (i) Higman linearization as in Corollary 5.18, (ii) multiplying on the left and right by invertible (block) matrices over the base field, as in Theorem 5.5 ,(iii) linear matrix factorization, and (iv) taking sub-blocks and multiplying by the inverses of the diagonal blocks of triangular full matrices. Each of these operations preserves polynomial overhead in ABP size; hence the entries of the resulting factors G, H are computable by $\mathrm{poly}(s)$-size noncommutative ABPs.

Let the degree of the coefficients in each entry of M is at most d. By Corollary 5.18, the Higman linearization step produces the matrices P, L, D in Equation 5.19 such that the coefficients appearing in L have degree bounded by $\mathrm{poly}(d)$ (and the same holds for the nonzero entries of P and D). Next, the r-decomposability transformation in Theorem 5.5 together with the explicit construction of the invertible matrices U, V from a shrunk subspace in Lemma C.2 gives the $\mathrm{poly}(\ell, m, d)$ bound on the degree of the coefficients. Finally, the linear factorization step applied to the resulting linear matrix is performed over $\mathbb{F}(\bar{y})$ and satisfies the required degree control by Lemma 5.7: the coefficients of the entries of the linear factors have degree at most $\mathrm{poly}(\ell, m, d)$. Subsequent block extraction and multiplication by the inverses of triangular diagonal blocks preserve this bound on the degree of the coefficients. Therefore, the degree of the coefficients in each entry of the factors G and H is at most $\mathrm{poly}(\ell, m, d)$. □

We have the following corollary which is required to prove Lemma 5.1 and Lemma 4.21.

Corollary 5.22. *Let $n \cdot k > m$ and $A_1, \ldots, A_n$ be $k \times m$ matrices, with each entry in these matrices being a noncommutative polynomial in $\mathbb{F}[\bar{y}]\langle X\rangle$, computable by size s and length d noncommutative ABPs. Then there exists a basis matrix $B_{r\times m}$ whose entries are noncommutative polynomials computed by size s ABPs and coefficient matrices $C_1, \ldots, C_n$ whose entries are noncommutative polynomials computed by size $\mathrm{poly}(s)$ ABPs, such that $A_i = C_i \cdot B \;\forall\, i \in [n]$. Moreover, if the maximum degree of the coefficients appearing in $A_1, \ldots, A_n$ is d, then the maximum degree of the coefficients appearing in C_i, B is $\mathrm{poly}(k, m, n, d)$.*

To prove Corollary 5.22, it suffices to exhibit a witness for a noncommutative polynomial matrix $A_{n\times m}$ of noncommutative rank r, where $r \leq \min(n, m)$, which Lemma 5.8 does.

Proof of the Corollary 5.22. Let $n \cdot k > m$ and let $A_1, \ldots, A_n$ be $k \times m$ matrices over $\mathbb{F}[\bar{y}]\langle X \rangle$, where every entry of every A_i is computable by size-s and length-d noncommutative ABPs.

Define the stacked matrix

$$A \ := \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} \in \mathbb{F}[\bar{y}]\langle \mathrm{X} \rangle^{(nk)\times m}.$$

Let

$$r \ := \ \mathrm{ncrank}(A).$$

By definition, $r \leq \min(nk, m)$, and since $n \cdot k > m$ we have $r \leq m$.

Step 1: Apply the rank witness factorization when $r < m$. By Lemma 5.8, there exist matrices

$$A' \in \mathbb{F}(\bar{y})\langle \mathrm{X} \rangle^{(nk)\times r}, \qquad A'' \in \mathbb{F}(\bar{y})\langle \mathrm{X} \rangle^{r\times m}$$

such that

$$A \ = \ A' \cdot A'', \tag{5.23}$$

and all entries of A' and A'' are computable by size-poly(s) noncommutative ABPs and the degree of the coefficients appearing in each entry of the factors is poly(s).

Step 2: Define the common basis matrix B. Set

$$B \ := \ A'' \in \mathbb{F}(\bar{y})\langle \mathrm{X} \rangle^{r\times m}.$$

Then B has r rows and m columns, exactly as required.

Step 3: Extract the coefficient matrices C_i. Write A' in n consecutive blocks of k rows each:

$$A' = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix}, \qquad C_i \in \mathbb{F}(\bar{y})\langle \mathrm{X} \rangle^{k\times r}.$$

Substituting this block decomposition into Equation 5.23 gives

$$\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} = \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} \cdot B.$$

Therefore, comparing block rows, for every $i \in [n]$ we have $A_i = C_i \cdot B$.

On the other hand if the rank $r = m$, then take the basis matrix $B = I_m$ and $C = A_{nk \times r}$.

Additional property: full column rank of the coefficient matrix. In the above decomposition, define the stacked coefficient matrix

$$C := \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} \in \mathbb{F}(\bar{y})\langle X \rangle^{(nk) \times r}.$$

Then C has full noncommutative column rank, i.e.

$$\mathrm{ncrank}(C) = r.$$

Indeed, since $A = C \cdot B$ and $\mathrm{ncrank}(A) = r$, we must have $\mathrm{ncrank}(C) \geq r$ (otherwise $\mathrm{ncrank}(A) \leq \mathrm{ncrank}(C) < r$), while trivially $\mathrm{ncrank}(C) \leq r$ because C has only r columns. Hence $\mathrm{ncrank}(C) = r$.

Equivalently, C admits a left inverse over the free skew field, i.e. there exists a matrix $C^{\dagger}$ (over the free skew field $\mathbb{F}(\bar{y})\langle\!\langle X \rangle\!\rangle$) such that

$$C^{\dagger} \cdot C = I_r.$$

ABP-size bound. By construction, each entry of $B = A''$ is computable by size-poly(s) ABPs, and each entry of each C_i (being a block of A') is also computable by size-poly(s) ABPs, as guaranteed by Lemma 5.8 in the case of rank $r < m$. When $r = m$, then the basis matrix is identity matrix, which is trivial to compute via ABP. Moreover, Lemma 5.8 provides the polynomially bounded degree of the coefficients. This completes the proof. □

5.3 Construction of witness matrices

In this section, we first give a proof of Lemma 5.1 using the machinery we developed in the previous subsection. We then use it to prove Lemma 4.21.

Lemma 5.1. (ABP identity witnesses over $\mathbb{F}(\bar{y})$) *Let A be a partially commutative ABP computing a polynomial in $\mathbb{F}[\bar{y}]\langle X_1 \sqcup X_2\rangle$ of depth d, which is homogeneous over X_1. For every $i \in [d]$, let $\overline{A_i}$ be as defined in Subsection 4.6.1. Suppose that A computes the zero polynomial. Then there exist matrices $\lambda'_0, \lambda'_1, \ldots, \lambda'_{d-1}$ and $T'_1, T'_2, \ldots, T'_{d-1}$, where $\lambda'_0 := 1$ and*

$$\lambda'_r \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_r \times w_r}, \qquad T'_{r+1} \in \mathbb{F}(\bar{y})\langle X_1 \sqcup X_2\rangle^{w_r \times w_{r+1}},$$

such that the following identities hold.

$$\lambda'_{d-i} \cdot \overline{A}_i = \overline{0} \qquad \forall\, i = 1, 2, \ldots, d. \tag{5.1}$$

$$\lambda'_{d-i} \cdot \overline{A}_i = T'_{d-i+1} \cdot \lambda'_{d-i+1} \cdot \overline{A}_{i-1} \qquad \forall\, i = 2, 3, \ldots, d. \tag{5.2}$$

Further, the coefficients' degree of each entry in the witness matrices λ'_i and T'_i is at most $\mathrm{poly}(s,d)$ and the ABP size complexity of each entry in the witness matrices λ'_i, T'_j is at most $\mathrm{poly}(s)$.

Proof. Let A be a partially commutative ABP over $X = X_1 \sqcup X_2 \sqcup \cdots \sqcup X_k$ (with $|X_i| = 1$ for $3 \le i \le k$) of depth d and layer widths $w_0, \ldots, w_d$, which is *homogeneous over* X_1 and computes the identically zero polynomial in $\mathbb{F}\langle X_{[k]}\rangle$. We assume $X_i = \{y_{i-2}\}$ for every $3 \le i \le k$. Let $\bar{y} := \{y_1, \ldots, y_{k-2}\}$ and our ABP A is defined over the field $\mathbb{F}(\bar{y})$.

For $i \in [d]$, recall the column vector

$$\overline{A}_i \;=\; \big(A[u_{d-i,1}, t],\, A[u_{d-i,2}, t],\, \ldots,\, A[u_{d-i,w_{d-i}}, t]\big)^\top \in \mathbb{F}\langle X_{[k]}\rangle^{w_{d-i}}.$$

In particular, $\overline{A}_d$ is the polynomial computed by A, and hence $\overline{A}_d = 0$.

Homogeneity convention. Since A is homogeneous over X_1 from our assumption, we view each edge label of A as a homogeneous linear form in the X_1-variables whose coefficients lie in $\mathbb{F}[\bar{y}]\langle X_2\rangle$. Equivalently, for every layer transition $r \to r+1$ there exist coefficient matrices

$$M_1, \ldots, M_n \in \mathbb{F}[\bar{y}]\langle X_2\rangle^{w_r \times w_{r+1}}$$

such that the one-layer ABP recurrence from layer r to layer $r+1$ can be written as

$$\overline{A}_{d-r} \;=\; \sum_{j=1}^{n} x_{1,j} \cdot \big(M_j \cdot \overline{A}_{d-r-1}\big), \tag{5.24}$$

where the product is matrix-vector multiplication over $\mathbb{F}[\bar{y}]\langle X_2\rangle$, and the $x_{1,j}$ appears as *left* multipliers.

Step 0: initialize λ'_0. Set

$$\lambda'_0 \ := \ 1 \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_0 \times w_0}.$$

Then $\lambda'_0 \cdot \overline{A}_d = \overline{A}_d = 0$, which is Item (3) in the statement.

Inductive construction. We construct, for $r = 0, 1, \ldots, d-2$, matrices

$$T'_{r+1} \in \mathbb{F}(\bar{y})\langle X_1 \sqcup X_2\rangle^{w_r \times w_{r+1}} \qquad \text{and} \qquad \lambda'_{r+1} \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_{r+1} \times w_{r+1}}$$

such that for every r we have both:

$$\lambda'_r \cdot \overline{A}_{d-r} = 0, \tag{5.25}$$

$$\lambda'_r \cdot \overline{A}_{d-r} = T'_{r+1} \cdot \lambda'_{r+1} \cdot \overline{A}_{d-r-1}. \tag{5.26}$$

Note that Equation 5.26 is exactly the desired transition identity, and together with Equation 5.25 it will imply the next vanishing statement $\lambda'_{r+1} \cdot \overline{A}_{d-r-1} = 0$.

Fix $r \in \{0, \ldots, d-2\}$ and assume the inductive hypothesis

$$\lambda'_r \cdot \overline{A}_{d-r} = 0. \tag{5.27}$$

Step 1: form the rectangular coefficient matrix over $\mathbb{F}(\bar{y})\langle X_2\rangle$. Using Equation 5.24 and left-multiplying by λ'_r, we get

$$\lambda'_r \cdot \overline{A}_{d-r} \ = \ \sum_{j=1}^{n} x_{1,j} \cdot \big((\lambda'_r M_j) \cdot \overline{A}_{d-r-1}\big). \tag{5.28}$$

By Equation 5.27, the left-hand side is 0. Since the variables in X_1 are noncommuting and appear as *leftmost* multipliers in Equation 5.28, homogeneity over X_1 implies that each coefficient must vanish:

$$(\lambda'_r M_j) \cdot \overline{A}_{d-r-1} = 0 \qquad \forall\, j \in [n]. \tag{5.29}$$

Define the stacked rectangular matrix

$$T \ := \ \begin{pmatrix} \lambda'_r \cdot M_1 \\ \lambda'_r \cdot M_2 \\ \vdots \\ \lambda'_r \cdot M_n \end{pmatrix} \ \in \ \mathbb{F}(\bar{y})\langle X_2\rangle^{(nw_r) \times w_{r+1}}. \tag{5.30}$$

Then Equation 5.29 is equivalent to

$$T \cdot \overline{A}_{d-r-1} = 0. \tag{5.31}$$

Step 2: apply Corollary 5.22 to extract a common left factor. Apply Corollary 5.22 over the ring $\mathbb{F}(\bar{y})\langle X_2\rangle$ to the matrices $\lambda'_r M_1, \ldots, \lambda'_r M_n \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_r \times w_{r+1}}$. We obtain:

- a basis matrix $B \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w'_{r+1} \times w_{r+1}}$,
- and coefficient matrices $C_1, \ldots, C_n \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_r \times w'_{r+1}}$,

such that for every $j \in [n]$,

$$\lambda'_r \cdot M_j \; = \; C_j \cdot B. \qquad \forall\, j \in [n]. \tag{5.32}$$

Step 3: define λ'_{r+1} and T'_{r+1}. Note that; it might be possible that; $w'_{r+1} < w_{r+1}$. In that case add 0 rows to make the number of rows of B w_{r+1}. Let this new matrix be B'. Similarly, add 0 columns to each C_j to get the number of columns w_{r+1}. Let these new matrices be C'_j.

$$\lambda'_{r+1} \; := \; B' \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_{r+1} \times w_{r+1}}, \qquad T'_{r+1} \; := \; \sum_{j=1}^{n} x_{1,j} \cdot C'_j \in \mathbb{F}(\bar{y})\langle X_1 \sqcup X_2\rangle^{w_r \times w_{r+1}}.$$

Then, combining Equation 5.24 with Equation 5.32, we obtain the propagation identity:

$$\begin{aligned} \lambda'_r \cdot \overline{A}_{d-r} = \sum_{j=1}^{n} x_{1,j} \cdot \big((\lambda'_r M_j) \cdot \overline{A}_{d-r-1} \big) = \sum_{j=1}^{n} x_{1,j} \cdot \big((C_j B) \cdot \overline{A}_{d-r-1} \big) = \left(\sum_{j=1}^{n} x_{1,j} \cdot C_j \right) \cdot \big(B \cdot \overline{A}_{d-r-1} \big) \\ = T'_{r+1} \cdot \lambda'_{r+1} \cdot \overline{A}_{d-r-1}. \end{aligned} \tag{5.33}$$

This proves Equation 5.26 for the current r.

Step 4: derive the next vanishing statement $\lambda'_{r+1} \cdot \overline{A}_{d-r-1} = 0$. We now derive the next vanishing statement, namely $\lambda'_{r+1} \cdot \overline{A}_{d-r-1} = 0$. We have two cases to consider here. The first case is when the stacked matrix T, defined in Equation 5.30 has non commutative rank $r < w_{i+1}$. For this case, recall from Equation 5.29; for every $j \in [n]$,

$$(\lambda'_r M_j) \cdot \overline{A}_{d-r-1} = 0.$$

In Step 2, we applied Corollary 5.22 to the family $\{\lambda'_r M_j\}_{j \in [n]}$ (each of dimension $w_r \times w_{r+1}$) and obtained a factorization of the form

$$\lambda'_r \cdot M_j \; = \; C_j \cdot B \qquad \forall\, j \in [n], \quad \text{where } B \in \mathbb{F}(\bar{y})\langle X_2\rangle^{r \times w_{r+1}}, \text{ and } C_j \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_r \times r}. \tag{5.34}$$

Substituting Equation 5.34 into the coefficient vanishing gives, for all $j \in [n]$,

$$0 \; = \; (\lambda'_r M_j) \cdot \overline{A}_{d-r-1} \; = \; (C_j B) \cdot \overline{A}_{d-r-1} \; = \; C_j \cdot \big(B \cdot \overline{A}_{d-r-1} \big) \; \forall\, j \in [n].$$

Hence

$$C \cdot \left(B \cdot \overline{A}_{d-r-1}\right) = 0 \quad \text{where} \quad C := \begin{pmatrix} C_1 \\ C_2 \\ \vdots \\ C_n \end{pmatrix} \in \mathbb{F}(\bar{y})\langle X_2\rangle^{(nw_r)\times r}. \tag{5.35}$$

The Corollary 5.22 shows that the stacked coefficient matrix C has full noncommutative column rank, r. Equivalently, C admits a left inverse over $\mathbb{F}(\bar{y})\langle\!\langle X_2\rangle\!\rangle$; i.e., there exists a matrix $C^\dagger$ (over the free skew field) such that

$$C^\dagger \cdot C \;=\; I_r.$$

Left-multiplying Equation 5.35 by $C^\dagger$ therefore gives

$$B \cdot \overline{A}_{d-r-1} = 0 \implies B' \cdot \overline{A}_{d-r-1} = 0 \implies \lambda'_{r+1} \cdot \overline{A}_{d-r-1} = 0. \tag{5.36}$$

On the other hand, suppose that the stacked matrix T defined in Equation 5.30 has full noncommutative column rank, i.e. $\mathrm{ncrank}(T) = w_{i+1}$ over $\mathbb{F}(\bar{y})\langle\!\langle X_2\rangle\!\rangle$. Then by Equation 5.29 we have $T \cdot \overline{A}_{d-r+1} = 0$. Each entry of $\overline{A}_{d-r+1}$ is a partially commutative ABP that is homogeneous over X_1 with coefficients in $\mathbb{F}(\bar{y})\langle X_2\rangle$. Accordingly, we may write $\overline{A}_{d-r+1} = \sum_m m \cdot v_m$, where the sum ranges over homogeneous noncommutative monomials m in X_1 and $v_m \in \mathbb{F}(\bar{y})\langle X_2\rangle^{w_{i+1}\times 1}$ is the coefficient vector of m. Substituting this expansion yields

$$0 = T \cdot \overline{A}_{d-r+1} = \sum_m m \cdot (T \cdot v_m). \tag{5.37}$$

Since the variables in X_1 are noncommuting and all monomials m appearing above have the same X_1-degree, the representation $\sum_m m \cdot (T \cdot v_m) = 0$ is unique; hence each coefficient must vanish, i.e., $T \cdot v_m = 0$ for all m. However, each v_m is a column vector over $\mathbb{F}(\bar{y})\langle X_2\rangle$, and since T has full noncommutative column rank over the free skew field, the equation $T \cdot v_m = 0$ implies a nontrivial linear dependence among the columns of T over that skew field. This contradicts the assumption that T has full column rank. This implies $\overline{A}_{d-r+1} = 0$, so trivially $\lambda'_{r+1} \cdot \overline{A}_{d-r+1} = I_{w_{r+1}\times w_{r+1}} \cdot \overline{A}_{d-r+1} = 0$.

Re-indexing to match Equation 5.2–Equation 5.3. Setting $i = d - r$ in Equation 5.33 gives

$$\lambda'_{d-i} \cdot \overline{A}_i = T'_{d-i+1} \cdot \lambda'_{d-i+1} \cdot \overline{A}_{i-1} \qquad \forall\, i = 2, 3, \ldots, d,$$

which is Equation 5.2. Moreover, Equation 5.36 yields Equation 5.3, namely

$$\lambda'_{d-i} \cdot \overline{A}_i = 0 \qquad \forall\, i = 1, 2, \ldots, d.$$

ABP-size of entries. Each entry of $\lambda'_r M_j$ is computed by a size $\mathrm{poly}(s)$ ABPs over $\mathbb{F}(\bar{y})\langle X_2\rangle$, and so are the entries of the stacked matrix T in Equation 5.30. Applying Corollary 5.22 preserves the size $\mathrm{poly}(s)$ ABPs computability of these entries, and therefore, $\lambda'_{r+1} = B$ and each C_j have polynomial size ABPs. Finally,

$$T'_{r+1} = \sum_{j=1}^{n} x_{1,j} \cdot C_j$$

is a homogeneous linear form in X_1 with ABP-computable coefficients in $\mathbb{F}(\bar{y})\langle X_2\rangle$, again of size $\mathrm{poly}(s)$. Therefore, Item (4) holds.

This completes the proof of Lemma 5.1. □

Remark 5.38. *To measure the ABP complexity in the Lemma 5.1, we are not accounting the complexity of the coefficients. We have accounted the complexity of the coefficients in the size of Frege proof given in Section 4.* ◊

Proof of witness identities over $\mathbb{F}[\bar{y}]$

Now we are ready to prove Lemma 4.21. We first recall the statement.

Lemma 4.21. (ABP identity witnesses over $\mathbb{F}[\bar{y}]$) *Let A, $\overline{A_i}$ for every $i \in [d]$, be as described in Subsection 4.6.1. Further suppose that A computes the zero polynomial. Then there exist matrices $\lambda''_0, \lambda''_1, \ldots, \lambda''_{d-1}$ and $T''_1, T''_2, \ldots, T''_{d-1}$ with*

$$\lambda''_r \in \mathbb{F}[\bar{y}]\,\langle X_2\rangle^{w_r \times w_r}, \qquad T''_{r+1} \in \mathbb{F}[\bar{y}]\,\langle X_1 \sqcup X_2\rangle^{w_r \times w_{r+1}},$$

such that the following identities hold.

$$\lambda''_{d-i} \cdot \overline{A}_i = \overline{0} \qquad \forall\, i = 1, 2, \ldots, d. \tag{4.21}$$

$$\lambda''_{d-i} \cdot \overline{A}_i = T''_{d-i+1} \cdot \lambda''_{d-i+1} \cdot \overline{A}_{i-1} \qquad \forall\, i = 2, 3, \ldots, d. \tag{4.22}$$

Here, λ'_0 is a polynomial of degree $\mathrm{poly}(s,d)$, and the degree of the coefficients of each entry in the witness matrices λ''_i, and T''_j is at most $\mathrm{poly}(s,d)$. Moreover, the entries of λ''_i, T''_j are computable by $\mathrm{poly}(s)$-size ABP.

Proof. Using Lemma 5.1, we first compute matrices $\lambda'_i \in \mathbb{F}(\bar{y})\,\langle X_2\rangle$ for $i = 1, \ldots, d-1$ and $T'_i \in \mathbb{F}(\bar{y})\,\langle X_1 \sqcup X_2\rangle$ for $i = 1, \ldots, d-1$ such that we have the following equations,

$$\begin{aligned} \lambda'_{d-i}\bar{A}_i = 0 \quad \forall\, i = 1, \ldots, d, \\ \lambda'_{d-i}\bar{A}_i = T'_{d-i+1}\lambda'_{d-i+1}\bar{A}_{i-1} \quad \forall\, i = 1, \ldots, d. \end{aligned} \tag{5.39}$$

Moreover, all coefficients appearing in the witness matrices $\{\lambda'_r, T'_{r+1}\}$ are rational functions in y of the form $p(y)/q(y)$ with $\deg p, \deg q \le \mathrm{poly}(s,d)$.

We now show how to clear the denominators.

Step-wise clearing (layer by layer). Fix $i \in \{2,3,\ldots,d\}$. Over $\mathbb{F}(y_1,\ldots,y_{k-2})$, we have the transition identity:

$$\lambda'_{d-i} \cdot \overline{A}_i \;=\; T'_{d-i+1} \cdot \lambda'_{d-i+1} \cdot \overline{A}_{i-1}. \tag{5.40}$$

Each entry of λ'_{d-i} and of $T'_{d-i+1}\lambda'_{d-i+1}$ is a rational function in y (with coefficients in $\mathbb{F}$) and hence may have denominators. Let $\bar{y} = \{y_1,\ldots,y_{k-2}\}$.

Let $p_i(\bar{y}), q_i(\bar{y}) \in \mathbb{F}[\bar{y}] \setminus \{0\}$ be the common multiple of all denominator appearing in λ'_i, T'_i respectively. We will define the λ''_i, T''_i matrices starting from $i = d-1$.

Base case : We have the equation $\lambda'_{d-1} \cdot \bar{A}_1 = 0$. This implies $p_{d-1}\lambda'_{d-1} \cdot \bar{A}_1 = 0$. Define $\lambda''_{d-1} := p_{d-1} \cdot \lambda'_{d-1} \in \mathbb{F}[\bar{y}]\,\langle X_2\rangle^{w_{d-1}\times w_{d-1}}$. Next, we want to define T''_{d-1}. Note that we have the following equation $\lambda'_{d-2}\bar{A}_2 = T'_{d-1}\lambda'_{d-1}\bar{A}_1$. Hence, multiplying both side by p_{d-1}, we get $p_{d-1}\lambda'_{d-2} \cdot \bar{A}_2 = T'_{d-1}(p_{d-1}\lambda'_{d-1}) \cdot \bar{A}_1$ and clearing the denominators of λ'_{d-2}, T'_{d-1}, we get that $(q_{d-1}p_{d-2}p_{d-1}\lambda'_{d-2}) \cdot \bar{A}_2 = (q_{d-1}p_{d-2}T'_{d-1})\lambda''_{d-1} \cdot \bar{A}_1$.

Define $T''_{d-1} = p_{d-2}(q_{d-1}T'_{d-1}) \in \mathbb{F}[\bar{y}]\,\langle X_1 \sqcup X_2\rangle^{w_{d-1}\times w_d}$ and $\lambda''_{d-2} = (q_{d-1}p_{d-2}p_{d-1}\lambda'_{d-2}) \in \mathbb{F}[\bar{y}]\,\langle X_2\rangle^{w_{d-2}\times w_{d-2}}$. Since $\lambda'_{d-2} \cdot \bar{A}_2 = 0$, $\lambda''_{d-2} \cdot \bar{A}_2 = 0$. Further note that multiplying by some polynomials in $\mathbb{F}[\bar{y}]$ does not increase the size of the ABP of each entry of λ''_{d-1} and T''_{d-1}.

Inductive case : Assume we have constructed λ''_{d-i} such that $\lambda''_{d-i} \cdot \bar{A}_i = 0$ and

$$\lambda''_{d-i} = \underbrace{(\prod_{j=1}^{i} p_{d-j} \prod_{j=1}^{i-1} q_{d-j})}_{P_{d-i}} \cdot \lambda'_{d-i}.$$

We have the equation $\lambda'_{d-i-1} \cdot \bar{A}_{i+1} = T'_{d-i} \cdot \lambda'_{d-i}\bar{A}_i$. Hence, multiplying both side by P_{d-i}, $P_{d-i}\lambda'_{d-i-1} \cdot \bar{A}_{i+1} = T'_{d-i} \cdot (P_{d-i} \cdot \lambda'_{d-i})\bar{A}_i$. Clearing the denominators of $\lambda'_{d-i-1}, T'_{d-i}$, we obtain the following equation: $(q_{d-i}p_{d-i-1}P_{d-i}\lambda'_{d-i-1}) \cdot \bar{A}_{i+1} = (q_{d-i}p_{d-i-1}T'_{d-i})\lambda''_{d-i} \cdot \bar{A}_i$.

We define $\lambda''_{d-i-1} \in \mathbb{F}[\bar{y}]\,\langle X_2\rangle^{w_{d-i-1}\times w_{d-i-1}}$ and $T''_{d-i} \in \mathbb{F}[\bar{y}]\,\langle X_1 \sqcup X_2\rangle^{w_{d-i}\times w_{d-i+1}}$ as follows:

$$\lambda''_{d-i-1} := (q_{d-i}p_{d-i-1}P_{d-i}\lambda'_{d-i-1}); \qquad T''_{d-i} := (q_{d-i}p_{d-i-1}T'_{d-i}).$$

Clearly the equation $\lambda''_{d-i-1} \cdot \bar{A}_{i+1} = 0$ is satisfied for this choice. So we can inductively define the following matrices,

$$\lambda''_{d-i} := (\prod_{j=1}^{i} p_{d-j} \prod_{j=1}^{i-1} q_{d-j}) \cdot \lambda'_{d-i}; \qquad T'_{d-i} := (q_{d-i}p_{d-i-1}T'_{d-i}); \qquad \lambda''_0 = (\prod_{j=1}^{d} p_{d-j} \prod_{j=1}^{d-1} q_{d-j}) \tag{5.41}$$

Note that both the matrices in the above equation has the degree of the coefficients st most $\mathrm{poly}(s,d)$ and each entry of λ''_i, T''_j is computable by $\mathrm{poly}(s)$ size ABP. □

References

[ACM24] Vikraman Arvind, Abhranil Chatterjee, and Partha Mukhopadhyay. Trading Determinism for Noncommutativity in Edmonds' Problem. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*, pages 539–559. IEEE, 2024.

[AJ25] V. Arvind and Pushkar S. Joglekar. On Efficient Noncommutative Polynomial Factorization via Higman Linearization, 2025. Pre-print available at `arXiv:2202.09883`.

[CF69] P. Cartier and D. Foata. Problémes combinatoires de commutation et réarrangements. *Lecture Notes in Mathematics*, 1969.

[CM23] Abhranil Chatterjee and Partha Mukhopadhyay. The Noncommutative Edmonds' Problem Re-visited. *CoRR*, abs/2305.09984, 2023. Pre-print available at `arXiv:2305.09984`.

[Coh95a] P. M. Cohn. *Skew Fields: Theory of General Division Rings*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1995.

[Coh95b] P. M. Cohn. Skew Fields: Theory of General Division Rings. In *Encyclopedia of Mathematics and its Applications 57*, 1995.

[Coh97] P. M. Cohn. The universal skew field of fractions of a tensor product of free rings. *Colloquium Mathematicae*, 72(1):1–8, 1997.

[CR79] Stephen A. Cook and Robert A. Reckhow. The Relative Efficiency of Propositional Proof Systems. *Journal of Symbolic Logic*, 44(1):36–50, 1979.

[FR04] Marc Fortin and Christophe Reutenauer. COMMUTATIVE/NONCOMMUTATIVE RANK OF LINEAR MATRICES AND SUBSPACES OF MATRICES OF LOW RANK. *Séminaire Lotharingien de Combinatoire [electronic only]*, 52, 01 2004.

[FSTW21] Michael A. Forbes, Amir Shpilka, Iddo Tzameret, and Avi Wigderson. Proof Complexity Lower Bounds from Algebraic Circuit Complexity. *Theory Comput.*, 17:1–88, 2021.

[GGdOW20] Ankit Garg, Leonid Gurvits, Rafael Mendes de Oliveira, and Avi Wigderson. Operator Scaling: Theory and Applications. *Found. Comput. Math.*, 20(2):223–290, 2020.

[GP18] Joshua A. Grochow and Toniann Pitassi. Circuit Complexity, Proof Complexity, and Polynomial Identity Testing: The Ideal Proof System. *J. ACM*, 65(6):37:1–37:59, 2018.

[Gro23] Joshua A. Grochow. Polynomial Identity Testing and the Ideal Proof System: PIT is in NP if and only if IPS can be p-simulated by a Cook–Reckhow proof system. arXiv:2306.02184 [cs.CC], 2023.

[HK91] Tero Harju and Juhani Karhumäki. The Equivalence Problem of Multitape Finite Automata. *Theor. Comput. Sci.*, 78(2):347–355, 1991.

[IKQS15] Gábor Ivanyos, Marek Karpinski, Youming Qiao, and Miklos Santha. Generalized Wong sequences and their applications to Edmonds' problems. *J. Comput. Syst. Sci.*, 81(7):1373–1386, 2015.

[IQS17] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Non-commutative Edmonds' problem and matrix semi-invariants. *Comput. Complex.*, 26(3):717–763, 2017.

[IQS18] Gábor Ivanyos, Youming Qiao, and K. V. Subrahmanyam. Constructive non-commutative rank computation is in deterministic polynomial time. *Comput. Complex.*, 27(4):561–593, 2018.

[KVV20] Igor Klep, Victor Vinnikov, and Jurij Volčič. Multipartite Rational Functions. *Documenta Math.*, 25:1285–1313, 2020.

[LTW18] Fu Li, Iddo Tzameret, and Zhengyu Wang. Characterizing Propositional Proofs as Noncommutative Formulas. *SIAM J. Comput.*, 47(4):1424–1462, 2018.

[Nis91] Noam Nisan. Lower Bounds for Non-Commutative Computation (Extended Abstract). In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing, May 5-8, 1991, New Orleans, Louisiana, USA*, pages 410–418, 1991.

[Raz13] Ran Raz. Tensor-Rank and Lower Bounds for Arithmetic Formulas. *J. ACM*, 60(6):40:1–40:15, 2013.

[Rec76] Robert A. Reckhow. *On the lengths of proofs in the propositional calculus*. Ph.d. thesis, University of Toronto, 1976. Technical Report #87.

[RS59] Michael O. Rabin and Dana S. Scott. Finite Automata and Their Decision Problems. *IBM J. Res. Dev.*, 3(2):114–125, 1959.

[RS05] Ran Raz and Amir Shpilka. Deterministic polynomial identity testing in non-commutative models. *Computational Complexity*, 14(1):1–19, 2005.

[Wor13] James Worrell. Revisiting the Equivalence Problem for Finite Multitape Automata. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II*, pages 422–433, 2013.

A Remaining details from Section 3

A.1 Boolean Solutions Exist if and only if Matrix Solutions Exist

Claim 3.5. *Let* $\{f_1 = 0, \ldots, f_m = 0 \; : \; f_i \in \mathbb{F}\langle X_1 \sqcup \cdots \sqcup X_p \rangle\}$ *(with* $|X_j| = n$ *for every* $j \in [q]$ *and* $|X_j| = 1$ *for every* $j \in [q+1, p]$*) be a set of equations that include the* boolean axioms *and the* commutator axioms*. Then the given system has a common solution in* $\mathbb{F}^{(n\times q)+(p-q)}$ *if and only if there exist* $a_{q+1}, \ldots a_p \in \mathbb{F}$ *and* $d \geq 1$*, set of matrices* $\{A_{1,1}, \ldots, A_{q,n}\} \subseteq \mathbb{F}^{d\times d}$ *which satisfy the given system.*

Proof. Suppose the given system has a common solution in $\mathbb{F}^{(n\times q)+(p-q)}$, say $\overline{a} = (a_{1,1}, \ldots, a_{q,n}, a_{q+1}, a_{q+2}, \ldots, a_p)$. Taking $d = 1$ and $A_{ij} = a_{ij}$ for every $(i,j) \in [q] \times [n]$, we trivially get the required statement.

For the converse direction, we give the full proof for $q = 1$ and observe that the proof follows analogously for larger q. Let $\overline{a} = (a_2, \ldots, a_p) \in \mathbb{F}^{p-1}$ and the set of $d \times d$ matrices $\overline{A} = (A_1, \ldots, A_n) \in (\mathbb{F}^{d\times d})^n$ be a common solution to the given system of equations.

Since the set of equations includes boolean axioms, $A_i^2 = A_i$ for every $i \in [n]$, the minimal polynomial of A_i must divide $x_i(x_i - 1)$. Thus the minimal polynomial has roots $0, 1$ with multiplicity 1, implying that each A_i is diagonalizable, with diagonal entries $0, 1$.

Moreover, the matrices satisfy the commutator axioms. That is, for every $i \neq j$, $A_iA_j = A_jA_i$. This implies that the matrices are simultaneously diagonalizable. That is, there exists an invertible matrix $R \in \mathbb{F}^{d\times d}$ such that for every i, $RA_iR^{-1} = D_i = \mathrm{Diag}(a_1^{(i)}, a_2^{(i)}, \ldots, a_d^{(i)})$ is a diagonal matrix with diagonal entries $a_j^{(i)} \in \{0, 1\}$. Observe that, for every $i \in [m]$,

$$f_i(A_1, \ldots, A_n, a_{2,1}, \ldots a_{p,1}) = 0 \implies R \cdot f_i(\overline{A}, \overline{a}) \cdot R^{-1} = 0 \implies f_i(RA_1R^{-1}, \ldots, RA_nR^{-1}, \overline{a}) = 0.$$

Now for every monomial over X_1 in f_i, say $m = \prod_{j\in T} x_j \in f_i$ for some $T \subseteq [n]$,

$$R \cdot m(\overline{A}) \cdot R^{-1} = R(\prod_{j\in T} A_j)R^{-1} = \prod_{j\in T}(R \cdot A_j \cdot R^{-1}).$$

Thus, for every $i \in [m]$, $f_i(D_1, \ldots, D_n, \overline{a}) = 0$ where each $D_j = \mathrm{Diag}(a_1^{(j)}, a_2^{(j)}, \ldots, a_d^{(j)})$ with $a_j^{(j)} \in \{0, 1\}$. Here, the left hand side is a $d \times d$ dimensional 0 matrix, with the (t,t)-th entry being $f_i(a_1^{(t)}, a_2^{(t)}, \ldots, a_n^{(t)}, \overline{a})$. Since this is true for every f_i, we obtain a set of solutions over $\mathbb{F}$ to the system of equations, namely $\left\{(a_1^{(t)}, \ldots, a_n^{(t)}, a_2, \ldots, a_p) \; : \; t \in [d]\right\}$.

For $q > 1$ the proof follows analogously since the matrices corresponding to different buckets commute (by definition) and we have commutator axioms for variables within each bucket. Hence all the matrices from the solution set mutually commute, and this gives us a unique invertible matrix R that diagonalizes every matrix. This proves the claim. $\square$

A.2 $\mathsf{PC}_{\mathsf{p,q}}$- IPS simulating *Frege*

Lemma 3.8. ($\mathsf{PC}_{\mathsf{p,q}}$- IPS simulates Schoenfield-Frege: analogue of [LTW18, Lemma 3.3]) *Let $\mathbb{F}$ be a field. Further, let $X = X_1 \sqcup \cdots \sqcup X_p$ be partially commutative variables where, for some $0 \le q \le p$, $|X_i| = n$ for every $i \in [q]$ and $|X_i| = 1$ for every $i \in [q+1, p]$. Here the variables inside each X_i are noncommuting and variables from different buckets commute.*

Let π be a tree-like *Frege (Schoenfield) proof of a propositional formula T, defined over X, from assumptions $\{F_1, \ldots, F_m\}$, and let the proof-lines be $\ell_1, \ell_2, \ldots, \ell_s$. Let $\mathrm{Tr}'(\cdot)$ be the algebraic translation from Boolean formulas to $\mathbb{F}\langle X \rangle$ as in Definition 3.6. For each $i \in [N]$ define the partially commutative algebraic translation $L_i := \mathrm{Tr}'(\ell_i) \in \mathbb{F}\langle X \rangle$ and let $F := \big(\mathrm{Tr}'(F_1), \ldots, \mathrm{Tr}'(F_m)\big)$.*

*Finally, let B be the set of all **Boolean axioms**: $x(x-1)$ for every $x \in X$ and let C be the set of all **partial commutator axioms**: $xy - yx$ for every $x, y \in X_i$, $i \in [p]$. Then for every $i \in [s]$, there exists a partially commutative algebraic formula $\Phi_i(X, Y, Z, W) \in \mathbb{F}\langle X, Y, Z, W \rangle$, where $Y = (y_1, \ldots, y_m)$ and Z, W are placeholder vectors for the Boolean and commutator axioms, such that:*

1. *$\Phi_i(X, \overline{0}, \overline{0}, \overline{0}) = 0$.*
2. *$\Phi_i(X, F, B, C) = L_i$.*
3. *$|\Phi_i| \le \Big(\sum_{\ell \in A_i} |L_\ell|\Big)^4$, where $A_i \subseteq [s]$ is the set of Frege proof-lines involved in deriving ℓ_i in the tree-like proof π (i.e., the indices of the sub-tree rooted at ℓ_i).*

In particular, Φ_s is a $\mathsf{PC}_{\mathsf{p,q}}$- IPS *proof of $\mathrm{Tr}'(T)$ from assumptions $\{\mathrm{Tr}'(F_1), \ldots, \mathrm{Tr}'(F_m)\}$, and its size is* $\mathrm{poly}(|\pi|)$.

Proof. Fix the (tree-like) proof π and its sequence of lines $\ell_1, \ldots, \ell_N$. As in [LTW18], we build Φ_i by induction on the derivation of ℓ_i.

Placeholders. The formula $\Phi_i(X, Y, Z, W)$ is a partially commutative formula in the variables X and in placeholder variables Y, Z, W, where:

- Y will be substituted by the translated assumptions $\mathrm{Tr}'(F_j)$;
- Z will be substituted by the Boolean axioms $x(x-1)$;
- W will be substituted by the commutator axioms $xy - yx$ for x, y in the same bucket.

Condition (1) says Φ_i vanishes when all placeholders are set to 0, and (2) says that after substitution we obtain exactly L_i.

Base cases. **(B1) ℓ_i is an assumption.** Suppose $\ell_i = F_j$ for some $j \in [m]$. Then $L_i = \mathrm{Tr}'(F_j)$. Define

$$\Phi_i(X, Y, Z, W) := y_j.$$

Clearly, $\Phi_i(X, \overline{0}, \overline{0}, \overline{0}) = 0$ and $\Phi_i(X, F, B, C) = \mathrm{Tr}'(F_j) = L_i$.

(B2) ℓ_i **is a Schoenfield axiom instance.** We treat explicitly the axiom scheme $A \to (B \to A)$. Note that $A \to (B \to A) \equiv \neg A \vee (\neg B \vee A)$, and hence by the definition of $\mathrm{Tr}'(\cdot)$,

$$L_i = \mathrm{Tr}'\big(\neg A \vee (\neg B \vee A)\big) = \mathrm{Tr}'(\neg A) \cdot \mathrm{Tr}'(\neg B \vee A) = (1 - a) \cdot \big((1 - b) \cdot a\big),$$

where $a := \mathrm{Tr}'(A)$ and $b := \mathrm{Tr}'(B)$. Thus,

$$L_i = (1 - a) \cdot (1 - b) \cdot a.$$

By Lemma 3.6 in [LTW18], there exists a partially commutative formula $\Psi_a(X, Z)$ such that

$$\Psi_a(X, \overline{0}) = 0, \qquad \Psi_a(X, B) = (1 - a) \cdot a, \qquad |\Psi_a| \le |a|^2.$$

Multiplying on the left by $(1 - b)$ yields a certificate for

$$(1 - b) \cdot (1 - a) \cdot a.$$

Define

$$\Theta(X, Z) := (1 - b) \cdot \Psi_a(X, Z).$$

Then $\Theta(X, \overline{0}) = 0$ and $\Theta(X, B) = (1 - b) \cdot (1 - a) \cdot a$.

Now swap the first two factors to obtain $(1 - a) \cdot (1 - b) \cdot a$. Using the analogue of Lemma 3.4 from [LTW18] with only intra-bucket commutator axioms, there exists a partially commutative formula $\varphi_{M,N}(X, W)$ such that

$$\varphi_{M,N}(X, \overline{0}) = 0, \qquad \varphi_{M,N}(X, C) = M - N, \qquad |\varphi_{M,N}| \le |M|^2 |N|^2,$$

where

$$M := (1 - b) \cdot (1 - a) \cdot a, \qquad N := (1 - a) \cdot (1 - b) \cdot a = L_i.$$

Finally define

$$\Phi_i(X, Y, Z, W) := \Theta(X, Z) \; + \; \varphi_{M,N}(X, W).$$

Then $\Phi_i(X, \overline{0}, \overline{0}, \overline{0}) = 0$ and after substitution we obtain

$$\Phi_i(X, F, B, C) = M + (M - N) = N = L_i.$$

The other Schoenfield axioms are handled analogously, as in [LTW18].

Induction step: modus ponens. Assume ℓ_i is derived by modus ponens from $\ell_j = A$ and $\ell_h = A \to B$, with $j, h < i$. By the definition of $\mathrm{Tr}'(\cdot)$, since $A \to B \equiv \neg A \vee B$, we have

$$L_j = \mathrm{Tr}'(A) = a, \qquad L_h = \mathrm{Tr}'(A \to B) = \mathrm{Tr}'(\neg A \vee B) = (1-a) \cdot \mathrm{Tr}'(B) = (1-a) \cdot c,$$

where $c := \mathrm{Tr}'(B)$. The conclusion is $\ell_i = B$, hence $L_i = c$.

By induction, we have Φ_j, Φ_h satisfying (2)-(3) in Definition 1.1 . Define

$$\Phi_i(X, Y, Z, W) \ := \ \Phi_j(X, Y, Z, W) \cdot \mathrm{Tr}'(B) \ + \ \Phi_h(X, Y, Z, W).$$

Then $\Phi_i(X, \overline{0}, \overline{0}, \overline{0}) = 0$ and after substituting $(Y, Z, W) = (F, B, C)$ we obtain

$$\begin{aligned} \Phi_i(X, F, B, C) &= \mathrm{Tr}'(A) \cdot \mathrm{Tr}'(B) \ + \ (1 - \mathrm{Tr}'(A)) \cdot \mathrm{Tr}'(B) \\ &= \big(\mathrm{Tr}'(A) + (1 - \mathrm{Tr}'(A))\big) \cdot \mathrm{Tr}'(B) \\ &= 1 \cdot \mathrm{Tr}'(B) \\ &= \mathrm{Tr}'(B) = L_i. \end{aligned}$$

The size bound follows exactly as in [LTW18] using tree-likeness and the fourth-power potential.

Conclusion. Applying the construction to $i = N$ gives a partially commutative IPS proof of $L_N = \mathrm{Tr}'(T)$ from assumptions $\mathrm{Tr}'(F_1), \ldots, \mathrm{Tr}'(F_m)$ and the Boolean/partial-commutator axioms, with size $\mathrm{poly}(|\pi|)$. □

B Remaining Details from Section 4

B.1 Frege Proof for Homogenisation of Partially Commutative Formulas

Lemma 4.6. *Let F be a partially commutative formula of size s and depth $O(\log s)$ computing a polynomial in $\mathbb{F}\langle X_1 \sqcup \ldots \sqcup X_k \rangle$ where $\mathbb{F} = \mathbf{GF}(2)$. Moreover, let $F_{(d_1,\ldots,d_k)}$ be the homogeneous component of the polynomial computed by F with degree signature $(d_1, \ldots, d_k)$. Then there exists a $s^{O(\log s)}$ size Frege proof of*

$$\widetilde{F}(X) \longleftrightarrow \bigoplus_{(d_1,\ldots,d_k)} \widetilde{F}_{(d_1,\ldots,d_k)}. \tag{4.6}$$

Proof. We first show the homogenization of [Raz13] in the partially commutative setting; then, we show how *Frege* can prove it.

Raz's Homogenization in the Partially Commutative Setting Let $X = X_1 \sqcup X_2 \sqcup \cdots \sqcup X_k$ be the variable set. Since the size of the formula F is s, the degree of the polynomial computed by F is at most s. Furthermore, the individual degree from each bucket can be at most s. Given any node $u \in F$, denote the product depth of u by δ_u. We assume the depth of F is $O(\log s)$. The idea is to track the increment of any monomial's degree signature along the path from $u \to r$ with r being the root. This is done by considering all monotone non increasing functions $N_u := \left\{D^u : \{0,1,\ldots,\delta_u\} \to \{0,1,\ldots,s+1\}^k\right\}$. The number of such functions is $|N_u| \leq \binom{s+\delta_u+2}{\delta_u+1}^k \leq s^{O(k\log s)}$. So, first we construct the homogeneous formula F' with $|N_u|$ many nodes for every node $u \in F$. We identify 0 with the root r in the function $D^u \in N_u$. For $(i_1,\ldots,i_k) \in [0,1,\ldots,s+1]^k$, $D^u_{i_1,\ldots,i_k}$ denotes all the functions $D^u \in N_u$ such that $D^u(\delta_u) = (i_1,\ldots,i_k)$. We want to prove for every node $u \in F$,

$$\left(\bigoplus_{(d_1,\ldots,d_k)} \widetilde{F'}_{(u,D^u_{d_1,\ldots,d_k})}\right) \longleftrightarrow \widetilde{F}_u. \tag{B.1}$$

If we prove Equation B.1 for the root r, we prove Equation 4.7 within Frege.

Construction of F' and Frege proof If $u \in F$ is a leaf, then for every $D^u \in N_u$, the node (u, D^u) is a leaf. If u is labeled by a field element and $D^u(u) = \underbrace{(0,\ldots,0)}_{k}$, then (u, D^u) is labeled by the same field element, and if $D^u(u) \neq (0,\ldots,0)$ then (u, D^u) is labeled by 0. If u is labeled from X_1 (respectively X_i for $i \in [k]$), then (u, D^u) is labeled by the same variable only if $D^u(u) = (1,0,\ldots,0)$ (respectively $D^u(u) = (0,\ldots,1,\ldots,0)$). Otherwise, we label it by 0. In this case, we can clearly prove Equation B.1 within Frege.

Inductive Case : Let $u, v, w \in F$ such that $u = v + w$. For every $D^u \in N_u$, let $D^v \in N_v$ be the function that agrees with D^u on $\{0,1,\ldots,\delta_u\}$ and $D^v(\delta_v) = D^u(\delta_u)$. Similarly, consider such $D^w \in N_w$. Then we have the following connection $(u, D^u) = (v, D^v) + (w, D^w)$ in F'. This implies

$$\begin{aligned} &\widehat{F'}_{(u,D^u)} := \widehat{F'}_{(v,D^v)} + \widehat{F'}_{(w,D^w)} \\ &\text{If } D^u(\delta_u) = (i_1,\ldots,i_k) \text{ then we have the following,} \\ &\widehat{F'}_{(u,D^u_{i_1,\ldots,i_k})} := \widehat{F'}_{(v,D^v_{i_1,\ldots,i_k})} + \widehat{F'}_{(w,D^w_{i_1,\ldots,i_k})}. \end{aligned} \tag{B.2}$$

Hence, we have the following tautology,

$$\widetilde{F'}_{(u,D^u_{i_1,\ldots,i_k})} \longleftrightarrow \widetilde{F'}_{(v,D^v_{i_1,\ldots,i_k})} \bigoplus \widetilde{F'}_{(w,D^w_{i_1,\ldots,i_k})}. \tag{B.3}$$

By the inductive hypothesis,

$$\begin{aligned} &\widetilde{F}_v \leftrightarrow \bigoplus_{(i_1,\ldots,i_k)} \widetilde{F'}_{(v,D^v_{i_1,\ldots,i_k})} \qquad \widetilde{F}_w \leftrightarrow \bigoplus_{(i_1,\ldots,i_k)} \widetilde{F'}_{(w,D^w_{i_1,\ldots,i_k})} \\ \implies &\widetilde{F}_v \oplus \widetilde{F}_w \leftrightarrow \bigoplus_{(i_1,\ldots,i_k)} \left(\widetilde{F'}_{(v,D^v_{i_1,\ldots,i_k})} \oplus \widetilde{F'}_{(w,D^w_{i_1,\ldots,i_k})}\right) \qquad \text{we can prove} \\ \implies &\widetilde{F}_u \longleftrightarrow \bigoplus_{(i_1,\ldots,i_k)} \widetilde{F}_{(u,D^u_{i_1,\ldots,i_k})} \qquad \text{this follows from Equation B.3.} \end{aligned} \tag{B.4}$$

Let $u = v \times w$ and $D^u \in N_u$ be such that $D^u(u) = (d_1, \ldots, d_k)$.
Then, $(u, D^u) = \sum_{i_1=0}^{d_1} \sum_{i_2=0}^{d_2} \cdots \sum_{i_k=0}^{d_k} (v, D^v_{i_1,\ldots,i_k})(w, D^w_{d_1-i_1,\ldots,d_k-i_k})$. That is,

$$\begin{aligned} \widehat{F'}_{(u,D^u_{i_1,\ldots,i_k})} &:= \sum_{j_1=0}^{i_1} \sum_{j_2=0}^{i_2} \cdots \sum_{j_k=0}^{i_k} \widehat{F'}_{(v,D^v_{j_1,\ldots,j_k})} \widehat{F'}_{(w,D^w_{i_1-j_1,\ldots,i_k-j_k})} \\ \implies \widetilde{F'}_{(u,D^u_{i_1,\ldots,i_k})} &\leftrightarrow \oplus_{j_1=0}^{i_1} (\cdots (\oplus_{j_k=0}^{i_k} (\widetilde{F'}_{(v,D^v_{j_1,\ldots,j_k})} \wedge \widetilde{F'}_{(w,D^w_{i_1-j_1,\ldots,i_k-j_k})}) \cdots). \end{aligned} \tag{B.5}$$

$$\begin{aligned} \widetilde{F}_u \longleftrightarrow \widetilde{F}_v \wedge \widetilde{F}_w &\leftrightarrow \Big(\bigoplus_{\substack{(i_1,\ldots,i_k) \\ 0 \le i_1,\ldots,i_k \le s_v+1}} \widetilde{F'}_{(v,D^v_{i_1,\ldots,i_k})} \Big) \wedge \Big(\bigoplus_{\substack{(j_1,\ldots,j_k) \\ 0 \le j_1,\ldots,j_k \le s_w+1}} \widetilde{F'}_{(w,D^w_{j_1,\ldots,j_k})} \Big) \\ &\leftrightarrow \bigoplus_{\substack{(j_1,\ldots,j_k) \\ 0 \le j_1,\ldots,j_k \le s_u+1}} (\oplus_{j_1=0}^{i_1} \cdots (\oplus_{j_k=0}^{i_k} (\widetilde{F'}_{(v,D^v_{j_1,\ldots,j_k})} \wedge \widetilde{F'}_{(w,D^w_{i_1-j_1,\ldots,i_k-j_k})})) \cdots) \\ &\leftrightarrow \bigoplus_{\substack{(i_1,\ldots,i_k) \\ 0 \le i_1,\ldots,i_k \le s_u+1}} \widetilde{F'}_{(u,D^u_{i_1,\ldots,i_k})} \quad \text{follows from Equation B.5.} \end{aligned} \tag{B.6}$$

□

B.2 Definition of the Translation Map between ABP and Formula

We want to define a map

$$g : \{\text{nodes of } A_F\} \to \{\text{nodes of } F\} \cup \{\emptyset\}$$

where $g(u) = v$ if the polynomial computed by F at v is the polynomial computed between u and t in A_F. The definition is by induction.

Base case. Let F be the formula $u = x$. The corresponding ABP A_F consists of a single edge $s \xrightarrow{x} t$. We define $g(s) := u$ and $g(t) := \emptyset$.

Inductive case (addition). Let $F = G + H$, where u, v, w are the roots of F, G, H, respectively, with $u = v + w$. Let A_G, A_H be the corresponding ABPs, and let g_1, g_2 be the associated maps with $g_1(s) = v$ and $g_2(s) = w$. We abuse notation by denoting the source and sink nodes of both ABPs by s and t.

The ABP A_F is obtained using the parallel composition of A_G and A_H. The map g is defined by

$$g(u_i) = \begin{cases} u, & \text{if } u_i \text{ is the source node,} \\ g_1(u_i), & \text{if } u_i \in A_G, \\ g_2(u_i), & \text{if } u_i \in A_H, \\ \emptyset, & \text{if } u_i \text{ is the sink node.} \end{cases} \tag{B.7}$$

Inductive case (multiplication). If $F = G \cdot H$, then A_F is obtained by sequentially composing A_G and A_H. The source of A_F is the source of A_G, and the sink of A_G is identified with the source of A_H. The map g is defined as the union of g_1 and g_2, except at the merged node u_t, where we set

$$g(u_t) := g_2(s),$$

with s denoting the source of A_H.

B.3 D′ as defined in Subsection 4.5.2 is Well-Defined

Lemma 4.15. (Well-definedness of the refined operator D') *Let F be a partially commutative homogeneous constant-free formula over the variable set $X = X_1 \sqcup \ldots \sqcup X_k$, and let d be the number of layers of the ABP A_F (hence the maximum possible X_1-degree in the constructions). Let D' be defined inductively on sub-formulas of F and on a parameter $p \in \{0, 1, \ldots, d\}$ as in the construction of Subsection 4.5.2.*

Then for every sub-formula F_u of F, every node u of F_u, and every $p \in \{0, \ldots, d\}$, the expression $D'(F_u, u, p)$ is well-defined. Moreover, for every triple (F_u, u, p) for which $D'(F_u, u, p)$ is defined, the output is an induced formula of F_u (obtained from F_u by substituting some sub-formulas by constants $0, 1$ and by replacing some sub-formulas by their degree-refined induced parts).

Proof. We define $D'(F_u, u, p)$ by induction on the structure (depth) of the subformula F_u, and within the same F_u by induction on the parameter p.

Base case (leaves). If F_u is a leaf, i.e. $F_u = x$ for some $x \in X$, then $D'(x, u, p)$ is defined explicitly by the leaf rule (and is either x or 0). Hence it is well-defined and is trivially an induced formula of F_u.

Inductive step. Assume F_u is an internal gate, and that for every strict subformula $F_{u'} \subsetneq F_u$, for every node $u' \in F_{u'}$ and every $p' \in \{0, \ldots, d\}$, the value $D'(F_{u'}, u', p')$ is already well-defined and is an induced formula of $F_{u'}$.

There are two cases.

Case 1: $F_u = G + H$. If $u \in G$ then $D'(G + H, u, p) := D'(G, u, p) + 0$, and if $u \in H$ then $D'(G + H, u, p) := 0 + D'(H, u, p)$. By the induction hypothesis, $D'(G, u, p)$ and $D'(H, u, p)$ are already well-defined induced formulas of G and H respectively; therefore $D'(G + H, u, p)$ is well-defined. It is obtained from F_u by replacing the other summand by 0, hence it is an induced formula of F_u.

Case 2: $F_u = G \cdot H$**.** We must justify the multiplication rule, in particular the occurrence of the term $F^{*(p-q)}_{\text{root}(H)}$.

Let r_H denote the root node of the subformula H (so H itself is F_{r_H}). By definition, for any integer $t \in \{0, \ldots, d\}$ we set

$$F^{*(t)}_{r_H} \;:=\; D'(H, r_H, t).$$

This is well-defined by the induction hypothesis because H is a strict subformula of F_u.

Now consider the definition of $D'(F_u, u, p)$:

(i) If $u \in H$. Then we simply set

$$D'(G \cdot H, u, p) \;:=\; D'(H, u, p),$$

which is well-defined by the induction hypothesis and is an induced formula of H (hence of F_u).

(ii) If $u \in G$. We define

$$D'(G \cdot H, u, p) \;:=\; \sum_{q=0}^{p} D'(G, u, q) \cdot F^{*(p-q)}_{r_H} \;=\; \sum_{q=0}^{p} D'(G, u, q) \cdot D'(H, r_H, p-q).$$

Each summand is well-defined by the induction hypothesis: $D'(G, u, q)$ is defined because G is a strict sub-formula of F_u, and $D'(H, r_H, p-q)$ is defined because H is a strict sub-formula of F_u. Therefore the entire sum is well-defined.

Finally we argue that this output is an induced formula of F_u. Indeed, each product term

$$D'(G, u, q) \cdot D'(H, r_H, p-q)$$

is obtained from the product $G \cdot H$ by replacing the left factor G by the induced sub-formula $D'(G, u, q)$ and replacing the right factor H by the induced sub-formula $D'(H, r_H, p-q)$, while keeping the top multiplication gate. Thus each summand is an induced formula of F_u, and the sum of induced formulas is again an induced formula of F_u (since we only introduce $+$ gates above induced sub-formulas and do not create any new variable occurrences outside F_u).

This completes the inductive proof that $D'(F_u, u, p)$ is well-defined in all cases and always yields an induced formula of F_u. □

C Remaining details of Section Section 5

C.1 Proof of linear matrix factorization

We first recall Lemma 5.7 from Subsection 5.1.

Lemma 5.7. (Linear Matrix Factorization). *Suppose $\mathbb{F}' = \mathbb{F}(\bar{y})$ be the field of rational functions in $\bar{y}$ over $\mathbb{F}$ and $X = \{x_1, x_2, \ldots, x_n\}$. Let $L \in \mathbb{F}'\langle X\rangle^{\ell \times m}$ be a noncommutative linear matrix of rank*

$r < \min\{\ell, m\}$ and d be the maximum degree of the coefficients of the entries of L. Then we can construct linear matrices $L_1 \in \mathbb{F}'\langle X\rangle^{\ell\times r}$ and $L_2 \in \mathbb{F}'\langle X\rangle^{r\times m}$ such that $L = L_1 \cdot L_2$ in deterministic $\text{poly}(\ell, m, n)$ *time. Moreover, the coefficients of the entries of L_1 and L_2 have degree at most* $\text{poly}(\ell, m, d)$.

As it is already mentioned, we divide the proof into two parts. Given such a linear matrix L, we first do a finer analysis of [IQS18] to construct a nontrivial shrunk subspace. We then exploit this subspace to obtain a zero block [FR04], which in turn yields a factorization of the input matrix. In the following lemma, we first show the construction of a nontrivial subspace.

Lemma C.1. *Let $\mathbb{F}' = \mathbb{F}(\bar{y})$ be the field of rational functions in $y_1, \ldots, y_k$ over $\mathbb{F}$, and let $L = \sum_{i=1}^{n} A_i x_i \in \mathbb{F}'\langle X\rangle^{\ell\times m}$ be a noncommutative linear matrix with* $\text{ncrank}(L) = r < \min\{\ell, m\}$. *Suppose the degree of the coefficients appearing in the entries of L is bounded by d. Then we can find a basis of a $(m-r)$-shrunk subspace $T \subseteq \mathbb{F}'^m$ such that the degree of the coefficients is bounded by* $\text{poly}(n, \ell, m, d)$.

Proof. For $k = 0$, the proof follows from [IQS18, Theorem 1.5]. For $k \geq 1$, the result is proved using a finer analysis of [IQS18] and [IKQS15], in particular, [IQS18, Theorem 5.10].

Define a matrix space $\mathcal{A} = \langle A_1, A_2, \ldots, A_n\rangle$. For any integer d', let us denote the matrix space $\mathcal{A}^{\{d'\}} = \mathbb{F}^{d'\times d'} \otimes \mathcal{A}$. We first describe the key ideas of [IQS18] in an informal way. The proof of [IQS18, Theorem 5.10] is obtained through an algorithm that runs for at most $\min\{\ell, m\}$ stages. Without loss of generality, we assume $m \leq \ell$. At the i^{th} stage of the algorithm, it maintains a matrix $B \in \mathbb{F}'^{\ell d'\times m d'}$ in $\mathcal{A}^{\{d'\}}$ of rank rd' where $r \leq \min\{\ell, \text{m}\}$ and $d' \leq r+1$. We call such a matrix B as a witness of noncommutative rank r of L. The algorithm then gradually tries to find witness matrices of higher ranks via the *second Wong sequence*. When it is unable to do so, it computes a shrunk subspace witnessing the fact that L is not of full noncommutative rank. By clearing the denominator, assume that the entries of B are polynomials in $\bar{y}$ and the maximum degree is bounded by D_i.

Since for our purpose, it is enough to argue the degree bound in an existential way, we do not need to investigate at each step whether the algorithm can increase the rank in a constructive manner. In particular, [IQS18] constructs matrices $C_1, \ldots, C_\ell \in \mathcal{A}$ such that

$$C_\ell B^{-1} C_{\ell-1} B^{-1} \cdots C_1 B^{-1}(0) \not\subseteq \text{Im}(B).$$

Note that for a subspace V, we denote by $B^{-1}V$ the preimage of B. By linearity, we can obtain $A_{i_1}, A_{i_2}, \ldots, A_{i_\ell} \in \{A_{i'} \otimes E_{j'k'} \mid i' \in [n] \text{ and } 1 \leq j', k' \leq d'\}$ such that

$$A_{i_\ell} B^{-1} A_{i_{\ell-1}} B^{-1} \cdots A_{i_1} B^{-1}(0) \not\subseteq \text{Im}(B).$$

Here $E_{j'k'}$ is an elementary matrix of dimension d'. Next, the algorithm in [IQS18] constructs a set of matrices $Z_1, \ldots, Z_n$ only using a basis of $\mathbb{F}'^{d''}$ (for a suitable integer d''). The dimension of the Z_i matrices is d''. Now construct the matrix $C' = A_{i_1} \otimes Z_1 + \cdots + A_{i_\ell} \otimes Z_\ell$. This step does not incur any degree blow-up. Define $B' = B \otimes I_{d''}$. If the rank of C' is already more than $rd'd''$, then the algorithm moves into the next subroutine with an update $B'' \leftarrow C'$. Otherwise it finds a shift $\lambda \in \mathbb{F}(\bar{y})$ such that the rank of $B' + \lambda C'$ is more than $rd'd''$. The computation of such a λ depends on avoiding the roots of a polynomial-degree polynomial in λ (essentially we need to ensure the invertibility of a submatrix). Thus we can choose a polynomial in $\mathbb{F}[\bar{y}]$ of polynomial degree for λ. This step incurs a degree blow-up from D_i to $D_i + \text{poly}(n, d, d', d'')$.

The next important steps are rounding and blow-up control. A self-contained and somewhat simpler description can be found in [CM23]. The rounding step ensures to produce a rank witness of rank $(r+1)d'd''$ (in other words, rounding to the next possible rank). This step uses concepts from cyclic division algebra. The initial witness matrices will be expressed as a linear combination of the generators of a suitable cyclic division algebra. The coefficients will be over $\mathbb{F}(\bar{y}, \omega)$ for a chosen root of unity adjoined with $\mathbb{F}$. Then gradually the coefficients will be chosen from the ground field $\mathbb{F}(\bar{y})$. This step again amounts to avoiding roots of univariate polynomials of polynomial-degree. Thus incurring $\text{poly}(n, d, d', d'')$ additive blow up in the degree of $\bar{y}$. The blow-up control steps bring down the blown-up dimension to at most m and again requires rounding using cyclic division algebra incurring $\text{poly}(n, d, d', d'')$ blow-up in the degree of y. So the final degree bound will be at most $D_i + (n, m, d)^{O(1)}$ which is bounded by $(n, m, d)^{O(1)}$.

Now when the algorithm can not increase the rank any more, the second Wong sequence computation outputs a shrunk subspace of the matrix space $\mathcal{A}^{\{d\}}$. Let the matrix obtained at this step is B and the goal is to obtain a basis of a shrunk subspace of dimension co-rank(B). Here we need to use the concept of pseudo-inverse of a matrix [IKQS15, Lemma 10]. When at the i^{th} step the rank can not be improved any more, we obtain subspaces $W_i = W_{i+1}$ and $W_i = (\mathcal{A}^{\{d\}} B')^i \ker(BB')$ where B' is the pseudo-inverse of B. Then the required shrunk subspace is $B'W_i$. Since the construction of pseudo-inverse involves simple linear algebraic operations, the final degree bound will be $(n, m, d)^{O(1)}$. □

The second part of the proof uses the shrunk subspace obtained to compute the factors. We do this again in two parts. First, we compute a nontrivial decomposition, which then implies the factors.

Lemma C.2. *Let $L = \sum_{i=1}^{n} A_i x_i \in \mathbb{F}[\bar{y}]\langle X \rangle^{\ell \times m}$ be a noncommutative linear matrix with* $\text{ncrank}(L) = r < \min\{\ell, m\}$. *Let $T \subseteq \mathbb{F}(\bar{y})^m$ be a $(m-r)$-shrunk subspace corresponding to L. Then, given a basis of T, one can construct invertible matrices $U \in \mathbb{F}(\bar{y})^{\ell \times \ell}$ and $V \in \mathbb{F}(\bar{y})^{m \times m}$ such that*

$$ULV = \left[\begin{array}{c|c} M & 0 \\ \hline B & C \end{array}\right],$$

where the zero block is of size $a \times b$ such that $a + b = \ell + m - r$.

Proof. We show the existence of invertible matrices $U \in \mathbb{F}(\bar{y})^{\ell \times \ell}$ and $V \in \mathbb{F}(\bar{y})^{m \times m}$ such that for every $i \in [n]$,

$$UA_iV = \left[\begin{array}{c|c} M_i & 0 \\ \hline B_i & C_i \end{array}\right]. \tag{C.3}$$

Let $\dim(T) = r'$ and let $B_T := \{u_1, \ldots, u_{r'}\} \subset \mathbb{F}(\bar{y})^m$ be a spanning set of T. We extend B_T to a basis of $\mathbb{F}(\bar{y})^m$ by adding $\{u_{r'+1}, \ldots, u_m\}$ to B_T and let $T' := \mathrm{span}_{\mathbb{F}(\bar{y})}\{u_{r'+1}, \ldots, u_m\}$. Thus $T \oplus T' = \mathbb{F}(\bar{y})^m$. We define the matrix $V := [u_1, \ldots, u_{r'}, u_{r'+1}, \ldots, u_m]$ to be the matrix whose columns are the vectors $\{u_1, \ldots, u_m\}$. Clearly, V is invertible over $\mathbb{F}(\bar{y})$.

Let W be a subspace of dimension $s < r'$ containing $\mathrm{span}_{\mathbb{F}(\bar{y})}\{\bigcup_{i=1}^n A_i \cdot T\}$.

Let $B_W := \{w_1, \ldots, w_s\} \subset \mathbb{F}(\bar{y})^\ell$ be a basis of W. We extend this basis to a basis of $\mathbb{F}(\bar{y})^\ell$ by adding $\{w_{s+1}, \ldots, w_\ell\}$ to B_W. Let U be the invertible matrix in $\mathbb{F}(\bar{y})^{\ell \times \ell}$ such that $U := [w_1, \ldots, w_\ell]^{-1}$. Assume the maximum degree (numerator and denominator) of each entry in V and U^{-1} is at most d. Then, the maximum degree between numerator and denominator of each entry is at most $O(\ell d)$.

Consider the matrix $A_i \cdot V$. For every $j \leq r'$, the j-th column of this matrix is $A_i \cdot u_j \in W$ by the definition of the shrunk subspace. Hence, for every $j \leq r'$, there exist $\alpha_1, \ldots, \alpha_s \in \mathbb{F}(\bar{y})$ such that $A_i u_j = \sum_{i=1}^s \alpha_i \cdot w_i$. Formally,

$$U^{-1} \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_s \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \left[\begin{array}{c|c|c|c|c} w_1 & w_2 & w_3 & \ldots & w_k \end{array}\right] \cdot \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_s \\ 0 \\ \vdots \\ 0 \end{pmatrix} = A_i u_j \implies U \cdot (A_i \cdot u_j) = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_s \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{C.4}$$

Thus, the last $\ell - s$ positions in $U \cdot (A_i u_j)$ are 0 for every $j \leq r'$.

Therefore, for every $i \in [n]$ we obtain

$$UA_iV = \left[\begin{array}{c|c} * & * \\ \hline 0 & * \end{array}\right], \tag{C.5}$$

where the zero block is of size $(\ell - s) \times r'$. Furthermore, T is a $(m - r)$-shrunk subspace of dimension r', therefore, $r' - s = m - r$. Hence, $\ell - s + r' = \ell + m - r$.

By further row and column operations, we obtain invertible matrices $U' \in \mathbb{F}(\bar{y})^{\ell \times \ell}$ and $V' \in \mathbb{F}(\bar{y})^{m \times m}$ such that

$$U'A_iV' = \left[\begin{array}{c|c} M_i & 0 \\ \hline B_i & C_i \end{array}\right] \quad \text{for every } i \in [n].$$

Consequently,

$$U'LV' = U'\left(\sum_i x_i A_i\right)V' = \left[\begin{array}{c|c} M & 0 \\ \hline B & C \end{array}\right]. \qquad \square$$

We now show that given the r-decomposability form Equation 5.6, the linear factors are easy to compute.

Corollary C.6. *Let $L = \sum_{i=1}^{n} x_i \cdot A_i \in \mathbb{F}(\bar{y})\langle X\rangle^{\ell\times m}$ be a noncommutative linear matrix with* $\text{ncrank}(L) = r$. *Then L admits a nontrivial factorization*

$$L = L_1 \cdot L_2,$$

where the factorization follows from the r-decomposability of L.

Proof. Since L is r-decomposable, we have invertible matrices $U \in \mathbb{F}(\bar{y})^{\ell\times\ell}$ and $V \in \mathbb{F}(\bar{y})^{m\times m}$ such that

$$U \cdot L \cdot V = \left[\begin{array}{c|c} M & 0 \\ \hline B & C \end{array}\right],$$

where the zero block is of size $i \times j$ such that $i + j = \ell + m - r$. Therefore, we can write:

$$L = U^{-1}\left[\begin{array}{c|c} M & 0 \\ \hline B & I_i \end{array}\right] \cdot \left[\begin{array}{c|c} I_j & 0 \\ \hline 0 & C \end{array}\right] \cdot V^{-1}$$

$$\text{where} \quad L_1 = U^{-1}\left[\begin{array}{c|c} M & 0 \\ \hline B & I_i \end{array}\right] \quad \text{and} \quad L_2 = \left[\begin{array}{c|c} I_j & 0 \\ \hline 0 & C \end{array}\right] \cdot V^{-1}. \qquad \square$$

We are now ready to prove Lemma 5.7.

Proof of Lemma 5.7. The proof combines the equivalent definitions in Theorem 5.5. If the linear matrix $L \in \mathbb{F}'\langle X\rangle$ has *low rank*, then it admits a shrunk subspace, and by Lemma C.2 one can choose a basis of this subspace with polynomially bounded y-degree. This yields an r-decomposition of L, and the desired factorization follows directly from Corollary C.6. $\square$

C.2 Toy Example for the factorisation

Let M be a 2×3 matrix in which each entry is a product of 2 matrices.

$$M = \begin{pmatrix} AB & CD & EF \\ A'B' & C'D' & E'F' \end{pmatrix} \text{ such that } A_{p\times q}B_{q\times r}, A'_{p'\times q}B'_{q\times r}$$
$$C_{p\times k}, D_{k\times n}, C'_{p'\times k}D'_{k\times n} \text{ and } E_{p\times m}F_{m\times s}, E'_{p'\times m}F'_{m\times s} \text{ are linear matrices.} \qquad \text{(C.7)}$$

Claim C.8. *There exists $t \in \mathbb{N}$ such that $(M \oplus I_t) = P \cdot L \cdot Q$ where the matrix P (Q) is an upper triangular (lower triangular) matrix with diagonal entries are 1 and every entry of P and Q is computable by* $\mathrm{poly}(S)$*-sized ABPs. Here $S = \mathrm{poly}(p, q, r, p', k, n, m, s)$.*

Proof. Consider $M \oplus I_m$ first.

$$M \oplus I_m = \begin{pmatrix} AB & CD & EF & 0 \\ A'B' & C'D' & E'F' & 0 \\ 0 & 0 & 0 & I_m \end{pmatrix}$$

$$= \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{p'} & -E' \\ 0 & 0 & I_m \end{pmatrix} \cdot \begin{pmatrix} AB & CD & EF & 0 \\ A'B' & C'D' & 0 & E' \\ 0 & 0 & -F' & I_m \end{pmatrix} \cdot \begin{pmatrix} I_r & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \\ 0 & 0 & Is & 0 \\ 0 & 0 & F' & I_m \end{pmatrix}$$

Using the previous equation, we linearize the matrix block $E'F'$.

Now we linearize the block EF by applying the same process to the matrix $\begin{pmatrix} AB & CD & EF & 0 \\ A'B' & C'D' & 0 & E' \\ 0 & 0 & -F' & I_m \end{pmatrix} = M_1$ (say). Let $P_1 = \begin{pmatrix} I_p & 0 & 0 \\ 0 & I_{p'} & -E' \\ 0 & 0 & I_m \end{pmatrix}$ and $Q_1 = \begin{pmatrix} I_r & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \\ 0 & 0 & I_s & 0 \\ 0 & 0 & F' & I_m \end{pmatrix}$. Then $(M \oplus I_m \oplus I_m) = (P_1 \oplus I_m)(M_1 \oplus I_m)(Q_1 \oplus I_m)$. Now if we consider the matrix $M_1 \oplus I_m$, then we get

$$M_1 \oplus I_m = \begin{pmatrix} AB & CD & EF & 0 & 0 \\ A'B' & C'D' & 0 & E' & 0 \\ 0 & 0 & -F' & I_m & 0 \\ 0 & 0 & 0 & 0 & I_m \end{pmatrix} \underset{\text{row-column swipe}}{=} P' \cdot \underbrace{\begin{pmatrix} 0 & 0 & I_m & -F' & 0 \\ A'B' & C'D' & E' & 0 & 0 \\ AB & CD & 0 & EF & 0 \\ 0 & 0 & 0 & 0 & I_m \end{pmatrix}}_{M_1'} \cdot Q'$$

where P', Q' are permutation matrices. Hence both $(P_1 \oplus I_m) \cdot P'$ and $Q' \cdot (Q_1 \oplus I_m)$ are invertible.

So

$$M_1' = \begin{pmatrix} 0 & 0 & I_m & -F' & 0 \\ A'B' & C'D' & E' & 0 & 0 \\ AB & CD & 0 & EF & 0 \\ 0 & 0 & 0 & 0 & I_m \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} I_m & 0 & 0 & 0 \\ 0 & I_{p'} & 0 & 0 \\ 0 & 0 & I_p & -E \\ 0 & 0 & 0 & I_m \end{pmatrix}}_{P_1'} \cdot \underbrace{\begin{pmatrix} 0 & 0 & I_m & -F' & 0 \\ A'B' & C'D' & E' & 0 & 0 \\ AB & CD & 0 & 0 & E \\ 0 & 0 & 0 & -F & I_m \end{pmatrix}}_{M_2} \cdot \underbrace{\begin{pmatrix} I_r & 0 & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 & 0 \\ 0 & 0 & I_m & 0 & 0 \\ 0 & 0 & 0 & I_s & 0 \\ 0 & 0 & 0 & F & I_m \end{pmatrix}}_{Q_1'}$$

Let us denote $P_2 = (P_1 \oplus I_m) \cdot P' \cdot P_1'$ and $Q_2 = Q_1' \cdot Q' \cdot (Q_1 \oplus I_m)$. Now the idea is to linearize the block CD in M_2 by applying the same process to $M_2 \oplus I_k$. Note that at the end of this process, we obtain $t = q + k + m$. □