On the Tension Between Full-Rankness and Self-Reducibility for Set-Multilinear Polynomials

Deepanshu Kush
Department of Computer Science and Technology, University of Cambridge, UK
Sidney Sussex College, Cambridge, UK

Abstract

In this paper, we rule out a natural programme for proving $\mathsf{VF} \neq \mathsf{VNP}$ via set-multilinear formula lower bounds. The programme combines two ingredients present in the literature: the $n^{\Omega(\log n)}$ set-multilinear formula lower bound of Kush and Saraf (CCC 2022) for any *full-rank* polynomial, and an IMM-style self-reducibility that propagates such a bound to small degree, where Raz's set-multilinearisation (J. ACM 2013) converts it to a general formula lower bound. Each ingredient has been realised separately, yet no polynomial family is known to combine them. We prove that no such family can exist: under any polynomial-width IMM-style self-reducibility – i.e., a small-width expression $g = \sum_{k=1}^{w} L_k \cdot R_k$ with each summand factoring across a balanced split – full-rankness forces width $n^{\Omega(d)}$, and even approximate full-rankness across a near-balanced split forces width $n^{\Omega(\sqrt{d})}$. This rules out the full-rank/self-reducible route to $\mathsf{VF} \neq \mathsf{VNP}$.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory; Theory of computation → Circuit complexity

Keywords and phrases Algebraic formula lower bounds, set-multilinear formulas, iterated matrix multiplication, rank methods, hardness escalation, self-reducibility, barriers

Digital Object Identifier 10.4230/LIPIcs.MFCS.2026.69

Funding This research is supported by Tom Gur's ERC Starting Grant 101163189 and UKRI Future Leaders Fellowship MR/X023583/1.

1 Introduction

1.1 Algebraic complexity theory and formula lower bounds

Algebraic complexity theory, initiated by Valiant [21], studies the minimum number of field operations required to compute a multivariate polynomial. The three most studied models are *algebraic circuits* (directed acyclic graphs of $+$ and $\times$ gates over a field), *algebraic branching programs* (layered DAGs whose edges are labelled by variables or constants, computing the sum over source-to-sink paths of edge-label products), and *algebraic formulas* (circuits in which every gate has fan-out 1, i.e., the underlying DAG is a tree). The corresponding polynomial-size and polynomial-degree complexity classes fit into the chain

$$\mathsf{VF} \subseteq \mathsf{VBP} \subseteq \mathsf{VP} \subseteq \mathsf{VNP},$$

the algebraic analogue of the Boolean chain $\mathsf{NC}^1 \subseteq \mathsf{NL} \subseteq \mathsf{P} \subseteq \mathsf{NP}$. Separating any two of these classes is of central importance in algebraic complexity theory; the conjecture that all four are strictly distinct is known as *Valiant's hypothesis* [21]. We refer the reader to [19, 18] for precise definitions of these classes and related exposition. Our focus in this paper is the extremal separation $\mathsf{VF} \neq \mathsf{VNP}$. Proving this separation is equivalent to exhibiting an explicit polynomial (i.e., one in VNP) that requires super-polynomial algebraic formula size. One natural candidate is the *iterated matrix multiplication* polynomial:

$$\mathrm{IMM}_{n,d}(X_1, \ldots, X_d) := (X_1 X_2 \cdots X_d)_{1,1}, \tag{1}$$

where each X_i is an $n \times n$ matrix of distinct variables. Since IMM is in fact complete[1] for VBP, a super-polynomial formula lower bound for $\mathrm{IMM}_{n,n}$ would yield the stronger separation $\mathsf{VF} \neq \mathsf{VBP}$.

1.2 Set-multilinear formulas and the large-degree regime

A polynomial $P \in \mathbb{F}[X_1, \ldots, X_d]$ is *set-multilinear* with respect to the partition $(X_1, \ldots, X_d)$ if every monomial contains exactly one variable from each X_i. A circuit, formula, or ABP is *set-multilinear* if every internal gate computes a set-multilinear polynomial. The classes $\mathsf{smVF} \subseteq \mathsf{smVBP} \subseteq \mathsf{smVP} \subseteq \mathsf{smVNP}$ are defined analogously to their general counterparts. Throughout, we write $\mathcal{L}_{\mathrm{sm}}(f)$ for the minimum size of a set-multilinear formula computing a set-multilinear polynomial f. Set-multilinear models offer an alluring route to general formula lower bounds, via a "hardness escalation" strategy made possible by the following result of Raz.

▶ **Theorem 1** (Raz's set-multilinearisation [15]). *Let P be a set-multilinear polynomial over sets $X_1, \ldots, X_d$ with $|X_i| = n$ and $d \leq O(\log n / \log\log n)$. If P has a general algebraic formula of size n^c for some constant c, then P has a set-multilinear formula of size $n^{O(c)}$. Contrapositively, a super-polynomial lower bound on the set-multilinear formula size of P at such a degree yields a super-polynomial lower bound on general formula size for P.*

Raz's proof gives the explicit size bound $O((\Delta+2)^d \cdot s)$, where $\Delta = O(\log s)$ is the depth after Brent-style depth reduction and s is the general formula size [15, Theorem 3]. For this to be polynomial when $s = n^c$, one needs $d \cdot \log\log n = O(\log n)$, whence the threshold $d \leq C \log n / \log\log n$ for a constant C depending on c. For our purposes, we wish to invoke this transfer at the *largest* degree in the allowable range – that is, at the boundary $d = \Theta(\log n / \log\log n)$ – so that the propagation gap one must bridge from degree n down to the transfer point is as small as possible. We henceforth refer to this boundary as the *Raz transfer point*.

Why target the large-degree regime?

Theorem 1 reduces general formula lower bounds to set-multilinear formula lower bounds only in the small-degree regime $d = O(\log n / \log\log n)$. In this regime, super-polynomial set-multilinear formula lower bounds are known – but only against *low-depth* formulas. The breakthrough of Limaye, Srinivasan, and Tavenas [10] proves a super-polynomial lower bound on the size of product-depth Δ set-multilinear formulas computing an explicit low-degree polynomial – in fact, $\mathrm{IMM}_{n,d}$ for d in the Raz transfer range – for any Δ up to $\Theta(\log\log d)$. A key enabling component of their argument is a *depth-preserving* refinement of Raz's Theorem: they show that any product-depth Δ general formula computing a set-multilinear polynomial of low enough degree can be converted into a product-depth $O(\Delta)$ *set-multilinear* formula with only a polynomial blow-up. Combined with their set-multilinear lower bound, this yields super-polynomial lower bounds against low-depth *general* formulas in the small-degree regime. Taken together, these results show that studying set-multilinear formula lower bounds in the low-degree regime is highly consequential for algebraic complexity theory. What we now observe is that the same is true of the *large-degree regime* – for two distinct reasons. First, it is in the large-degree regime that we have the strongest known lower bounds

[1] Via p-projections. See [18] for a precise definition of such reductions.

against *unbounded-depth* set-multilinear formulas, which matches the unrestricted depth property of the class VF.[2] Second, this regime admits a *hardness-propagation* mechanism – IMM-style self-reducibility – that in principle transfers large-degree lower bounds down to the small-degree regime where Theorem 1 can take over; this is the subject of Section 1.3. We first elaborate on the unbounded-depth lower bounds. Kush and Saraf [7, 8] prove $n^{\Omega(\log n)}$ unbounded-depth set-multilinear formula lower bounds for explicit polynomial families in VNP (the Nisan–Wigderson design polynomial, see Appendix B.1), in VP (the Dyck-word polynomial of [8, Sec. 4.1], see Appendix B.2), and even in VBP [8][3] (see also [6] for an alternate construction). The key thing to note is that all these bounds hold only for $d = n^{\Omega(1)}$. For the IMM polynomial itself, Tavenas, Limaye, and Srinivasan [20] prove a $(\log n)^{\Omega(\log d)}$ unbounded-depth set-multilinear formula lower bound – weaker, but structurally significant for reasons we explain in Section 1.3; this bound too is, of course, non-trivial (and even super-polynomial, as it becomes $n^{\Omega(\log \log n)}$) only in the large-degree regime, i.e., for $d = n^{\Omega(1)}$. None of these lower bounds, taken in isolation, implies a general formula lower bound: the degree is far outside the range of Raz's Theorem. The missing ingredient is a mechanism to *propagate* a lower bound at large degree down to small degree, where Raz's transfer takes over. IMM admits exactly such a mechanism – a *polynomial-width midpoint decomposition* – which is the focus of the next subsection.

1.3 Why the large-degree regime: the role of self-reducibility

$\mathrm{IMM}_{n,d}$ admits a natural recursive decomposition, valid for every even d:

$$\mathrm{IMM}_{n,d}(X_1, \ldots, X_d) \;=\; \sum_{k=1}^{n} (X_1 \cdots X_{d/2})_{1,k} \cdot (X_{d/2+1} \cdots X_d)_{k,1}. \tag{2}$$

Each summand is a product of two $(1,k)$- *and* $(k,1)$-*variants* of $\mathrm{IMM}_{n,d/2}$ on disjoint variable sets. Each variant is computed by a set-multilinear formula of the same size (up to re-indexing the output entry) as $\mathrm{IMM}_{n,d/2}$ itself. Consequently:

$$\mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,d}) \;\le\; 2n \cdot \mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,d/2}), \tag{3}$$

where $\mathcal{L}_{\mathrm{sm}}(\cdot)$ denotes the minimum set-multilinear formula size (as above). Iterating (3) from $\mathrm{IMM}_{n,n}$ down to $\mathrm{IMM}_{n,1}$ halves the degree at each step, and after $\log n$ halvings yields $\mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,n}) \le (2n)^{\log n} \cdot \mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,1}) = n^{(1+o(1))\log n}$, recovering the familiar divide-and-conquer upper bound. Read contrapositively, iterating the recursion (3) from $\mathrm{IMM}_{n,n}$ down to $\mathrm{IMM}_{n,d}$ for any d dividing n (applying (3) $\log(n/d)$ times) propagates lower bounds:

$$\mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,d}) \;\ge\; \frac{\mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,n})}{(2n)^{\log(n/d)}}. \tag{4}$$

[2] By a Brent-style depth reduction [1], any size-s formula can be rebalanced to depth $O(\log s)$ at polynomial cost, so "unbounded-depth" here can equivalently be taken to mean depth $O(\log s)$. For polynomial size formulas, this can still be much larger than the depth $O(\log \log d)$ at which the LST bound applies.

[3] See [8] for the definition of this VBP polynomial. We emphasise that this does *not* imply a corresponding lower bound for IMM: although IMM is VBP-complete and so the VBP polynomial of [8] reduces to IMM by p-projections, this reduction destroys set-multilinearity, so set-multilinear lower bounds for this polynomial do not transfer to IMM.

On the strength required for propagation

The midpoint recursion (3) carries a factor of $2n$ at each halving, so propagating a lower bound from $d = n$ down to the Raz transfer point $d_0 = \Theta(\log n / \log\log n)$ via (4) incurs a multiplicative loss of $(2n)^{\log(n/d_0)}$. The precise strength of the large-degree bound needed to survive this loss depends on the particular self-reduction used to propagate; different self-reductions of IMM incur different losses.[4] Crucially, the barrier we prove in this paper is a *structural* obstruction that holds regardless of the quantitative strength of any lower bound for full-rank set-multilinear polynomials, so this comparison does not affect our conclusions; we therefore do not belabour it here. The current best lower bound on the set-multilinear formula size of any full-rank polynomial is the $d^{\Omega(\log d)}$ bound of Kush and Saraf [7], which at $d = n$ is $n^{\Omega(\log n)}$.

The core of the propagation

The identity (2) has two essential features that drive (3):

- **Bipartite structure:** the decomposition separates the blocks into two halves (here $\{1, \ldots, d/2\}$ and $\{d/2 + 1, \ldots, d\}$), with each summand a product of polynomials over these two disjoint halves.
- **Polynomial width:** the number of summands is polynomial in n (in (2), exactly n).

Together these constitute a structural property that can be stated abstractly, independently of IMM. Previous treatments (e.g., [20]) are informal accounts and work directly with the IMM identity (2) without extracting this property; isolating it cleanly lets us state a single barrier that applies to any set-multilinear polynomial family.

▶ **Definition 2** (S-decomposition). *Let $S \subseteq [d]$ be a subset of blocks, and write $\overline{S} := [d] \setminus S$. For a tuple of variable sets $(X_1, \ldots, X_d)$ and $T \subseteq [d]$, let $\overline{X}(T) := (X_i)_{i \in T}$ denote the sub-tuple indexed by T. An* S-decomposition of width w *of a set-multilinear polynomial $g \in \mathbb{F}_{\mathrm{sm}}[X_1, \ldots, X_d]$ is an expression*

$$g \;=\; \sum_{k=1}^{w} L_k \cdot R_k, \qquad L_k \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(S)],\; R_k \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(\overline{S})].$$

The minimum width *of an S-decomposition of g is the least w for which such an expression exists. The* midpoint case *is $S = \{1, \ldots, d/2\}$.*

The identity (2) is thus a midpoint S-decomposition of $\mathrm{IMM}_{n,d}$ of width n – the canonical instance of this notion. More generally, any S-decomposition with $|S| = d/2$ yields an analogous halving recursion $\mathcal{L}_{\mathrm{sm}}(g_{n,d}) \leq 2w \cdot \mathcal{L}_{\mathrm{sm}}(g_{n,d/2})$, provided the L_k, R_k are themselves formulas of the same family at degree $d/2$; the propagation analysis above applies verbatim. The broader range of balanced $|S|$, i.e., $|S| \in [\varepsilon d, (1-\varepsilon)d]$ for constant $\varepsilon > 0$, admits the same propagation conclusion with only cosmetic quantitative changes. Our main theorems rule out polynomial-width S-decompositions of any *full-rank* set-multilinear polynomial for every such S, not only the midpoint case.

[4] For instance, the midpoint recursion (3) loses a factor that is *additive* in the exponent of n (roughly $\log(n/d_0) \approx \log n$ per the full descent), whereas other balanced self-reductions of IMM behave differently. We do not dwell on this quantitative point, as our barrier is structural and independent of it.

1.4 The two-ingredient recipe, and why each known example gives only one

To prove VF $\neq$ VNP via this route, one needs a set-multilinear polynomial family $\{g_{n,d}\}$ achieving simultaneously:

(I1) A technique delivering a strong set-multilinear formula lower bound $\mathcal{L}_{\mathrm{sm}}(g_{n,n}) \geq n^{\Omega(\log n)}$ in the large-degree regime.

(I2) A hardness-propagation mechanism from high to low degree – i.e., a polynomial-width midpoint decomposition (or more generally, an S-decomposition with $|S|$ bounded away from 0 and d and L_k, R_k drawn from the family itself).

Both ingredients have been realised *separately* in the literature. We stress at the outset that our barrier does not merely rule out the literal combination of (I1) and (I2) as stated, but the combination of their natural strengthenings – full-rankness in place of (I1), and an S-decomposition in place of (I2) – which is what one would actually need to instantiate the programme (see Section 1.5).

(I1): Full-rankness yields strong set-multilinear formula lower bounds

The *min-partition rank* method used in [7, 8], building on earlier partial-derivative and rank-based techniques of Nisan–Wigderson [12], Raz [13, 14], and Limaye, Srinivasan, and Tavenas [20, 9, 10], yields the following qualitative conclusion for every full-rank polynomial (i.e., a polynomial whose coefficient matrix has maximal rank under every balanced bipartition of the d blocks; see Definition 9).

▶ **Fact 3.** *For every full-rank set-multilinear polynomial family $\{P_{n,d}\}$ with $d = \Theta(n)$, every set-multilinear formula computing $P_{n,d}$ has size at least $n^{\Omega(\log n)}$.*

More precisely, [7] implies the lower bound $d^{(\log d)/120}$, which at $d = n$ is $n^{\Omega(\log n)}$. Explicit examples of full-rank set-multilinear polynomials abound (each over a suitable field extension): the Nisan–Wigderson design-based polynomial in VNP [12, 7, 8] (see Appendix B.1), the Dyck-word polynomial in VP of [8, Sec. 4.1] (see Appendix B.2), the VBP polynomial of [8] based on [2], and the recent VBP construction of [6]. By Fact 3, each witnesses an $n^{\Omega(\log n)}$ set-multilinear formula lower bound.

(I2): IMM admits a width-n midpoint decomposition

The identity (2) exhibits $\mathrm{IMM}_{n,d}$ as a width-n midpoint S-decomposition – the canonical instance of the structural property isolated in Definition 2.

Why neither ingredient alone suffices

For the known full-rank polynomials listed under (I1), no midpoint decomposition of polynomial width is known – and, as we will show, none exists. Consequently, the strong lower bound at degree $d = n$ does not propagate to the small-degree regime where Raz's set-multilinearisation applies. For $\mathrm{IMM}_{n,n}$, (I2) is available via (2), but (I1) is not: it is straightforward to see that IMM is *not* full-rank. Indeed, the identity (2) directly exhibits $\mathrm{IMM}_{n,d}$ as a sum of n rank-1 tensors under the midpoint bipartition, so the coefficient matrix at the contiguous balanced word has rank at most n. A matching lower bound is straightforward:

▶ **Lemma 4** (IMM is not full-rank with respect to the contiguous midpoint word). *Let $u_0 \in \{+k, -k\}^d$ be the contiguous balanced word assigning $+k$ to the first $d/2$ blocks and $-k$ to the last $d/2$ blocks. Then*

$$\operatorname{rank}(M_{u_0}(\mathrm{IMM}_{n,d})) \;=\; n, \qquad \textit{hence} \qquad \operatorname{relrk}_{u_0}(\mathrm{IMM}_{n,d}) \;=\; n^{1-d/2}.$$

Proof. *Upper bound.* By (2), the coefficient matrix $M_{u_0}(\mathrm{IMM}_{n,d})$ is a sum of n rank-1 matrices, so its rank is at most n. *Lower bound.* The n polynomials $\{(X_1 \cdots X_{d/2})_{1,k}\}_{k\in[n]}$ are linearly independent in $\mathbb{F}_{\mathrm{sm}}[X_1, \ldots, X_{d/2}]$ (they have pairwise-disjoint monomial supports, distinguished by the final column index). Likewise for the R_k's. Thus $M_{u_0}(\mathrm{IMM}_{n,d}) = \sum_{k=1}^{n} \mathbf{u}_k \mathbf{v}_k^\top$ where $\{\mathbf{u}_k\}_k$ and $\{\mathbf{v}_k\}_k$ are each linearly independent, giving rank exactly n. Normalising by $n^{d/2}$ gives $\operatorname{relrk}_{u_0}(\mathrm{IMM}_{n,d}) = n^{1-d/2}$. ◀

Lemma 4 evaluates IMM's rank at a single word u_0 and gives a vanishingly small value. In Appendix C, we in fact give an exact characterisation of $\operatorname{rank}(M_u(\mathrm{IMM}_{n,d}))$ for *every* balanced word u (Theorem 21): the rank equals $n^{\kappa(u)-1}$, where $\kappa(u)$ is the number of maximal constant runs in u. Thus the contiguous word u_0 (with $\kappa = 2$) yields the minimum rank n; a fully alternating word (with $\kappa = d$) yields the maximum rank n^{d-1}, i.e., *full-rankness*. IMM's rank profile is therefore highly non-uniform – full-rank at some words, nearly minimal at others – in sharp contrast to the uniformly full-rank polynomials of (I1). This non-uniformity is what intrinsically obstructs Fact 3 for IMM, and the current best unbounded-depth set-multilinear formula lower bound for $\mathrm{IMM}_{n,n}$ remains a much weaker $n^{\Omega(\log\log n)}$ [20].

1.5 The natural programme, and our result

In light of the preceding discussion, a natural and appealing programme for VF $\neq$ VNP suggests itself:

> *Construct an explicit set-multilinear polynomial family $\{g_{n,d}\}$ (drawing on* VNP *or beyond if necessary) that is* simultaneously *full-rank and admits a polynomial-width S-decomposition.*

Such a family would inherit the $n^{\Omega(\log n)}$ lower bound from Fact 3 at degree n, which could then be propagated via (4) to small degree, yielding VF $\neq$ VNP (even VF $\neq$ VBP if the family is in VBP) through Raz's set-multilinearisation. We prove that this programme is *unconditionally impossible* – not for any quantitative reason, but for a structural reason that leaves no room for improvement. At the heart of our argument is a single rank inequality relating the width of any S-decomposition of g to the relative rank of g at any balanced word (Lemma 13). Specialising this inequality to exact and approximate full-rankness yields our two main theorems, stated informally below.

▶ **Theorem 5** (Strong barrier, informal; see Theorem 14). *Every full-rank set-multilinear polynomial over d blocks of n variables has S-decomposition width at least $n^{\min(|S|,\, d-|S|)}$, for every $S \subseteq [d]$. In particular, at the midpoint $|S| = d/2$ the width is at least $n^{d/2}$ – exponential in d.*

▶ **Theorem 6** (Robust barrier, informal; see Theorem 15). *Even if a set-multilinear polynomial g is only* approximately *full-rank – with relative rank bounded below by a constant on a constant fraction of balanced words – its S-decomposition width at every balanced-enough S is still at least $n^{\Omega(\sqrt{d})}$, super-polynomial in n.*

The strong barrier, specialised to the midpoint, recovers "full-rankness forces midpoint-decomposition width $\geq n^{d/2}$." The robust barrier says that approximate full-rankness suffices, and rules out any rescue by moving the split off the midpoint or choosing a non-contiguous S. Via Proposition 12, polynomial-width midpoint decomposition drives IMM-style hardness propagation; consequently, the combined requirement – "full-rankness of $\{g_{n,d}\}$" together with "useful-for-escalation self-reducibility of $\{g_{n,d}\}$" – cannot be realised by any polynomial family.

1.6 Scope and related barriers

Our result is a barrier on a specific and well-motivated programme. We emphasise its scope.

What the barrier does *not* say

Our barrier does not rule out:

- Proving $n^{\Omega(\log n)}$ set-multilinear formula lower bounds for $\mathrm{IMM}_{n,n}$ directly. Since IMM is not full-rank (Lemma 4), any such bound must go beyond the full-rank paradigm. The current best is $n^{\Omega(\log \log n)}$ of [20].
- Self-reducibility at extreme splits. The case $|S| = 1$ is always available – one may expand any set-multilinear g along a single block X_i, writing $g = \sum_{k=1}^{n} x_{i,k} \cdot g_k$ with g_k set-multilinear on the remaining blocks, an S-decomposition of width n for $S = \{i\}$. This survives the barrier (as $\min(|S|, d - |S|) = 1$), but the corresponding recursion $s(d) \leq w(n) \cdot s(d-1)$ iterates too slowly – over $\Theta(d)$ levels – to yield any useful hardness propagation.

Relation to other barriers

To our knowledge, ours is the first barrier to the particular strategy of combining a full-rank lower bound with IMM-style self-reducibility. It is independent of two existing barriers in the literature:

- Efremenko, Garg, Oliveira, and Wigderson [3] show that rank methods – linear flattenings composed with matrix rank – cannot prove tensor rank or Waring rank lower bounds beyond $n^{\lfloor d/2 \rfloor}$. That barrier targets a lower-bound *technique* and applies to a different complexity measure; ours targets a *design constraint* on the hard polynomial for set-multilinear formula lower bounds.
- A recent result of the author [6] exhibits a full-rank multilinear polynomial computable by a polynomial-size set-multilinear ABP, showing that the min-partition rank method cannot prove super-polynomial smVBP lower bounds. That barrier says full-rankness alone is insufficient to separate smVBP from higher set-multilinear classes via the min-partition rank method; ours says full-rankness is incompatible with IMM-style self-reducibility in the set-multilinear *formula* setting.

Related work on IMM lower bounds and conditional separations

Two recent papers propose new *conditional* avenues to algebraic complexity separations, both complementary to the full-rank/self-reducibility programme our barrier rules out.

- Rossman [16] introduces a complexity measure on even-order tensors called *riffle rank*, conjectures that the d-fold identity tensor has riffle rank $n^{d-o(\log d)}$, and shows that such a bound would imply an $n^{(1-o(1))\log d}$ set-multilinear formula lower bound for $\mathrm{IMM}_{n,d}$ at some growing $d(n) = \omega(1)$, thereby separating VF from VBP (using self-reducibility to

propagate to the Raz transfer range). This is a direct conjectural route to a strong lower bound on IMM that eludes the full-rank paradigm: IMM is the target, and IMM is not full-rank.

- Rossman and Zhu [17] establish an analogous connection between *multiquadratic sum-of-squares* and commutative arithmetic formulas, in the spirit of the biquadratic/non-commutative connection of Hrubeš–Wigderson–Yehudayoff [5].[5] The SoS complexity of a d-multiquadratic polynomial f (quadratic in each of d blocks of n variables) is the minimum s such that $f = \sum_{i=1}^{s} g_i^2$ with each g_i being d-multilinear. Rossman and Zhu show that an $n^{d-o(\log d)}$ lower bound on the SoS complexity of an explicit d-multiquadratic polynomial, for $\omega(1) \le d(n) \le O(\log n / \log\log n)$, would separate VF from VNP.

Both of these programmes target essentially the same separation as ours, but route around the full-rank paradigm entirely: Rossman's via a new rank measure tailored to IMM; Rossman–Zhu's via a rank measure on $n^d \times n^d$ matrices (partial transpose rank, a generalisation of sum-of-squares complexity) that directly yields lower bounds at *small* degree via an analogue of Raz's tensor-rank theorem [15] – thereby bypassing the large-degree-to-small-degree propagation step altogether. Our Corollary 16 does not obstruct either programme – on the contrary, it underscores why techniques of this flavour may be needed: the natural full-rank-plus-self-reducibility strategy is unconditionally blocked.

Organisation

Section 2 sets up notation, recalls the relative-rank measure, and formalises how polynomial-width midpoint decomposition drives the lower bound propagation from large degree to small. Section 3 states the main theorems along with the underlying key inequality. Section 4 discusses the resulting picture. Appendix A contains the proofs of the main results; Appendix B works out how two concrete full-rank polynomials from [8] fail to admit polynomial-width S-decompositions; Appendix C gives a characterisation of the rank profile of $\mathrm{IMM}_{n,d}$ at arbitrary words.

2 Preliminaries

2.1 Set-multilinear polynomials

We work over an arbitrary field $\mathbb{F}$ (one may take $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$). Let $X_1, \dots, X_d$ be d disjoint sets of variables, with $|X_i| = n_i$. A polynomial f in these variables is *set-multilinear* with respect to $(X_1, \dots, X_d)$ if every monomial contains exactly one variable from each X_i. We write $\mathbb{F}_{\mathrm{sm}}[X_1, \dots, X_d]$ for the space of such polynomials, and recall the notation $\overline{X}(S) := (X_i)_{i \in S}$ from Definition 2; thus $\mathbb{F}_{\mathrm{sm}}[\overline{X}(S)]$ denotes the space of polynomials set-multilinear with respect to $\{X_i : i \in S\}$.

2.2 Words and relative rank

Following [10, 8], we parameterise bipartitions of blocks by *words*.

▶ **Definition 7** (Word). *A* word *of length d is a tuple $w = (w_1, \dots, w_d) \in (\mathbb{Z} \setminus \{0\})^d$. We write $\mathcal{P}_w := \{i \in [d] : w_i > 0\}$ and $\mathcal{N}_w := \{i \in [d] : w_i < 0\}$. For $S \subseteq [d]$, $w_S := \sum_{i \in S} w_i$ and $w|_S$ denotes the restriction of w to S. The* total imbalance *is $w_{[d]}$; w is* balanced *if $w_{[d]} = 0$.*

[5] Hrubeš–Wigderson–Yehudayoff showed that a suitable super-linear lower bound on the sum-of-squares complexity of an explicit biquadratic polynomial implies an exponential lower bound for *non-commutative* arithmetic circuits.

In our applications we specialise to words with $|w_i| = k$ for all i and variable sets of size $n := 2^k$. Balanced such words are exactly the elements of $\{+k, -k\}^d$ with $|\mathcal{P}_w| = |\mathcal{N}_w| = d/2$.

▶ **Definition 8** (Coefficient matrix and relative rank [10, 8]). *Let w be a word with $|w_i| = k$ for all i, and $n = 2^k$. Let $\mathcal{M}_w^{\mathcal{P}}$ (resp. $\mathcal{M}_w^{\mathcal{N}}$) be the set of all set-multilinear monomials over $(X_i)_{i\in\mathcal{P}_w}$ (resp. $(X_i)_{i\in\mathcal{N}_w}$). For a set-multilinear f, the* coefficient matrix $M_w(f) \in \mathbb{F}^{\mathcal{M}_w^{\mathcal{P}} \times \mathcal{M}_w^{\mathcal{N}}}$ *has rows indexed by $\mathcal{M}_w^{\mathcal{P}}$ and columns by $\mathcal{M}_w^{\mathcal{N}}$, with the (m_1, m_2) entry equal to the coefficient of the monomial $m_1 m_2$ in f (written $[m_1 m_2]\, f$; this is 0 unless $m_1 m_2$ is a set-multilinear monomial over all d blocks). The* relative rank *is*

$$\mathrm{relrk}_w(f) \;:=\; \frac{\mathrm{rank}(M_w(f))}{2^{\frac{1}{2}\sum_i |w_i|}} \;=\; \frac{\mathrm{rank}(M_w(f))}{n^{d/2}}.$$

For balanced w, $M_w(f)$ is a square $n^{d/2} \times n^{d/2}$ matrix and $\mathrm{relrk}_w(f) \in [0, 1]$. Otherwise $M_w(f)$ is rectangular and $\mathrm{relrk}_w(f) \le 2^{-|w_{[d]}|/2}$.

▶ **Definition 9** (Full-rankness). *A set-multilinear g is* full-rank *if $\mathrm{relrk}_w(g) = 1$ for every balanced word $w \in \{+k, -k\}^d$.*

▷ Claim 10 (Standard properties of relrk [10]). Let w be a word with $|w_i| = k$ for all i, and f, g set-multilinear.

(1) Imbalance. $\mathrm{relrk}_w(f) \le 2^{-|w_{[d]}|/2}$.

(2) Sub-additivity. $\mathrm{relrk}_w(f + g) \le \mathrm{relrk}_w(f) + \mathrm{relrk}_w(g)$.

(3) Multiplicativity. If $f = f_1 \cdots f_t$ with $f_i \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(S_i)]$ and $(S_1, \ldots, S_t)$ a partition of $[d]$, then $\mathrm{relrk}_w(f) = \prod_{i=1}^t \mathrm{relrk}_{w|_{S_i}}(f_i)$.

All three properties are proved in [10, Claim 2.3]. Property (1) follows from the dimension bound on $M_w(f)$; (2) from sub-additivity of matrix rank; (3) from the tensor factorisation $M_w(f) = \bigotimes_i M_{w|_{S_i}}(f_i)$, which holds since the f_i's have disjoint variable supports.

2.3 S-decomposition width equals bipartite rank

Recall from Definition 2 that the minimum width of an S-decomposition of g is the least w with $g = \sum_{k=1}^w L_k \cdot R_k$, L_k on the blocks of S and R_k on those of $\overline{S}$. We first observe that this minimum width equals the rank of a specific flattening of g.

▶ **Proposition 11** (Width equals flattening rank). *Let $S \subseteq [d]$ and let u_S^+ denote the word with $(u_S^+)_i = +k$ for $i \in S$ and $-k$ for $i \in \overline{S}$. The minimum width of an S-decomposition of g equals* $\mathrm{rank}(M_{u_S^+}(g))$.

Proof. Identify polynomials in $\mathbb{F}_{\mathrm{sm}}[\overline{X}(S)]$ with their coefficient vectors indexed by set-multilinear monomials on S, and similarly for $\overline{X}(\overline{S})$. Under this identification, the coefficient vector of a product $L \cdot R$ with $L \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(S)]$, $R \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(\overline{S})]$ is the outer product $\mathbf{u}\,\mathbf{v}^\top$, where $\mathbf{u}, \mathbf{v}$ are the coefficient vectors of L, R. Consequently, a width-w S-decomposition $g = \sum_{j=1}^w L_j \cdot R_j$ corresponds to a rank-w factorisation $M_{u_S^+}(g) = \sum_{j=1}^w \mathbf{u}_j \mathbf{v}_j^\top$, and vice versa. The minimum width of an S-decomposition thus equals the minimum w over such factorisations – which is precisely $\mathrm{rank}(M_{u_S^+}(g))$. ◀

Proposition 11 exhibits the minimum S-decomposition width as a rank measure. Our key inequality (Lemma 13) lower-bounds this rank by invoking the relative-rank machinery at *other* words (i.e., $u \neq u_S^+$).

2.4 Midpoint decomposition drives lower bound propagation

We now formalise how a polynomial-width midpoint decomposition enables IMM-style hardness propagation from large to small degree.

▶ **Proposition 12** (Midpoint decomposition drives propagation). *Let $\{g_{n,d}\}$ be a family of set-multilinear polynomials, indexed by n, d both powers of 2 with $d \le n$. Suppose that for every such n, d with $d \ge 2$, the polynomial $g_{n,d}$ admits a midpoint decomposition $g_{n,d} = \sum_{k=1}^{w(n)} L_k \cdot R_k$ in which each L_k, R_k is computable by a set-multilinear formula of size at most $\mathcal{L}_{\mathrm{sm}}(g_{n,d/2})$. Then*

$$\mathcal{L}_{\mathrm{sm}}(g_{n,d}) \;\le\; 2\,w(n) \cdot \mathcal{L}_{\mathrm{sm}}(g_{n,d/2}), \tag{5}$$

and consequently, for every power of 2 with $d \le n$,

$$\mathcal{L}_{\mathrm{sm}}(g_{n,d}) \;\ge\; \frac{\mathcal{L}_{\mathrm{sm}}(g_{n,n})}{(2w(n))^{\log(n/d)}}. \tag{6}$$

Proof. From the midpoint decomposition $g_{n,d} = \sum_{k=1}^{w(n)} L_k \cdot R_k$, we construct a set-multilinear formula for $g_{n,d}$ as follows. Each L_k has a set-multilinear formula of size at most $\mathcal{L}_{\mathrm{sm}}(g_{n,d/2})$ (by hypothesis); likewise for R_k. Connecting each L_k and R_k by a product gate and taking a $w(n)$-way sum gives a formula of total size at most $w(n) \cdot (\mathcal{L}_{\mathrm{sm}}(L_k) + \mathcal{L}_{\mathrm{sm}}(R_k)) \le 2\,w(n) \cdot \mathcal{L}_{\mathrm{sm}}(g_{n,d/2})$. This proves (5). Iterating $\log(n/d)$ times from $g_{n,n}$ down to $g_{n,d}$ yields (6). ◀

Proposition 12 isolates the quantitative content of IMM-style self-reducibility: a polynomial-width midpoint decomposition turns a large-degree set-multilinear formula lower bound into one at small degree, with a $(2w(n))^{\log(n/d)}$ multiplicative loss. When $w(n) = \mathrm{poly}(n)$, this loss is quasi-polynomial in n. The same mechanism applies, with only cosmetic changes, to any *balanced* S-decomposition: fix $\varepsilon \in (0, 1/2]$, and suppose each $g_{n,d}$ admits a width-$w(n)$ S-decomposition at some $|S| \in [\varepsilon d, (1-\varepsilon)d]$, with L_k, R_k each computable by a set-multilinear formula of size at most $\mathcal{L}_{\mathrm{sm}}(g_{n,\lceil(1-\varepsilon)d\rceil})$. One step now reduces degree from d to $\lceil(1-\varepsilon)d\rceil$ rather than $d/2$, and iterating to reach the Raz transfer point takes $\Theta_\varepsilon(\log(n/d))$ levels; for any constant $\varepsilon > 0$, the cumulative loss is still $(2w(n))^{O(\log(n/d))}$, and the same propagation conclusion holds. The main results of this paper rule out this recipe: no *full-rank* set-multilinear polynomial family can admit a polynomial-width S-decomposition at any balanced S.

3 The barrier

Both main results follow from the following rank inequality.

▶ **Lemma 13** (Key inequality). *Let $n = 2^k$ with $k \ge 1$, $d \ge 2$, and let g be set-multilinear over $X_1, \ldots, X_d$ with $|X_i| = n$. Let $S \subseteq [d]$, and suppose g admits an S-decomposition of width w. Then for every balanced word $u \in \{+k, -k\}^d$,*

$$w \;\ge\; \mathrm{relrk}_u(g) \cdot 2^{|u_S|}.$$

The lemma is a short consequence of the multiplicativity, sub-additivity, and imbalance properties of relative rank (Claim 10); the proof appears in Appendix A.

3.1 Strong and robust barriers

Lemma 13 is an "every balanced word" inequality. Its two natural instantiations correspond to exact and approximate full-rankness, and together constitute the main results of this paper.

▶ **Theorem 14** (Strong barrier). *If g is full-rank and $S \subseteq [d]$ is any subset, then every S-decomposition of g has width at least $n^{\min(|S|,d-|S|)}$.*

The proof specialises Lemma 13 to a word that places all of S on the $+k$ side; full-rankness forces the relative-rank factor to 1, and the imbalance factor becomes $n^{|S|}$. See Appendix A.

▶ **Theorem 15** (Robust barrier). *Let $\alpha, \rho \in (0,1]$ and $\varepsilon \in (0, 1/2]$. Suppose at least a ρ-fraction of balanced words $u \in \{+k, -k\}^d$ satisfy $\mathrm{relrk}_u(g) \geq \alpha$. Then for every $S \subseteq [d]$ with $|S| \in [\varepsilon d, (1-\varepsilon)d]$, every S-decomposition of g has width at least*

$$w \;\geq\; \alpha \cdot n^{c(\rho,\varepsilon)\sqrt{d}}$$

for some constant $c(\rho, \varepsilon) > 0$ depending only on ρ and ε.

The proof applies Lemma 13 to a random balanced word u: a second-moment argument (Paley–Zygmund combined with Hoeffding's hypergeometric-to-binomial comparison) shows that $|u_S| = \Omega(k\sqrt{d})$ with positive probability, even after intersecting with the event $\mathrm{relrk}_u(g) \geq \alpha$. See Appendix A.

3.2 The programmatic barrier

▶ **Corollary 16** (Programmatic barrier). *No family $\{g_{n,d}\}$ of set-multilinear polynomials (explicit or otherwise) simultaneously satisfies:*

- **(i)** *For all n, d and at least a ρ-fraction of balanced words $u \in \{+k, -k\}^d$, $\mathrm{relrk}_u(g_{n,d}) \geq \alpha$, for some fixed constants $\alpha, \rho > 0$;*
- **(ii)** *For all n and all even d, $g_{n,d}$ admits an S-decomposition of width $w(n) = n^{o(\sqrt{d})}$ for some $S \subseteq [d]$ with $|S| \in [\varepsilon d, (1-\varepsilon)d]$ for a fixed $\varepsilon > 0$.*

In particular, no such family admits a polynomial-width midpoint decomposition together with approximate full-rankness.

Via Proposition 12 and Fact 3, Corollary 16 means: there is no set-multilinear polynomial family on which the full-rank machinery of [14, 7, 8] can be combined with IMM-style lower bound propagation from large degree to small degree to yield $\mathsf{VF} \neq \mathsf{VNP}$.

4 Summary and Open Direction

The strategy of lifting a large-degree set-multilinear formula lower bound to a general formula lower bound via Raz's theorem has two ingredients: a way to prove the large-degree bound, and a way to propagate it to small degree where Raz's transfer applies. Both are understood in isolation, yet no explicit polynomial family has ever combined them. We show that no family *can*: the two ingredients are *unconditionally incompatible*.

Formalising IMM-style self-reducibility

A prerequisite for the barrier is a precise notion of "IMM-style self-reducibility" – which prior work (e.g., [20]) treats only informally via the IMM identity. We propose that one natural abstraction is the S-decomposition (Definition 2), and two propositions support this. First,

Proposition 11 identifies the minimum S-decomposition width of a polynomial with the rank of its flattening at a specific balanced word, establishing S-decomposition as a rank-theoretic object rather than a syntactic one. Second, Proposition 12 shows that this rank object drives hardness propagation: a polynomial-width midpoint decomposition, and more generally a polynomial-width balanced S-decomposition, propagates a large-degree lower bound to a small-degree one with only quasi-polynomial loss. The IMM identity (2) is the canonical instance of a width-n midpoint decomposition.

The barrier

Equipped with this framework, the main results of this paper (Theorems 14 and 15), both following from a single underlying rank inequality (Lemma 13), rule out the combination of the two ingredients. Any family admitting a polynomial-width balanced S-decomposition is either not full-rank at any balanced word, or not even approximately full-rank on a constant fraction of balanced words. The $\mathsf{VF} \neq \mathsf{VNP}$ strategy of combining a full-rank lower bound with IMM-style hardness propagation is therefore closed: any avenue along these lines must either *move beyond the full-rank paradigm* in the lower-bound step, or *strengthen the propagation mechanism* beyond polynomial-width S-decomposition. As an instance of the latter, IMM also admits a *compositional* self-reduction[6] – grouping the d matrices into $\sqrt{d}$ blocks of $\sqrt{d}$ exhibits $\mathrm{IMM}_{n,d}$ as a composition of two copies of $\mathrm{IMM}_{n,\sqrt{d}}$, giving $\mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,d}) \leq \mathcal{L}_{\mathrm{sm}}(\mathrm{IMM}_{n,\sqrt{d}})^2$ – which lies outside the single-split S-decomposition framework studied here; whether it too is incompatible with full-rankness (which would require controlling the relative rank of a composition of set-multilinear polynomials) is left to future work.

Open direction: direct lower bounds for IMM

Prove formula lower bounds for $\mathrm{IMM}_{n,d}$ that go beyond the full-rank paradigm. By Lemma 4, IMM is not full-rank, so any such result must exploit structural features of IMM other than its relative rank at isolated words. One such feature – exhibited by Theorem 21 in Appendix C – is the full *rank profile* of IMM: the rank of $M_u(\mathrm{IMM}_{n,d})$ grows geometrically in the number of sign-alternations of u, ranging from n at the contiguous word to full-rank n^{d-1} at alternating words. Current techniques evaluate rank at a single word (or average over a distribution of words) and so see only part of this picture; a lower-bound method for IMM may need to engage with the global shape of the rank profile across multiple words simultaneously. Concrete conjectural routes have recently been proposed: Rossman's *riffle rank* programme [16], and the *multiquadratic sum-of-squares* programme of Rossman–Zhu [17]. Corollary 16 does not rule out such techniques; it only closes off the particular strategy of transplanting the full-rank machinery to an IMM-like self-reducible polynomial.

References

1 Richard P. Brent. The parallel evaluation of general arithmetic expressions. *Journal of the ACM*, 21(2):201–206, 1974. doi:10.1145/321812.321815.

2 Zeev Dvir, Guillaume Malod, Sylvain Perifel, and Amir Yehudayoff. Separating multilinear branching programs and formulas. In *Proceedings of the 44th STOC*, pages 615–624. ACM, 2012. doi:10.1145/2213977.2214034.

[6] We thank the anonymous reviewer who pointed this out to us.

3 Klim Efremenko, Ankit Garg, Rafael Oliveira, and Avi Wigderson. Barriers for rank methods in arithmetic complexity. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *LIPIcs*, pages 1:1–1:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ITCS.2018.1.

4 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi:10.1080/01621459.1963.10500830.

5 Pavel Hrubeš, Avi Wigderson, and Amir Yehudayoff. Non-commutative circuits and the sum-of-squares problem. *Journal of the American Mathematical Society*, 24(3):871–898, 2011. doi:10.1090/S0894-0347-2011-00694-2.

6 Deepanshu Kush. An unconditional barrier for proving multilinear algebraic branching program lower bounds. *Electron. Colloquium Comput. Complex.*, TR26-043, 2026. URL: https://eccc.weizmann.ac.il/report/2026/043.

7 Deepanshu Kush and Shubhangi Saraf. Improved low-depth set-multilinear circuit lower bounds. In *37th Computational Complexity Conference (CCC 2022)*, volume 234 of *LIPIcs*, pages 38:1–38:16, 2022. doi:10.4230/LIPIcs.CCC.2022.38.

8 Deepanshu Kush and Shubhangi Saraf. Near-optimal set-multilinear formula lower bounds. In *38th Computational Complexity Conference (CCC 2023)*, volume 264 of *LIPIcs*, pages 15:1–15:33, 2023. doi:10.4230/LIPIcs.CCC.2023.15.

9 Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. On the partial derivative method applied to lopsided set-multilinear polynomials. In Shachar Lovett, editor, *37th Computational Complexity Conference, CCC 2022, Philadelphia, PA, USA, July 20-23, 2022*, volume 234 of *LIPIcs*, pages 32:1–32:23. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPIcs.CCC.2022.32.

10 Nutan Limaye, Srikanth Srinivasan, and Sébastien Tavenas. Superpolynomial lower bounds against low-depth algebraic circuits. *J. ACM*, 72(4):26:1–26:35, 2025. doi:10.1145/3734215.

11 Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994. doi:10.1016/S0022-0000(05)80043-1.

12 Noam Nisan and Avi Wigderson. Lower bounds on arithmetic circuits via partial derivatives. *Comput. Complex.*, 6(3):217–234, 1997. doi:10.1007/BF01294256.

13 Ran Raz. Separation of multilinear circuit and formula size. *Theory of Computing*, 2(6):121–135, 2006. doi:10.4086/TOC.2006.V002A006.

14 Ran Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. *J. ACM*, 56(2):8:1–8:17, 2009. doi:10.1145/1502793.1502797.

15 Ran Raz. Tensor-rank and lower bounds for arithmetic formulas. *J. ACM*, 60(6):40:1–40:15, 2013. doi:10.1145/2535928.

16 Benjamin Rossman. Riffle rank. In Luciano N. Grippo and Martín D. Safe, editors, *Proceedings of the XIII Latin-American Algorithms, Graphs and Optimization Symposium, LAGOS 2025, Buenos Aires, Argentina, November 10-14, 2025*, volume 273 of *Procedia Computer Science*, pages 215–222. Elsevier, 2025. doi:10.1016/j.procs.2025.10.301.

17 Benjamin Rossman and Davidson Zhu. Multi-quadratic sum-of-squares lower bounds imply VNC $^1 \neq$ VNP. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference, ITCS 2026, Bocconi University, Milan, Italy, January 27-30, 2026*, LIPIcs, pages 113:1–113:22. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2026. doi:10.4230/LIPIcs.ITCS.2026.113.

18 Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity, 2016. Available at https://github.com/dasarpmar/lowerbounds-survey.

19 Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science*, 5(3–4):207–388, 2010. doi:10.1561/0400000039.

20 Sébastien Tavenas, Nutan Limaye, and Srikanth Srinivasan. Set-multilinear and non-commutative formula lower bounds for iterated matrix multiplication. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 416–425. ACM, 2022. doi:10.1145/3519935.3520044.

21 Leslie G. Valiant. Completeness classes in algebra. In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC)*, pages 249–261. ACM, 1979. doi:10.1145/800135.804419.

A Proofs of the main results

We include the proofs of Lemma 13, Theorem 14, and Theorem 15. The latter two are implications of the former.

A.1 Proof of Lemma 13 (key inequality)

Proof of Lemma 13. Write $g = \sum_{j=1}^{w} L_j R_j$ with $L_j \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(S)]$, $R_j \in \mathbb{F}_{\mathrm{sm}}[\overline{X}(\overline{S})]$, where $\overline{S} := [d] \setminus S$. Let $u_L := u|_S$, $u_R := u|_{\overline{S}}$. Since u is balanced, $u_{[d]} = u_S + u_{\overline{S}} = 0$, so

$$|u_S| \;=\; |u_{\overline{S}}|. \tag{7}$$

For each j, since $(S, \overline{S})$ is a partition of $[d]$ and L_j, R_j have disjoint variable supports, multiplicativity (Claim 10(3)) gives

$$\mathrm{relrk}_u(L_j R_j) \;=\; \mathrm{relrk}_{u_L}(L_j) \cdot \mathrm{relrk}_{u_R}(R_j).$$

By the imbalance bound (Claim 10(1)) applied to L_j and R_j individually, combined with (7),

$$\mathrm{relrk}_{u_L}(L_j) \;\le\; 2^{-|u_S|/2}, \qquad \mathrm{relrk}_{u_R}(R_j) \;\le\; 2^{-|u_S|/2}.$$

Multiplying, $\mathrm{relrk}_u(L_j R_j) \le 2^{-|u_S|}$. By sub-additivity (Claim 10(2)),

$$\mathrm{relrk}_u(g) \;\le\; \sum_{j=1}^{w} \mathrm{relrk}_u(L_j R_j) \;\le\; w \cdot 2^{-|u_S|},$$

and rearranging yields the claimed inequality. ◀

A.2 Proof of Theorem 14 (strong barrier)

Proof of Theorem 14. Without loss of generality $|S| \le d - |S|$ (otherwise swap S and $\overline{S}$). Define a word $u \in \{+k, -k\}^d$ by setting $u_i = +k$ for every $i \in S$, $u_i = +k$ for an arbitrary choice of $d/2 - |S|$ positions in $\overline{S}$, and $u_i = -k$ for the remaining $d/2$ positions. Then u is balanced (exactly $d/2$ positions are $+k$), and $|u_S| = k|S|$ since every position in S receives $+k$. Hence $2^{|u_S|} = 2^{k|S|} = n^{|S|}$. Since g is full-rank, $\mathrm{relrk}_u(g) = 1$, and Lemma 13 gives $w \ge n^{|S|}$, as required. ◀

A.3 Proof of Theorem 15 (robust barrier)

Proof of Theorem 15. Fix S with $|S| = s \in [\varepsilon d, (1-\varepsilon)d]$. For a uniformly random balanced $u \in \{+k, -k\}^d$, let $Y := |\{i \in S : u_i = +k\}|$. Then Y follows the hypergeometric distribution with parameters $(d, d/2, s)$, i.e., the number of successes in s draws without replacement from a population of d containing $d/2$ successes. Standard calculation gives $\mathbb{E}[Y] = s/2$ and

$$\mathrm{Var}(Y) \;=\; \frac{s(d-s)(d/2)(d/2)}{d^2(d-1)} \;=\; \Theta\Big(\tfrac{s(d-s)}{d}\Big) \;=\; \Theta(d),$$

the last equality using $s \in [\varepsilon d, (1-\varepsilon)d]$. Since $u_S = k(2Y - s)$, we have $\mathrm{Var}(u_S) = 4k^2\mathrm{Var}(Y) = \Theta(k^2 d)$. We now need an *anti-concentration* bound: the event $|u_S| \geq \gamma k\sqrt{d}$ should hold with positive probability for some $\gamma > 0$. Since $|u_S| = 2k|Y - s/2|$, this amounts to a lower bound on the probability that Y deviates from its mean by $\Omega(\sqrt{d}) = \Omega(\sqrt{\mathrm{Var}(Y)})$. Let $Z := (Y - s/2)^2$, so $\mathbb{E}[Z] = \mathrm{Var}(Y) =: \sigma^2$. By the Cauchy–Schwarz inequality (the one-line derivation known as the Paley–Zygmund inequality), for any $\theta \in (0,1)$,

$$\Pr\left[Z \geq \theta\,\mathbb{E}[Z]\right] \;\geq\; (1-\theta)^2 \cdot \frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]} \;=\; (1-\theta)^2 \cdot \frac{\sigma^4}{\mathbb{E}\left[(Y - s/2)^4\right]}.$$

The ratio $\mathbb{E}[(Y - s/2)^4]/\sigma^4$ is the fourth standardised moment – or *kurtosis* – of the hypergeometric distribution. To bound it, we use Hoeffding's classical comparison [4, Theorem 4] between hypergeometric and binomial distributions: for every continuous convex function f,

$$\mathbb{E}[f(Y)] \;\leq\; \mathbb{E}[f(Y')],$$

where $Y' \sim \mathrm{Bin}(s, 1/2)$ is the analogous binomial (independent-trials) random variable with the same mean. Applying this with $f(t) = (t - s/2)^4$ (convex), and using the identity $\mathbb{E}[(Y' - s/2)^4] = (3 - 2/s)(s/4)^2 \leq 3s^2/16$ for $\mathrm{Bin}(s, 1/2)$ (a standard fourth-moment calculation for a sum of i.i.d. $\pm 1/2$ variables),

$$\mathbb{E}[(Y - s/2)^4] \;\leq\; \mathbb{E}[(Y' - s/2)^4] \;\leq\; \tfrac{3s^2}{16}.$$

On the other hand, since $d - s \geq \varepsilon d$,

$$\sigma^2 \;=\; \frac{s(d-s)(d/2)^2}{d^2(d-1)} \;\geq\; \frac{s \cdot \varepsilon d \cdot d^2/4}{d^2 \cdot d} \;=\; \frac{\varepsilon s}{4}.$$

Combining,

$$\frac{\mathbb{E}[(Y - s/2)^4]}{\sigma^4} \;\leq\; \frac{3s^2/16}{(\varepsilon s/4)^2} \;=\; \frac{3}{\varepsilon^2} \;=:\; C_\varepsilon.$$

Taking $\theta = 1/2$ in the Paley–Zygmund inequality yields

$$\Pr\left[|Y - s/2| \;\geq\; \tfrac{\sigma}{\sqrt{2}}\right] \;\geq\; \frac{1}{4C_\varepsilon}.$$

Since $\sigma^2 = \mathrm{Var}(Y) = \Theta_\varepsilon(d)$, there exist constants $\gamma_0(\varepsilon) > 0$ and $\delta_0(\varepsilon) > 0$ such that $\Pr[|Y - s/2| \geq \gamma_0\sqrt{d}] \geq \delta_0$, equivalently $\Pr[|u_S| \geq 2k\gamma_0\sqrt{d}] \geq \delta_0$. For $0 < \gamma \leq 2\gamma_0$, the event $\{|u_S| \geq \gamma k\sqrt{d}\}$ becomes less restrictive as γ decreases, and its probability $\delta(\gamma, \varepsilon) := \Pr[|u_S| \geq \gamma k\sqrt{d}]$ can be made arbitrarily close to 1 by taking $\gamma \to 0$. Now pick $\gamma = \gamma(\rho, \varepsilon) > 0$ small enough that $\delta(\gamma, \varepsilon) > 1 - \rho$. By a union bound, a uniformly random balanced u satisfies both $\mathrm{relrk}_u(g) \geq \alpha$ and $|u_S| \geq \gamma k\sqrt{d}$ with probability at least $\rho + \delta(\gamma, \varepsilon) - 1 > 0$. In particular, some such u exists; applying Lemma 13 to this u gives

$$w \;\geq\; \mathrm{relrk}_u(g) \cdot 2^{|u_S|} \;\geq\; \alpha \cdot 2^{\gamma k\sqrt{d}} \;=\; \alpha \cdot n^{\gamma\sqrt{d}},$$

using $n = 2^k$. Setting $c(\rho, \varepsilon) := \gamma(\rho, \varepsilon)$ completes the proof. ◀

B Failure of known full-rank polynomials to admit polynomial S-decompositions

We illustrate Corollary 16 on two concrete full-rank polynomials from [7, 8]: the Nisan–Wigderson polynomial (based on [11]) and the Dyck-word polynomial (based on [13]). By Fact 3, both witness $n^{\Omega(\log n)}$ set-multilinear formula lower bounds; the discussion below shows why neither delivers $\mathsf{VF} \neq \mathsf{VNP}$ through the programme of Section 1.5.

B.1 The Nisan–Wigderson polynomial

For a prime power n, identify $\mathbb{F}_n$ with $[n]$, and for variable sets $X_j = \{x_{i,j} : i \in [n]\}$, $j \in [d]$, define

$$\mathrm{NW}_{n,d}(\overline{X}) := \sum_{\substack{f \in \mathbb{F}_n[z] \\ \deg f < d/2}} \prod_{j=1}^{d} x_{f(j),\,j}.$$

▶ **Fact 17** ([8], Claim 3.2). $\mathrm{NW}_{n,d}$ *is full-rank.*

By Theorem 14, for any $S \subseteq [d]$, every S-decomposition of $\mathrm{NW}_{n,d}$ has width at least $n^{\min(|S|,d-|S|)}$. For $|S| = d/2$, this is $n^{d/2}$, and is in fact attained:

▶ **Proposition 18** (Matching upper bound for $\mathrm{NW}_{n,d}$). $\mathrm{NW}_{n,d}$ *admits a midpoint decomposition of width exactly* $n^{d/2}$.

Proof. Split each summand at the midpoint: $\prod_{j=1}^{d} x_{f(j),j} = L_f \cdot R_f$, with $L_f := \prod_{j=1}^{d/2} x_{f(j),j}$ and $R_f := \prod_{j=d/2+1}^{d} x_{f(j),j}$. The number of f's with $\deg f < d/2$ is $n^{d/2}$. ◀

The moral: the natural expression of $\mathrm{NW}_{n,d}$ distributes its rank mass uniformly over a combinatorial set (degree-$< d/2$ polynomials over $\mathbb{F}_n$) of cardinality $n^{d/2}$; by full-rankness, no algebraic rearrangement can collapse this.

B.2 The Dyck polynomial

The Dyck-word polynomial $F_{n,d}$ of [8, Sec. 4.1] is defined over variable sets $\overline{X}, \overline{Y}$ (each Y_i of size n), by $F_{n,d} := f_{1,d}$, where

$$f_{i,j} = \begin{cases} y_{i,j}\, y_{j,i}\, (X_i \cdot X_j) & \text{if } j = i+1, \\ 0 & \text{if } j - i \text{ even}, \\ y_{i,j}\, y_{j,i}\, (X_i \cdot X_j)\, f_{i+1,j-1} + \sum_{r=i+1}^{j-1} f_{i,r}\, f_{r+1,j} & \text{otherwise}, \end{cases}$$

with $X_i \cdot X_j = \sum_{\ell=1}^{n} x_{i,\ell}\, x_{j,\ell}$.

▶ **Fact 19** ([8], Lemma 4.5). *Over a field of characteristic zero, and viewed as a polynomial in* $\overline{X}$ *over the field* $\mathbb{F}(\overline{Y})$*, the polynomial* $F_{n,d}$ *is full-rank.*

By Theorem 14 applied over $\mathbb{F}(\overline{Y})$ (permissible since relative rank is preserved under scalar field extension), every midpoint decomposition of $F_{n,d}$ in $\overline{X}$ has width at least $n^{d/2}$. We add a word on why this is not visible from the natural recursion for $F_{n,d}$. Unfolding the definition expresses $F_{n,d}$ as a sum over balanced bracketings (Dyck words) u of length d, each weighted by a corresponding $\overline{Y}$-monomial. The *outermost* matching bracket of such a u splits $[d]$ into a prefix and a suffix at some odd position $2r-1$, contributing a product across that split; but as u ranges over all Dyck words, this outermost split occurs at *every* valid position, not just the midpoint. Thus the natural recursion does not present $F_{n,d}$ as a single midpoint decomposition of small width; it presents it as a sum, over all admissible split positions, of decompositions at those positions (informally, an entire "spectrum" of S-decompositions, one S per split position). Theorem 14 shows that no clever regrouping can convert this into a polynomial-width decomposition at any *single* balanced split: full-rankness forces width $n^{d/2}$ at every such split.

B.3 Summary

In both cases the mechanism is captured abstractly by Theorem 14: full-rankness at the contiguous balanced word u_0 forces the minimum midpoint-decomposition width to equal $\mathrm{rank}(M_{u_0}(g)) = n^{d/2}$, incompatible with the polynomial-width requirement. No choice of the L_k, R_k can evade this.

C The rank profile of IMM at general words

Lemma 4 records the rank of $\mathrm{IMM}_{n,d}$ at the contiguous word u_0. We state a characterisation at *every* word $u \in \{+,-\}^d$ in terms of the combinatorics of u, and sketch the proof; full details appear in the full version. We use the standard parameterisation: $\mathrm{IMM}_{n,d}$ is defined on blocks $Y_1, \ldots, Y_d$ with $|Y_1| = |Y_d| = n$, $|Y_i| = n^2$ for interior i, and

$$\mathrm{IMM}_{n,d}(Y_1, \ldots, Y_d) \;=\; \sum_{(i_1,\ldots,i_{d-1}) \in [n]^{d-1}} y_{1,i_1} \cdot \prod_{t=2}^{d-1} y_{t,i_{t-1},i_t} \cdot y_{d,i_{d-1}},$$

the $(1,1)$-entry of the matrix product. Monomials are in bijection with tuples $(i_1, \ldots, i_{d-1}) \in [n]^{d-1}$, the *path indices*.

▶ **Definition 20** (Run structure of a word). *For $u \in \{+,-\}^d$, write $u = B_1 \cdots B_{\kappa(u)}$ as a concatenation of maximal monochromatic runs. The* run count $\kappa(u) \geq 1$ *satisfies: $\kappa(u) - 1$ is the number of sign changes in u read left to right.*

As in Section 2, $M_u(\mathrm{IMM}_{n,d})$ has rows indexed by set-multilinear monomials over $(Y_i)_{i \in \mathcal{P}_u}$ and columns by those over $(Y_i)_{i \in \mathcal{N}_u}$.

▶ **Theorem 21** (Rank of IMM at general words). *For every word $u \in \{+,-\}^d$ with $\kappa(u) \geq 1$,*

$$\mathrm{rank}(M_u(\mathrm{IMM}_{n,d})) \;=\; n^{\kappa(u)-1}.$$

In particular, the contiguous word u_0 has $\kappa = 2$, recovering rank n (Lemma 4); the alternating word has $\kappa = d$, giving rank n^{d-1} (full rank); a constant-sign word has $\kappa = 1$, giving rank 1.

Proof sketch. Each of the $d-1$ *boundary positions* between consecutive blocks is *shared* if its two adjacent blocks have opposite signs and *internal* otherwise; there are exactly $\kappa(u) - 1$ shared positions. Every nonzero entry of $M_u(\mathrm{IMM}_{n,d})$ equals 1 and corresponds to a unique path index. A row monomial and a column monomial combine to a valid path-monomial iff they agree on all shared positions (these boundary indices appear on both sides), so grouping by the shared-assignment tuple $\vec{a} \in [n]^{\kappa-1}$ blocks the nonzero entries into the $\mathcal{R}(\vec{a}) \times \mathcal{C}(\vec{a})$. Within a block the row monomial is freely determined by the internal positions between two $\mathcal{P}$-blocks and the column monomial by those between two $\mathcal{N}$-blocks; these are disjoint and unconstrained, so the block is all-ones of rank 1. Distinct $\vec{a}$ give disjoint row and column sets. Hence $M_u(\mathrm{IMM}_{n,d})$ is a direct sum of $n^{\kappa-1}$ rank-1 blocks. ◀

▶ Remark 22 (Consequence for S-decompositions of IMM). By Proposition 11, the minimum S-decomposition width of $\mathrm{IMM}_{n,d}$ at $(S, \overline{S})$ equals $n^{\kappa(u_S^+)-1}$. The midpoint decomposition (2) ($\kappa = 2$) achieves the minimum value n; any non-contiguous bipartition strictly increases κ and hence the width, so the natural midpoint split is optimal.