Electronic Colloquium on Computational Complexity, Report No. 90 (2026)

Multilinear Formula Lower Bounds for Sparse Determinants

Pruthvi Boyapati
CSTAR, International Institute of Information Technology, Hyderabad, India
Suryajith Chillara
CSTAR, International Institute of Information Technology, Hyderabad, India
Pratyush Vempati
CSTAR, International Institute of Information Technology, Hyderabad, India

Abstract

Raz (2009) proved that multilinear formulas computing the determinant of a generic $n \times n$ matrix require size $n^{\Omega(\log n)}$. A fundamental question in understanding this lower bound is identifying which structural properties of the determinant drive this hardness. To answer this, we prove the existence of $n \times n$ symbolic matrices with only $\Theta(n \log^6 n)$ nonzero entries—reducing the variable count by a factor of $n/\log^6 n$—such that any multilinear formula computing their determinants still requires size $n^{\Omega(\log n)}$. Our construction uses rectangle sampling from the complete bipartite graph to generate sparse matrices that simultaneously maintain perfect matchings (ensuring nonzero determinant) while exhibiting *diagonal imbalance* under random vertex permutations—a geometric property we identify as the key driver of factor imbalance in Raz's framework.

This demonstrates that Raz's partial derivatives method is remarkably robust to sparsification, and suggests that the fundamental source of multilinear hardness for determinant lies in expansion-like combinatorial structure rather than density. Our techniques combine concentration inequalities for dependent random variables with insights from random graph theory.

2012 ACM Subject Classification Theory of computation → Algebraic complexity theory

Keywords and phrases Determinants, Multilinear polynomials, Formula lower bounds

Acknowledgements We would like to thank the anonymous referees of CCC 2026 for their careful readings and for their valuable comments and insights. SC thanks Srikanth Srinivasan for informative discussions about VBP-hardness of sparse determinants.

1 Introduction

The computational complexity of the determinant and permanent of a generic $n \times n$ matrix has been a central question in algebraic complexity theory. Valiant [18] established fundamental completeness results in his algebraic complexity framework: the permanent is VNP-complete under p-projections, while the determinant lies in the more restricted class VBP (polynomial-size algebraic branching programs) and it characterizes VP through quasi-polynomial projections. Furthermore, Valiant [19] showed that computing the permanent of $\{0, 1\}$ matrices is #P-complete, establishing a direct connection between algebraic and Boolean complexity. The separation between determinant and permanent complexity parallels the P versus NP question in Boolean complexity.

While the determinant can be computed in polynomial time via Gaussian elimination over fields, computing determinants over general commutative rings or in parallel complexity classes requires more sophisticated techniques. Berkowitz [1] developed an elegant algorithm that computes the characteristic polynomial (and hence the determinant) of an $n \times n$ matrix using $O(\log^2 n)$ parallel time with polynomially many processors, based on iterated matrix multiplication and the Cayley-Hamilton theorem. Mahajan and Vinay [9] provided a combinatorial characterization of the determinant through *closed walk sequences* (clow

ISSN 1433-8092

sequences), leading to an alternative iterated matrix multiplication approach that places the determinant computation in the complexity class GapL. Both algorithms demonstrate that despite the apparent sequential nature of Gaussian elimination, the determinant can be computed through carefully structured iterated matrix products, in parallel. In general, iterated matrix products are also referred to as *Algebraic Branching Programs* (ABPs) and they characterize the complexity class VBP.

A polynomial $f(x_1, \ldots, x_n)$ is *multilinear* if the degree of each variable x_i in every monomial is at most one. A formula is *syntactically multilinear* if every gate in the formula computes a multilinear polynomial. Both the determinant and permanent are multilinear polynomials, yet classical algorithms for computing the determinant (such as Mahajan-Vinay or Berkowitz's algorithm) crucially rely on intermediate computations that involve non-multilinear terms, which eventually cancel to produce the final multilinear result. This suggests that restricting to purely multilinear computation models might make these polynomials harder to compute. Indeed, Raz [12, 14] established the first superpolynomial lower bounds for multilinear formulas, proving that any multilinear formula computing the determinant or permanent of an $n \times n$ matrix requires size $n^{\Omega(\log n)}$. Raz [13] then proved that multilinear circuits are superpolynomially more powerful than multilinear formulas (see [15] for a simplified exposition). The lower bounds of Raz and subsequent work apply to syntactically multilinear formulas, where the multilinearity constraint is enforced at every gate.

While the determinant has a polynomial-sized ABP in the general setting, it is not known whether determinant can be computed by polynomial-sized multilinear ABPs. Raz and Yehudayoff [15] conjectured that multilinear ABPs might be exponentially more powerful than multilinear formulas. This conjecture was resolved by Dvir, Malod, Perifel, and Yehudayoff [6], who constructed an explicit polynomial family computable by polynomial-size multilinear ABPs but requiring superpolynomial multilinear formula size. Raz and Yehudayoff [15] proved that for every constant k, there exist polynomials computable by multilinear formulas of depth $k + 1$ that require superpolynomial size when restricted to depth k, establishing a hierarchy for constant depths. This hierarchy was made *near-optimal* by Chillara, Engels, Limaye, and Srinivasan [4].

Shpilka and Yehudayoff [17, Open Problem 14] raised the following fundamental question – are general formulas strictly more powerful than multilinear formulas? Over large fields, Chillara, Limaye and Srinivasan [5] showed that the general formulas outperform multilinear formulas of depth $o(\log n)$. Extending this to $O(\log n)$ depth would resolve Shpilka and Yehudayoff's question as [5] also showed that the depth reduction of formulas to $O(\log n)$ depth ([3, 2]) is optimal. It is important to note that [5] provides evidence that the multilinearity restriction fundamentally limits computational power, and that even natural algebraic problems like iterated matrix multiplication over small matrices become significantly harder when intermediate cancellations of non-multilinear terms are forbidden.

1.1 Motivation and Our Results

A central challenge in algebraic complexity is understanding *why* certain restricted computation models admit lower bounds while others remain intractable. Raz's groundbreaking proof [12, 14] that multilinear formulas require size $n^{\Omega(\log n)}$ to compute the determinant is our strongest lower bound for an explicit polynomial family in formula complexity. However, the proof's reliance on multiple properties of the determinant—dense connectivity, quadratic variable count, permutation-based symmetries—makes it unclear which aspects are essential and which are incidental.

Understanding the essential properties that drive multilinear hardness is valuable for two reasons. First, it guides the search for new lower bounds: if we can isolate the minimal structural requirements, we can identify other polynomials that might admit similar techniques. Second, it provides insight into the broader question of whether general formulas are fundamentally more powerful than multilinear formulas—an important open problem raised by Shpilka and Yehudayoff [17]. While recent work by Chillara, Limaye, and Srinivasan [5] shows that general formulas can outperform shallow multilinear formulas, the question remains open for logarithmic-depth formulas.

Our work takes the following approach: we systematically stress-test Raz's technique by removing structural properties to determine which are essential. Specifically, we investigate whether *density*—the fact that the determinant involves n^2 variables in a fully connected matrix—is necessary for the lower bound. If hardness persists for sparse determinants with only $O(n \text{ polylog } n)$ variables, this would demonstrate that density is not the core driver, and would suggest that Raz's method captures deeper combinatorial properties.

▶ **Theorem 1** (Lower Bound for Sparse Determinants). *There exists a family of $n \times n$ sparse symbolic matrices $(X_G)_{n\in\mathbb{N}}$ (constructed via probabilistic sampling) with the following properties:*

1. *Each row and column has $\Theta(\log^6 n)$ nonzero entries,*
2. *The underlying bipartite graph G has a perfect matching, ensuring $\mathsf{Det}(X_G) \neq 0$,*
3. *Any multilinear formula computing $\mathsf{Det}(X_G)$ requires size $s = n^{\Omega(\log n)}$.*

Our result demonstrates that Raz's technique is remarkably robust: reducing the variable count from n^2 to $\Theta(n \log^6 n)$—a factor of $n/\log^6 n$—still yields a superpolynomial lower bound. Despite the weakening, we get the best possible lower bound of $n^{\Omega(\log n)}$.

▶ Remark 2. All results stated for determinant also hold for the permanent.

1.2 Proof overview

Since we build on the proof strategy of Raz [12], we shall present the overview of their proof and then point out the locations where our proof deviates from theirs.

Raz's partial derivative measure:

Let f be a polynomial defined on the variable set X. Let $\Gamma : X \mapsto Y \sqcup Z$ be an equi-partition of X into two sets Y and Z. Let $\mathcal{M}_Y$ be the set of all multilinear monomials over the variables $\Gamma^{-1}(Y)$, and $\mathcal{M}_Z$ be the set of all multilinear monomials over the variables $\Gamma^{-1}(Z)$. Let M_f be a matrix whose rows and columns are indexed by elements of $\mathcal{M}_Y$ and $\mathcal{M}_Z$ respectively. An entry of this matrix M_f indexed by row m_Y and column m_Z is the coefficient of the monomial $m_Y \cdot m_Z$ in f. Naturally, rank of this matrix is well defined. Raz's complexity measure of the polynomial f is the rank of the matrix M_f. We shall henceforth denote it as $\mu(f)$. It is clear that $\mu(f) \leq 2^{|X|/2}$. Raz [12] made a fundamental observation that if f has a *multiplicative structure*, then under a random partition Γ, the rank could be much lower than the maximum possible rank with good probability.

Let $f = f_1 \times f_2$ where f_1 and f_2 are defined over the variable disjoint variable sets X_1 and X_2 where $X_1 \sqcup X_2 = X$. Let Y_1, Z_1 and Y_2, Z_2 be the partitions imposed by Γ on X_1 and X_2 respectively. It follows that $M_f = M_{f_1} \otimes M_{f_2}$ and thus $\mu(f) = \mu(f_1) \cdot \mu(f_2)$. If $|Y_1| \neq |Z_1|$ then it is easy to see that $\mu(f_1)$ is at most $2^{(|X_1|-1)/2}$ and same would be the case for $\mu(f_2)$. Thus, under this structure of f and the partition Γ that imposes inequality in the factors f_1 and f_2, $\mu(f)$ can at most be $2^{\frac{|X|}{2}-1}$. Instead of two factors, if f was expressed as a product of t many factors and if at least ℓ of these t factors had this "discrepancy" then we get that $\mu(f)$ is at most $2^{\frac{|X|-\ell}{2}}$. Instead of asking for $|Y_i| \neq |Z_i|$, we could also ask for a difference of at least k between $|Y_i|$ and $|Z_i|$ for a "discrepancy" parameter k. For this, we need non-trivial guarantees on the size of X_i's depending on k.

It was also shown[1] that every multilinear formula F of size s that computes a polynomial f on n variables admits a product-decomposition of the following form.

$$f = \sum_{i=1}^{s+1} f_i \text{ where } f_i = \prod_{j=1}^{t} f_{ij} .$$

where $t = \Theta(\log n)$ and for each $j \in [t]$,

$$\left(\frac{1}{3}\right)^j |X| \leq |X_{ij}| \leq \left(\frac{2}{3}\right)^j |X| .$$

That is, at least $\Omega_\delta(\log n)$ factors $X_{i,j}$ are such that $|X_{ij}| \geq |X|^\delta$. This gives us the leeway to work with a suitable discrepancy factor k. To prove a lower bound, we need to show that there is a equi-partition of the variables under which the polynomial of interest attains full rank and the multilinear formula of small size attains rank that is far from full. Raz [13] notes that technique can be applied to a *full-rank polynomial* directly but not to dthe eterminant.

Since the determinant is a polynomial on n^2 many variables, Raz's measure on all n^2 variables would yield a matrix of with $2^{n^2/2}$ many rows and columns. But this matrix can never attain full rank or a rank that is near-full as the number of monomials in the determinant polynomial are at most $2^{n \log(n)}$. Towards this, Raz [12] embeds m 2×2 matrices (with two variables per matrix kept alive and rest are set to 1) along a generalized diagonal (for some $m = o(n)$) at random and sets those variables in the rest of the $(n-2m) \times (n-2m)$ matrix that correspond to the same generalized diagonal to 1, and all the other variables to 0. The determinant of such an embedding would "eventually" be of the form $\prod_{i=1}^{m}(y_i - z_i)$. The rank attained by the matrix of the restricted determinant would be 2^m.

On the formula side, Raz shows that each factor under this random restriction attains an imbalance with a high probability. Further he shows that each summand has rank deficiency with a probability that is inverse quasi-polynomial. Putting all of these together with careful probabilistic arguments, Raz [12] concludes that a multilinear formula of size smaller than $n^{O(\log n)}$ cannot compute determinant (and permanent).

Unlike Raz [14], our polynomial of interest is not explicit but is a determinant of a probabilistically constructed sparse matrix X_G. We construct X_G through rectangle sampling: we independently sample each of the $N = \binom{n}{2}^2$ rectangles (that is, 2×2 sub-matrices) in the complete $n \times n$ generic matrix with probability $p = \frac{\log^6 n}{n^3}$ (equivalently sampling $K_{2,2}$'s from a complete bipartite graph $K_{n,n}$), and retain an entry (edge) if it appears in at least one sampled rectangle. This yields a sparse matrix X_G and the corresponding sparse graph G with $\Theta(n \log^6 n)$ edges where each vertex has degree $\Theta(\log^6 n)$ which is a union of sampled rectangles $\mathcal{R}$.

A crucial property we seek from G is that under a random permutation $\sigma = (\sigma_L, \sigma_R)$ of vertices, for any large set S of adversarially chosen edges (with $|S| \geq |E_G|^{0.75}$), a constant fraction of rectangles exhibit *diagonal imbalance* relative to S—meaning exactly one of the two opposite diagonal corners belongs to S. This geometric property, which we prove

[1] The current presentation of formulas through their *log-product decomposition* is due to [17]. Though this form of presentation does not appear in [12], it is just a rephrasing of Raz's ideas.

via McDiarmid's inequality with bounded differences, creates the asymmetry that drives imbalance and thus rank deficiency of a factor in our lower bound proof.

Another crucial property that we need is the non-zeroness of $\mathsf{Det}(X_G)$. We verify that via Hall's marriage condition. We present two approaches for this. One approach is through rigorous analysis across various size regimes (small, moderate, and near-full sets) to ensure $\mathsf{Det}(X_G) \neq 0$ with high probability. The other approach is through graph coupling arguments from the study of random graphs [7, Chapter 1]. Through this we establish that a sparse matrix exists with the required structure with a very high probability.

We then apply a carefully designed random restriction to both the determinant $\mathsf{Det}(X_G)$ and any purported multilinear formula computing it— we sample $m = \Theta(n^{1/3})$ disjoint[2] rectangles uniformly from the collection $\mathcal{R}$, and for each sampled rectangle, we randomly choose one of two diagonal configurations: top-left and bottom-right corners become either active variables (y_i, z_i) or (z_i, y_i), while the other diagonal is set to all 1s. This creates a partition of variables into sets Y and Z of size m each, with the remaining $|E_G| - 2m$ variables set according to a perfect matching in the residual graph. The key technical result (Theorem 23) shows that for any factor in the formula decomposition with at least $N^{0.75}$ variables, the number of sampled rectangles contributing diagonal imbalance to that factor is quite substantial with high probability. Since each rectangle's configuration is chosen independently, the (Y, Z) partition for that factor follows an approximate binomial distribution with sufficient concentration, making k-balance (where $||Y| - |Z|| < 2k$ for a suitable parameter k) occur with small probability.

The lower bound emerges from analyzing the formula decomposition structure. Any formula of size s decomposes into $s + 1$ summands, each being a product of $\Theta(\log n)$ factors, where crucially $\Theta(\log n)$ factors in each product have at least $N^{0.75}$ variables. Conditioned on the sampled rectangles, the balance events across these large factors form a *read-2 family* with respect to the independent configuration choices, since each rectangle's two active variables can affect at most two factors. Applying the read-2 Chernoff bound (Theorem 16) with the small average balance probability, we show that with probability at least $1 - \exp(-\Omega(\log^2 n))$, at least half of the large factors in any single summand exhibit k-imbalance. By Proposition 11, this multiplicative rank deficiency gives each summand rank at most $2^m \cdot 2^{-\Omega(n^{1/32} \log n)}$. Taking a union bound over all $s+1$ summands, if $s = n^{O(\log n)}$, then with probability $1 - o(1)$, the total rank is bounded by $(s + 1) \cdot 2^m \cdot 2^{-\Omega(n^{1/32} \log n)} \ll 2^m$. However, the restricted determinant maintains full rank 2^m after the restriction (since the residual graph retains a perfect matching by Theorem 29), yielding a contradiction. Therefore, any multilinear formula computing $\mathsf{Det}(X_G)$ requires size $s = n^{\Omega(\log n)}$.

Our analysis crucially relies on the read-k tail bounds of [8].

1.3 Discussion

In this section, at the risk of being redundant, we discuss key design choices in our construction.

A natural question is: why not simply use the Erdős-Rényi model $G_{n,n,p}$ for a suitably chosen edge probability $p = \Theta(\frac{\log^6 n}{n})$? After all, we showed in Section 3 that such graphs have perfect matchings with high probability. The critical obstacle is the second-round restriction. In Raz's framework, we must sample vertex-disjoint rectangles and assign their corners to variables while ensuring the *residual graph* (after removing sampled vertices) retains a perfect matching. Under the $G_{n,n,p}$ model, the event that a sufficient number

[2] Sampled rectangles should not share rows or columns, analogous to [14]

of vertex-disjoint rectangles exist as induced subgraphs *and* the residual graph maintains a perfect matching has extremely low probability. Rectangle sampling resolves this. By directly sampling rectangles as atomic units with probability $\frac{\log^6 n}{n^3}$, we create a graph G that is the union of sampled rectangles. This design has two crucial advantages. First, rectangles are guaranteed to exist: the sampled set $\mathcal{R}$ of $\Theta(n \log^6 n)$ rectangles is available for second-round sampling by construction—we simply sample $m = \Theta(n^{1/3})$ rectangles uniformly from $\mathcal{R}$. Second, we have distribution invariance under deletion (Theorem 29): when we remove vertices corresponding to sampled rectangles, the residual graph has a similar distribution as rectangle sampling on the smaller $(n-2m) \times (n-2m)$ complete bipartite graph, implying the residual graph has a perfect matching with the same high probability as the original construction. Our approach is directly guided by Raz's method [14], which embeds m disjoint 2×2 block-diagonal matrices along a generalized diagonal of the dense $n \times n$ matrix. We achieve the same structural property—vertex-disjoint rectangles available for restriction—while working in a sparse regime, and the rectangle sampling framework is precisely what enables this.

Another natural approach would be to use structured sparse matrices such as band matrices, block-diagonal matrices, or matrices with support on expander graphs. These constructions have the advantage of being fully explicit and deterministic. However, structured constructions face a fundamental challenge: they provide an adversary with exploitable structure. Multilinear formulas decompose polynomials as $f = \sum_{i=1}^{s+1} \prod_{j=1}^{t} f_{ij}$, where the variable sets X_{ij} are determined by the formula's structure. An adversarial formula could potentially align factors with structured sparsity patterns, ensuring that each factor intersects few edges and thereby avoiding diagonal imbalance, or exploit symmetries in deterministic constructions to achieve balance across many factors simultaneously.

Randomness provides robustness against adversarial decompositions. Our random construction via rectangle sampling ensures that for any large variable set $S \subseteq E_G$ with $|S| \geq N^{0.75}$ (regardless of how the formula chooses S), under a random permutation σ, a constant fraction of rectangles exhibit diagonal imbalance relative to S (Theorem 21). The random permutation $\sigma = (\sigma_L, \sigma_R)$ of vertices is essential for creating diagonal imbalance in the sparse regime. Without it, corner positions (top-left, bottom-right, etc.) would be fixed, allowing an adversarial variable set S to align with the predetermined rectangle structure. Since these sparse matrices have at least 1 but at most $O(\log^6 n)$ rectangles per edge (versus $(n-1)^2$ in the dense case), such alignment could eliminate diagonal imbalance entirely. Random permutation ensures every variable has equal probability of occupying each corner position, symmetrizing the structure so that for any large set S, approximately half the rectangles containing S-edges exhibit diagonal imbalance (Theorem 21), regardless of how the formula chooses S.

The choice of sampling probability $p = \frac{\log^6 n}{n^3}$ yields edge density $\Theta(\frac{\log^6 n}{n})$ and degree $\Theta(\log^6 n)$ per vertex. The exponent 6 is not fundamental—it can be improved. Here any $\omega(\log n)$ parameter would suffice (and still not necessary) for the construction, with appropriate adjustments to the concentration bounds. The exponent 6 provides comfortable margins in our probabilistic arguments and most importantly make way for a cleaner exposition.

We also note that sparse determinants inherit VBP-hardness from general determinants through a natural reduction. Given any $n \times n$ determinant, the Mahajan-Vinay characterization [9] expresses it as the entry of an iterated matrix product, yielding an ABP of polynomial size s and width w. To create a sparse determinant encoding this ABP, we stagger the computation: replace each high-degree node in the ABP with a binary tree of intermediate nodes, ensuring the resulting graph has maximum degree $O(1)$. This transformation produces a sparse matrix where each row and column has constantly many nonzero entries, with size increasing from s to $(s \cdot w)^{O(1)}$ (polynomial) and width from w to $O(w \log s)$ (logarithmic increase). Since this reduction is a polynomial-size projection, sparse determinants are VBP-hard under polynomial projections. Consequently, if sparse determinants admitted general formulas of size $n^{o(\log n)}$, we would have VF = VBP, contradicting the widely believed hierarchy VF $\subsetneq$ VBP $\subsetneq$ VP.

2 Preliminaries

Throughout this paper, we work with an $n \times n$ symbolic matrix X where entries $x_{i,j}$ are formal variables. We identify this matrix with a complete bipartite graph $K_{n,n}$ where the left partition represents rows and the right partition represents columns.

▶ **Definition 3** (Rectangle). *A rectangle is a 2×2 submatrix of X. Given rows i, i' and columns j, j' with $i < i'$ and $j < j'$, the rectangle R consists of four entries: $(i, j), (i, j'), (i', j), (i', j')$. The complete bipartite graph $K_{n,n}$ contains exactly $N = \binom{n}{2}^2$ distinct rectangles.*

▶ **Definition 4** (Corner Positions). *For a rectangle defined by rows $\{i, i'\}$ (with $i < i'$) and columns $\{j, j'\}$ (with $j < j'$), we define four corner positions. The top-left position is (i, j). The top-right position is (i, j'). The bottom-left position is (i', j). The bottom-right position is (i', j'). These positions depend on an ordering $\sigma = (\sigma_L, \sigma_R)$ of the left and right vertices respectively.*

2.1 Edmonds' Criterion and Determinant Non-vanishing

▶ **Definition 5** (Edmonds' Graph). *The Edmonds' graph of an $n\times n$ symbolic matrix $A = \{x_{i,j}\}$ is the bipartite graph $G_A = (V_1 \sqcup V_2, E)$, where $V_1 = \{u_1, \ldots, u_n\}$, $V_2 = \{v_1, \ldots, v_n\}$, such that $(u_i, v_j) \in E$ if and only if $x_{i,j} \neq 0$.*

▶ **Lemma 6** (Edmonds' Criterion). *Let A be an $n \times n$ symbolic matrix, and let G_A be its Edmonds' graph. Then, $\mathsf{Det}(A) \neq 0$ if and only if G_A has a perfect matching.*

2.2 Algebraic complexity classes

Arithmetic circuits are a standard model of computation in the algebraic complexity theory. An arithmetic circuit is a directed acyclic graph where internal nodes are labeled with arithmetic operations ($+$ or $\times$) and leaves are labeled with variables or field constants, computing a polynomial in the natural way. The size of a circuit is the number of edges, and the fundamental question in algebraic complexity theory is to understand which polynomials require large circuits to compute.

Let $(f_n)_{n\in\mathbb{N}}$ be a polynomial family where f_n has n variables and degree $\mathrm{poly}(n)$. We say $(f_n) \in \mathsf{VF}$ if each f_n is computable by a polynomial-size arithmetic formula (tree-shaped circuit with fan-out one at each gate); $(f_n) \in \mathsf{VP}$ if each f_n is computable by a polynomial-size arithmetic circuit of polynomial degree; and $(f_n) \in \mathsf{VNP}$ if $f_n(x) = \sum_{e\in\{0,1\}^m} g_n(x, e)$ for some $(g_n) \in \mathsf{VP}$ with $m = \mathrm{poly}(n)$ variables. The known containments $\mathsf{VF} \subseteq \mathsf{VP} \subseteq \mathsf{VNP}$ parallel the Boolean hierarchy, with both believed to be strict.

Algebraic branching programs (ABPs), also known as algebraic branching programs, compute polynomials through iterated matrix multiplication along paths from source to sink. It is is known that VBP lies between formulas and general circuits: $\mathsf{VF} \subseteq \mathsf{VBP} \subseteq \mathsf{VP}$, where both containments are believed to be strict.

2.3 Raz's Partial Derivatives Measure

▶ **Definition 7** (Variable Partition and Coefficient Matrix). *Let f be a multilinear polynomial over variables X. A partition $\sigma : X \mapsto Y \sqcup Z$ divides X into two disjoint sets with $|Y| = |Z| = |X|/2$ (assuming $|X|$ is even). Let $\mathcal{M}_Y$ denote the set of all multilinear monomials over $\sigma^{-1}(Y)$, and $\mathcal{M}_Z$ the set of all multilinear monomials over $\sigma^{-1}(Z)$. The coefficient matrix M_f^σ has rows indexed by $\mathcal{M}_Y$ and columns indexed by $\mathcal{M}_Z$, where the entry (m_Y, m_Z) equals the coefficient of the monomial $m_Y \cdot m_Z$ in the expansion of f. We define the complexity measure $\mu_\sigma(f) = rank(M_f^\sigma)$.*

▶ **Observation 8.** *For any partition σ of a multilinear polynomial f over $|X|$ variables, we have $\mu_\sigma(f) \leq 2^{|X|/2}$, with equality when f is full-rank under σ.*

▶ **Proposition 9** (Multiplicativity Under Products). *If $f = f_1 \times f_2$ where f_1 and f_2 are multilinear polynomials defined over disjoint variable sets X_1 and X_2 with $X = X_1 \sqcup X_2$, then $M_f^\sigma = M_{f_1}^{\sigma|_{X_1}} \otimes M_{f_2}^{\sigma|_{X_2}}$ and consequently $\mu_\sigma(f) = \mu_{\sigma|_{X_1}}(f_1) \cdot \mu_{\sigma|_{X_2}}(f_2)$.*

2.4 Imbalance and Rank Deficiency

▶ **Definition 10** (k-Imbalance and Balance). *For a multilinear polynomial f over variable set X and partition $\sigma : X \mapsto Y \sqcup Z$, we say that f exhibits k-imbalance under σ if $||Y| - |Z|| \geq 2k$, and that f is k-balanced under σ if $||Y| - |Z|| < 2k$.*

▶ **Proposition 11** (Rank Loss from Imbalance). *If f is a multilinear polynomial over $|X|$ variables exhibiting k-imbalance under partition σ, then $\mu_\sigma(f) \leq 2^{(|X|-2k)/2} = 2^{|X|/2} \cdot 2^{-k}$. If $f = \prod_{i=1}^{t} f_i$ where each f_i is defined over disjoint variable set X_i, and at least ℓ factors exhibit k-imbalance, then $\mu_\sigma(f) \leq 2^{(|X|-2\ell k)/2} = 2^{|X|/2} \cdot 2^{-\ell k}$.*

Proof. For a single factor with k-imbalance, we have $||Y| - |Z|| \geq 2k$. Without loss of generality, assume $|Y| \leq |Z|$, so $|Y| = (|X| - \Delta)/2$ and $|Z| = (|X| + \Delta)/2$ where $\Delta \geq 2k$. The coefficient matrix has dimension $2^{|Y|} \times 2^{|Z|}$, so rank is at most $\min(2^{|Y|}, 2^{|Z|}) = 2^{|Y|} \leq 2^{(|X|-2k)/2}$. For products, by Proposition 9, each imbalanced factor contributes a multiplicative factor of at most 2^{-k} to the rank. ◀

By Proposition 9, the product structure over disjoint variable sets ensures that the rank measure μ_σ behaves multiplicatively: $\mu_\sigma(f_1 \times f_2) = \mu_\sigma(f_1) \cdot \mu_\sigma(f_2)$. Under a random partition Γ, this multiplicativity is decisive: each factor exhibiting k-imbalance contributes a multiplicative factor of 2^{-k} to the rank (Proposition 11), and with ℓ such factors, the rank suffers an exponential deficiency of $2^{-\ell k}$. Our approach therefore requires that multilinear formulas admit decompositions exposing product structure over disjoint variable sets—a structural property already established in the literature [17, 16].

▶ **Theorem 12** (Formula Decomposition [17, 16]). *Every multilinear formula F of size s computing a multilinear polynomial f over N variables admits a decomposition $f = \sum_{i=1}^{s+1} f_i$ where $f_i = \prod_{j=1}^{t} f_{ij}$ with $t = \Theta(\log N)$ and for each level $j \in [t]$, the variable set X_{ij} satisfies $\left(\frac{1}{3}\right)^j N \leq |X_{ij}| \leq \left(\frac{2}{3}\right)^j N$.*

2.5 Concentration Inequalities

We would need the following concentration inequalities in our work.

▶ **Theorem 13** (Chernoff Bound [11]). *Let* $X_1, \ldots, X_n$ *be independent random variables with* $X_i \in \{0,1\}$ *and* $\Pr[X_i = 1] = p_i$*. Let* $X = \sum_{i=1}^n X_i$ *and* $\mu = \mathbb{E}[X] = \sum_{i=1}^n p_i$*. Then for any* $\delta > 0$*:*

$$\Pr[X \geq (1+\delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{2+\delta}\right) \leq \exp\left(-\frac{\delta^2\mu}{3}\right) \quad \text{for } 0 < \delta \leq 1,$$
$$\Pr[X \leq (1-\delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{2}\right).$$

▶ **Theorem 14** (McDiarmid's Inequality [10]). *Let* $X_1, \ldots, X_n$ *be independent random variables and let* $f : \mathcal{X}^n \to \mathbb{R}$ *satisfy the bounded differences property: there exist constants* $c_1, \ldots, c_n$ *such that for all* $i \in [n]$ *and all* $x_1, \ldots, x_n, x_i' \in \mathcal{X}$*,*

$$|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \leq c_i.$$

Then for any $t > 0$*,*

$$\Pr\left[|f(X_1, \ldots, X_n) - \mathbb{E}[f(X_1, \ldots, X_n)]| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

▶ **Definition 15** (Read-k Family of Events). *Let* $Y_1, \ldots, Y_r$ *be random events determined by functions of independent random variables* $X_1, \ldots, X_m$*. We say that the collection forms a read-*k *family if each* X_j *influences at most* k *of the events* Y_i*.*

▶ **Theorem 16** ([8]). *Let* $Y_1, \ldots, Y_r$ *be a family of read-*k *indicator variables with* $\Pr[Y_i = 1] = p_i$*, and let* p *be the average of* $p_1, \ldots, p_r$*. Then for any* $\varepsilon > 0$*,*

$$\mathbb{P}\left[Y_1 + \ldots + Y_r \geq (p+\varepsilon)r\right] \leq e^{-D(p+\varepsilon \,\|\, p)\cdot r/k}$$

and

$$\mathbb{P}\left[Y_1 + \ldots + Y_r \leq (p-\varepsilon)r\right] \leq e^{-D(p-\varepsilon \,\|\, p)\cdot r/k}.$$

Here, $D(q \,\|\, p)$ is the *Kullback-Leibler divergence* defined as

$$D(q \,\|\, p) = q\log\left(\frac{q}{p}\right) + (1-q)\log\left(\frac{1-q}{1-p}\right),$$

where $D(q \,\|\, p) \geq 2(q-p)^2$.

3 Sparse Matrix Construction via Rectangle Sampling

Let X be a generic $n \times n$ symbolic matrix corresponding to the complete bipartite graph $K_{n,n}$. Through controlled deletion of edges, we generate a sparse bipartite graph G on n vertices in each partition. We define our polynomial of interest as $\mathsf{Det}(X_G)$, the determinant of the matrix obtained by setting $x_{i,j} = 0$ for all edges $(i,j) \notin E_G$ and leaving $x_{i,j}$ as a formal variable for all $(i,j) \in E_G$. This polynomial possesses the following key properties: it is defined over $|E_G| = \Theta(n\log^6 n)$ variables, it is nonzero due to the existence of perfect matchings in G, and the combinatorial structure of G inherited from rectangle sampling creates the necessary asymmetry and thus the hardness for our lower bound technique.

3.1 The Sampling Process

▶ **Definition 17** (Rectangle Sampling). *We sample a set of rectangles $\mathcal{R}$ by including each of the $N = \binom{n}{2}^2$ rectangles independently with probability $p = \frac{\log^6 n}{n^3}$. An edge (i, j) survives in the graph G if and only if it appears in at least one sampled rectangle.*

The expected number of rectangles is given by linearity of expectation as $\mathbb{E}[|\mathcal{R}|] = N \cdot p = \binom{n}{2}^2 \cdot \frac{\log^6 n}{n^3} \approx n \log^6 n$. Since rectangles are sampled independently, standard Chernoff bounds yield the following result.

▶ **Proposition 18** (Concentration of $|\mathcal{R}|$). *With probability at least $1 - \exp(-\Omega(n \log^6 n))$, we have $|\mathcal{R}| = \Theta(n \log^6 n)$.*

3.2 Edge Survival Probabilities

Each edge (i, j) appears in exactly $(n-1)^2$ rectangles (by choosing one additional row and one additional column in $(n-1)$ ways each). The probability that edge (i, j) survives is given by the following calculation. We have

$$\mathbb{P}\left[(i, j) \text{ survives}\right] = 1 - (1 - p)^{(n-1)^2} \approx 1 - e^{-(n-1)^2 p} \approx (n-1)^2 p \approx \frac{\log^6 n}{n}.$$

The expected number of edges follows by linearity of expectation as $\mathbb{E}[|E_G|] \approx n^2 \cdot \frac{\log^6 n}{n} = n \log^6 n$.

▶ **Proposition 19** (Concentration of $|E_G|$ via Read-4 Chernoff). *With probability at least $1 - \exp(-\Omega(n \log^6 n))$, we have $|E_G| = \Theta(n \log^6 n)$. Similarly, each vertex degree is $\Theta(\log^6 n)$ with probability at least $1 - \exp(-\Omega(\log^6 n))$.*

Proof. Let $Y_1, \ldots, Y_{n^2}$ be indicator random variables for edges and $X_1, \ldots, X_N$ be indicator random variables for rectangles. Each edge appears in exactly four corner positions across all rectangles it belongs to, which forms a read-4 structure. The edge indicator Y_e depends on the rectangle indicators for rectangles containing e, and each rectangle indicator X_R influences exactly four edge indicators corresponding to the four corners of R. Applying the read-4 Chernoff bound with the expected number of edges $\mathbb{E}[|E_G|] = \Theta(n \log^6 n)$ yields the stated concentration. The vertex degree bound follows similarly by considering the read-2 structure restricted to edges incident to a fixed vertex. ◀

Let N_e be the number of rectangles an edge $e = (i, j) \in [n] \times [n]$ belongs to after the sampling process. Observe then that $\mathbb{E}\left[N_e\right] = (n-1)^2 \cdot p \approx \frac{\log^6 n}{n}$ and hence $N_e = \Theta\left(\frac{\log^6 n}{n}\right)$ with a high probability. However, note that $N_e \geq 1$ when edge e survives.

We now present two methods to establish that the sparse matrix we have constructed has a non-zero determinant. The first relies on rigorously establishing that Hall's Marriage Theorem holds across the resulting Edmond's graph, while the second relies on a threshold function argument for the perfect matching on the individual edge probabilities.

3.3 Non-zero Determinant via Hall's Marriage Theorem

▶ **Definition 20** (Diagonal Rectangle Sets). *For a set $S \subseteq E_G$, and a random permutation $\sigma = (\sigma_L, \sigma_R)$ of the left and right vertices, let $\mathcal{R}_{S,diag}(\sigma)$ denote the set of rectangles in $\mathcal{R}$ where exactly one of the two opposite corners (TL and BR, or TR and BL) belongs to S under the ordering σ.*

▶ **Theorem 21** (Concentration of Diagonal Rectangles). *Let $S \subseteq E_G$ with $|S| \geq |E_G|^{0.75}$. Then with probability at least $1 - 2\exp(-\Omega(n^{0.5}\log^{-3} n))$ over the random permutation $\sigma = (\sigma_L, \sigma_R)$, we have $|\mathcal{R}_{S,diag}(\sigma)| \geq (1-\delta) \cdot \frac{|S|}{2}$ for any constant $\delta > 0$.*

Proof. For each rectangle $R \in \mathcal{R}$, define $W_R(\sigma) = \mathbf{1}[\text{exactly one of TL, BR of } R \text{ is in } S]$. Then $|\mathcal{R}_{S,\text{diag}}(\sigma)| = \sum_{R\in\mathcal{R}} W_R(\sigma)$.

For computing the expectation, consider a fixed rectangle R defined by rows $\{i, i'\}$ and columns $\{j, j'\}$, with four edges $e_1 = (i, j)$, $e_2 = (i, j')$, $e_3 = (i', j)$, $e_4 = (i', j')$. The permutation σ randomly determines which pair becomes TL and BR versus TR and BL, with each pairing having probability one half. Therefore, $\mathbb{E}[W_R(\sigma)] = \frac{1}{2}[\mathbf{1}[e_1 \in S \oplus e_4 \in S] + \mathbf{1}[e_2 \in S \oplus e_3 \in S]]$, where $\oplus$ denotes XOR.

Summing over all rectangles and rearranging by edges gives

$$\begin{aligned}\mathbb{E}[|\mathcal{R}_{S,\text{diag}}(\sigma)|] &= \frac{1}{2}\sum_{R\in\mathcal{R}}[\mathbf{1}[e_1 \in S \oplus e_4 \in S] + \mathbf{1}[e_2 \in S \oplus e_3 \in S]] \\ &= \frac{1}{2}\sum_{e\in E_G}\sum_{\substack{R\in\mathcal{R}\\ e\in R}} \mathbf{1}[e \in S \oplus \text{opposite}(e, R) \in S].\end{aligned}$$

For any subset S of edges, let $N_e^{(S)}$ count the number of rectangles in which opposite corner of e is in S. For $e \in S$, this counts rectangles where the opposite corner is not in S, contributing $N_e - N_e^{(S)}$. For $e \notin S$, this counts rectangles where the opposite corner is in S, contributing $N_e^{(S)}$. Since $\sum_{e\notin S} N_e^{(S)} = \sum_{e\in S} N_e^{(S)}$, we have

$$\mathbb{E}[|\mathcal{R}_{S,\text{diag}}(\sigma)|] = \frac{1}{2}\left[\sum_{e\in S}(N_e - N_e^{(S)}) + \sum_{e\in S} N_e^{(S)}\right] = \frac{1}{2}\sum_{e\in S} N_e \geq \frac{|S|}{2},$$

using the fact that $N_e \geq 1$ for all surviving edges $e \in S$.

For concentration via McDiarmid's inequality, changing a single coordinate of σ affects at most $c = O(\log^6 n)$ rectangles (bounded by the maximum vertex degree). By McDiarmid's inequality applied to $2n$ coordinates, we have

$$\Pr[||\mathcal{R}_{S,\text{diag}}(\sigma)| - \mathbb{E}[|\mathcal{R}_{S,\text{diag}}(\sigma)|]| > t] \leq 2\exp\left(-\frac{2t^2}{2nc^2}\right).$$

Setting $t = \delta \cdot \mathbb{E}[|\mathcal{R}_{S,\text{diag}}(\sigma)|] = \Theta(|S|) = \Theta(n^{0.75}\log^{4.5} n)$ yields

$$\Pr[\text{deviation}] \leq 2\exp\left(-\Omega\left(\frac{n^{1.5}\log^9 n}{n\log^{12} n}\right)\right) = 2\exp(-\Omega(n^{0.5}\log^{-3} n)).$$

◀

3.4 Sampling Rectangles for Round 2

The preceding theorem establishes concentration for $|\mathcal{R}_{S,\text{diag}}(\sigma)|$ under a fixed permutation σ. We now analyze what happens when we sample m rectangles uniformly at random from $\mathcal{R}$ in the second round. This is the key connection between our construction and the factor imbalance needed for the lower bound.

▶ **Definition 22** (Sampled Diagonal Rectangles). *Let $\mathbf{R} = (R_1, \dots, R_m)$ denote m rectangles sampled uniformly and independently from $\mathcal{R}$ with replacement. For a set $S \subseteq E_G$ and permutation σ, define $Y_S(\sigma, \mathbf{R}) = |\{i \in [m] : R_i \in \mathcal{R}_{S,diag}(\sigma)\}|$ as the number of sampled rectangles that belong to $\mathcal{R}_{S,diag}(\sigma)$.*

▶ **Theorem 23** (Concentration of Sampled Diagonal Rectangles)**.** *Let $S \subseteq E_G$ with $|S| \geq |E_G|^{0.75}$. Sample $m = \Theta(n^{1/3})$ rectangles uniformly and independently from $\mathcal{R}$ with replacement, and let σ be a uniformly random permutation. Then with probability at least $1 - \exp(-\Omega(n^{1/12} \log^{-1.5} n))$ over both σ and $\mathbf{R}$, we have $Y_S(\sigma, \mathbf{R}) \geq (1 - \delta') \cdot \frac{m|S|}{2|\mathcal{R}|}$ for any constant $\delta' > 0$.*

Proof. We condition on σ and analyze the distribution of Y_S given σ. For the conditional distribution, given a fixed permutation σ, the rectangles $R_1, \ldots, R_m$ are sampled independently and uniformly from $\mathcal{R}$. Therefore, $Y_S \mid \sigma = \sum_{i=1}^{m} \mathbf{1}[R_i \in \mathcal{R}_{S,\text{diag}}(\sigma)]$ is a sum of m independent Bernoulli random variables, each with success probability $\frac{|\mathcal{R}_{S,\text{diag}}(\sigma)|}{|\mathcal{R}|}$. Thus, $\mathbb{E}[Y_S \mid \sigma] = m \cdot \frac{|\mathcal{R}_{S,\text{diag}}(\sigma)|}{|\mathcal{R}|}$. By the Chernoff bound, for any $\delta_1 > 0$, we have

$$\Pr\left[Y_S < (1 - \delta_1) \cdot \mathbb{E}[Y_S \mid \sigma] \;\middle|\; \sigma\right] \leq \exp\left(-\frac{\delta_1^2 \mathbb{E}[Y_S \mid \sigma]}{2}\right).$$

For the expectation over σ, using the law of total expectation gives $\mathbb{E}_{\sigma,\mathbf{R}}[Y_S] = \mathbb{E}_\sigma[\mathbb{E}_{\mathbf{R}}[Y_S|\sigma]] = m \cdot \frac{\mathbb{E}_\sigma[|\mathcal{R}_{S,\text{diag}}(\sigma)|]}{|\mathcal{R}|} \geq \frac{m|S|}{2|\mathcal{R}|}$, where the last inequality follows from Theorem 21.

For the two-stage concentration argument, from Theorem 21, for any constant $\delta_2 > 0$, with probability at least $1 - 2\exp(-\Omega(n^{0.5} \log^{-3} n))$ over σ, we have $|\mathcal{R}_{S,\text{diag}}(\sigma)| \geq (1 - \delta_2) \cdot \frac{|S|}{2}$.

Let $\mathcal{E}_1$ denote the event that $\mathbb{E}[Y_S|\sigma] \geq (1 - \delta_2) \cdot \mathbb{E}[Y_S]$, and let $\mathcal{E}_2$ denote the event that $Y_S \geq (1 - \delta_1) \cdot \mathbb{E}[Y_S|\sigma]$. We want to show that $Y_S \geq (1 - \delta') \cdot \mathbb{E}[Y_S]$ with high probability, which follows from $\mathcal{E}_1 \cap \mathcal{E}_2$ by choosing $\delta' = \delta_1 + \delta_2 - \delta_1\delta_2$.

We decompose the failure probability as $\Pr[(\mathcal{E}_1 \cap \mathcal{E}_2)^c] \leq \Pr[\mathcal{E}_1^c] + \Pr[\mathcal{E}_2^c]$. From Theorem 21, we have $\Pr[\mathcal{E}_1^c] \leq 2\exp(-\Omega(n^{0.5} \log^{-3} n))$. To bound $\Pr[\mathcal{E}_2^c]$, we further decompose as $\Pr[\mathcal{E}_2^c] \leq \Pr[\mathcal{E}_2^c|\mathcal{E}_1] + \Pr[\mathcal{E}_1^c]$.

For bounding $\Pr[\mathcal{E}_2^c|\mathcal{E}_1]$, conditioning on $\mathcal{E}_1$ means we condition on $\sigma \in \mathcal{E}_1$. For any such σ, we have $\mathbb{E}[Y_S|\sigma] \geq (1 - \delta_2) \cdot \frac{m|S|}{2|\mathcal{R}|}$. With $|S| \geq |E_G|^{0.75} = \Theta(n^{0.75} \log^{4.5} n)$, $|\mathcal{R}| = \Theta(n \log^6 n)$, and choosing $m = \Theta(n^{1/3})$, we have

$$\mathbb{E}[Y_S|\sigma] \geq \Theta\left(\frac{n^{1/3} \cdot n^{0.75} \log^{4.5} n}{n \log^6 n}\right) = \Theta\left(\frac{n^{13/12} \log^{4.5} n}{n \log^6 n}\right) = \Theta(n^{1/12} \log^{-1.5} n).$$

For large n where this expectation is $\Omega(n^{1/12})$, by the Chernoff bound we have $\Pr[\mathcal{E}_2^c|\sigma] \leq \exp(-\Omega(n^{1/12} \log^{-1.5} n))$. Since this bound holds uniformly for all $\sigma \in \mathcal{E}_1$, we obtain $\Pr[\mathcal{E}_2^c|\mathcal{E}_1] \leq \exp(-\Omega(n^{1/12} \log^{-1.5} n))$.

For the final bound, we have

$$\begin{aligned}
\Pr[(\mathcal{E}_1 \cap \mathcal{E}_2)^c] &\leq \Pr[\mathcal{E}_1^c] + Pr[\mathcal{E}_1^c] + \Pr[\mathcal{E}_2^c|\mathcal{E}_1] \\
&\leq 2\exp(-\Omega(n^{0.5} \log^{-3} n)) + \exp(-\Omega(n^{1/12} \log^{-1.5} n)) \\
&= \exp(-\Omega(n^{1/12} \log^{-1.5} n)),
\end{aligned}$$

where the dominant term is the last exponential. ◀

3.5 Verification of Hall's Marriage Condition

To rigorously establish that our randomly sampled sparse graphs have perfect matchings with high probability, we verify Hall's marriage condition. For every subset $S \subseteq V_1$, the neighborhood $N(S) = \{v \in V_2 : \exists u \in S, (u, v) \in E_G\}$ must satisfy $|N(S)| \geq |S|$. We partition the analysis by set size into three regimes: small sets ($|S| \leq \frac{n}{\log n}$), moderate sets ($\frac{n}{\log n} < |S| \leq n - \frac{n}{\log n}$), and near-full sets ($n - \frac{n}{\log n} < |S| \leq n$).

3.5.1 Quantifying Neighborhood Coverage

For a fixed set $S \subseteq V_1$ of size k and a set $T \subseteq V_2$ of size ℓ, let $M_{T,S}$ denote the number of rectangles incident on at least one row in S and at least one column in T. Let $C_{i,T}$ be the set of rectangles incident on row i with at least one column in T. Then we have $|C_{i,T}| = (n-1)\left[\binom{n}{2} - \binom{n-\ell}{2}\right]$ and $|C_{i,T} \cap C_{j,T}| = \binom{n}{2} - \binom{n-\ell}{2}$. By inclusion-exclusion, we obtain

$$\begin{aligned} M_{T,S} = \left|\bigcup_{i\in S} C_{i,T}\right| &\geq \sum_{i\in S}|C_{i,T}| - \sum_{\substack{i,j\in S\\ i\neq j}}|C_{i,T}\cap C_{j,T}| \\ &= k(n-1)\left[\binom{n}{2} - \binom{n-\ell}{2}\right] - \binom{k}{2}\left[\binom{n}{2} - \binom{n-\ell}{2}\right] \\ &= \underbrace{k\left[\binom{n}{2} - \binom{n-\ell}{2}\right]\left(n - \frac{k+1}{2}\right)}_{\alpha(n,k,\ell)}. \end{aligned}$$

The probability that $T \cap N(S) = \emptyset$ (no column from T neighbors S) is at most $\Pr[T \cap N(S) = \emptyset] \leq (1-p)^{M_{T,S}} \leq \exp(-p \cdot \alpha(n,k,\ell))$. Hall's condition fails for set S if and only if there exists T with $|T| \geq n-k+1$ such that $T \cap N(S) = \emptyset$. Therefore, $\Pr[|N(S)| < k] \leq \sum_{\ell \geq n-k+1}\binom{n}{\ell}\exp(-p\cdot\alpha(n,k,\ell))$.

3.5.2 Small Sets: $|S| \leq \frac{n}{\log n}$

▶ **Theorem 24** (Hall's Condition for Small Sets)**.** *With probability at least* $1 - \exp(-\Omega(\log^6 n))$*, Hall's condition holds for all sets* S *with* $1 \leq |S| \leq \frac{n}{\log n}$.

Proof. Simplifying the binomial terms gives $\binom{n}{2} - \binom{n-\ell}{2} = \frac{n(n-1)-(n-\ell)(n-\ell-1)}{2} = \frac{\ell(2n-\ell-1)}{2}$. For $\ell \geq n-k+1$, the expression $\ell(2n-\ell-1)$ is minimized at $\ell = n-k+1$, giving $\alpha(n,k,\ell) \geq k \cdot \frac{(n-k+1)(n+k-2)}{2} \cdot \left(n - \frac{k+1}{2}\right)$.

For $k \leq \frac{n}{\log n}$, we have $\alpha(n,k,\ell) \geq k \cdot \left(n - \frac{n}{\log n}\right) \cdot n \cdot \left(n - \frac{n}{2\log n}\right) \geq \frac{kn^3}{8}$. Therefore, the failure probability satisfies:

$$\begin{aligned} \Pr\left[\exists S : |S| = k \leq \frac{n}{\log n}, |N(S)| < k\right] &\leq \sum_{k=1}^{n/\log n}\binom{n}{k}\sum_{\ell\geq n-k+1}\binom{n}{\ell}\exp(-p\cdot\alpha(n,k,\ell)) \\ &\leq \sum_{k=1}^{n/\log n}\binom{n}{k}\cdot k\cdot\binom{n}{n-k}\cdot\exp\left(-\Omega\left(\frac{k\log^6 n\cdot n^3}{n^3}\right)\right) \\ &\leq \sum_{k=1}^{n/\log n} k^3 n^{2k}\cdot\exp(-\Omega(k\log^6 n)) \\ &\leq \sum_{k=1}^{n/\log n}\exp(-\Omega(k\log^6 n)) \leq \exp(-\Omega(\log^6 n)). \end{aligned}$$

◀

3.5.3 Moderate Sets: $\frac{n}{\log n} < |S| \leq n - \frac{n}{\log n}$

For the moderate size regime, we analyze sets with $\frac{n}{\log n} < k \leq n - \frac{n}{\log n}$.

▷ Claim 25. For $\frac{n}{\log n} < k \leq n - \frac{n}{\log n}$ and $\ell \geq n-k+1 > \frac{n}{\log n}$, we have $\alpha(n,k,\ell) \geq \Omega\left(\frac{n^4}{\log^2 n}\right)$.

Proof. From the general lower bound, $\alpha(n,k,\ell) \geq k \cdot \frac{(n-k+1)(n+k-2)}{2} \cdot \left(n - \frac{k+1}{2}\right)$. Minimizing each factor in the relevant regime yields $\alpha(n,k,\ell) \geq \frac{n}{\log n} \cdot \frac{n}{\log n} \cdot n \cdot \frac{n}{2} = \Omega\left(\frac{n^4}{\log^2 n}\right)$. ◀

▶ **Theorem 26** (Hall's Condition for Moderate Sets). *With probability at least* $1 - \exp(-\Omega(n \log^4 n))$, *Hall's condition holds for all sets S with* $\frac{n}{\log n} < |S| \leq n - \frac{n}{\log n}$.

Proof. By a union bound and Claim 25, we have

$$\begin{aligned}
&\Pr\left[\exists S : \frac{n}{\log n} < |S| \leq n - \frac{n}{\log n}, |N(S)| < |S|\right] \\
&\leq \sum_{k=n/\log n}^{n-n/\log n} \sum_{\ell \geq n-k+1} \binom{n}{k}\binom{n}{\ell} \exp(-p \cdot \alpha(n,k,\ell)) \\
&\leq 4^n \cdot \mathrm{poly}(n) \cdot \exp\left(-\frac{\log^6 n}{n^3} \cdot \Omega\left(\frac{n^4}{\log^2 n}\right)\right) = 4^n \cdot \mathrm{poly}(n) \cdot \exp(-\Omega(n \log^4 n)) \\
&= \exp(-\Omega(n \log^4 n)).
\end{aligned}$$

◀

3.5.4 Near-Full Sets: $n - \frac{n}{\log n} < |S| \leq n$

▶ **Theorem 27** (Hall's Condition for Near-Full Sets). *With probability at least* $1 - \exp(-\Omega(\log^6 n))$, *Hall's condition holds for all sets S with* $n - \frac{n}{\log n} < |S| \leq n$.

Proof. For sets $S \subseteq V_1$ with $|S| = k > n - \frac{n}{\log n}$, consider the complementary set $\overline{S} = V_1 \setminus S$ with $|\overline{S}| = n - k < \frac{n}{\log n}$. By the symmetry of the bipartite graph construction, Hall's condition for S is equivalent to verifying that the complementary vertices have sufficiently large neighborhoods.

Specifically, if $|N(\overline{S})| \geq |\overline{S}|$, then the vertices in $V_2 \setminus N(\overline{S})$ can only be matched to vertices in S. Since $|V_2 \setminus N(\overline{S})| \geq n - |N(\overline{S})| \geq n - n + |\overline{S}| = |\overline{S}|$ implies that when $|N(\overline{S})| < n - |\overline{S}| + 1 = k$, we would have Hall's condition violated for S. However, if Hall's condition holds for $\overline{S}$, meaning $|N(\overline{S})| \geq |\overline{S}|$, then Hall's condition is satisfied for S as well.

The key observation is that the probability analysis for small sets in Theorem 24 applies directly to $\overline{S}$, since $|\overline{S}| < \frac{n}{\log n}$. By that theorem, with probability at least $1 - \exp(-\Omega(\log^6 n))$, all sets of size at most $\frac{n}{\log n}$ satisfy Hall's condition. This includes all complementary sets $\overline{S}$ for $|S| > n - \frac{n}{\log n}$.

Therefore, by symmetry and the analysis of small sets, with probability at least $1 - \exp(-\Omega(\log^6 n))$, Hall's condition holds for all near-full sets S with $n - \frac{n}{\log n} < |S| \leq n$. ◀

Combining Theorem 24, Theorem 26, and Theorem 27 via union bound gives the following:

▶ **Corollary 28** (Perfect Matching Existence). *With probability at least* $1 - \exp(-\Omega(\log^6 n))$, *the randomly sampled sparse graph G has a perfect matching, and hence* $\mathsf{Det}(X_G) \neq 0$.

▶ **Theorem 29** (Perfect Matching After Random Vertex Deletion). *Let* $m = \Theta(n^{1/3})$, *and let S_L and S_R denote the random sets of deleted rows and columns. With probability at least* $1 - \exp(-\Omega(\log^6 n))$ *over the construction of G and the choice of deletion of $2m$ rows and $2m$ columns* $\mathcal{D} = (S_L, S_R)$, *the graph* $G_{rest} = G|_{V_1 \setminus S_L, V_2 \setminus S_R}$ *has a perfect matching, and hence the restricted determinant remains non-zero.*

Proof. We apply the law of total expectation, conditioning on the specific deleted vertices. Let $\mathcal{D} = (S_L, S_R)$ denote a deterministic choice of which rows and columns are deleted. We write $\Pr[G_{\text{rest}}$ has perfect matching$] = \mathbb{E}_{\mathcal{D}}[\Pr[G_{\text{rest}}$ has perfect matching $\mid \mathcal{D}]]$.

For the inner probability conditioned on a fixed deletion pattern $\mathcal{D} = (S_L, S_R)$, by the analysis above, Hall's marriage condition provides the criterion for perfect matching existence. We established in Theorem 24, Theorem 26, and Theorem 27 that Hall's condition holds for all subsets of V_1 with high probability. Specifically, we showed that with probability at least $1 - \exp(-\Omega(\log^6 n))$ over the sparse matrix construction, every subset $S \subseteq V_1$ satisfies $|N(S)| \geq |S|$ in the graph G.

Given a fixed deletion pattern $\mathcal{D}$ with $|S_L| = |S_R| = 2m$ where $m = \Theta(n^{1/3}) \ll n$, consider the restricted graph $G|_{V_1 \setminus S_L, V_2 \setminus S_R}$. For Hall's condition to hold in this restricted graph, we need to verify that for every subset $T \subseteq V_1 \setminus S_L$, the neighborhood $N_{G_{\text{rest}}}(T)$ in the restricted graph satisfies $|N_{G_{\text{rest}}}(T)| \geq |T|$.

The key observation is that the neighborhood in the restricted graph can only be smaller than in the original graph. That is, $N_{G_{\text{rest}}}(T) = N_G(T) \cap (V_2 \setminus S_R)$, so $|N_{G_{\text{rest}}}(T)| = |N_G(T) \setminus S_R|$. For Hall's condition to fail in the restricted graph, we would need $|N_G(T) \setminus S_R| < |T|$, which means $|N_G(T) \cap S_R| > |N_G(T)| - |T|$.

However, our analysis in Section 3.5 established Hall's condition with substantial margin. Specifically, for small sets with $|T| \leq \frac{n}{\log n}$, Theorem 24 showed that the probability of neighborhood deficiency is exponentially small in $\log^6 n$. For moderate sets, Theorem 26 provided even stronger exponential bounds in $n \log^4 n$. For near-full sets, Theorem 27 established the result via symmetry with small sets.

Since the deletion affects at most $2m = \Theta(n^{1/3})$ vertices from each partition, and our Hall's condition analysis holds with high probability uniformly over the structure of G, the restricted graph maintains sufficient neighborhood expansion to satisfy Hall's condition. The margin provided by our concentration bounds in Section 3.5 accommodates the removal of $o(n)$ vertices without violating the matching condition. To see this explicitly, we can invoke Corollary 28 with sets each of size $n - 2m$ and in the analysis with-in, we avoid the rows and columns specified by $\mathcal{D}$.

For the outer expectation over deletion patterns $\mathcal{D}$, we note that the second-round sampling of rectangles determines $\mathcal{D}$ through the union of vertices in sampled rectangles. By the earlier invoking of Corollary 28, regardless of which specific $2m$ rows and columns are involved in the sampled rectangles, the conditional probability that the remaining graph has a perfect matching remains bounded below by $1 - \exp(-\Omega(\log^6 n))$. This follows because the rectangle sampling described earlier creates a graph with robust expansion properties that persist under any deletion pattern removing $o(n)$ vertices.

Therefore, by the law of total expectation,

$$\begin{aligned}\Pr[G_{\text{rest}} \text{ has perfect matching}] &= \mathbb{E}_{\mathcal{D}}[\Pr[G_{\text{rest}} \text{ has perfect matching} \mid \mathcal{D}]] \\ &\geq 1 - \exp(-\Omega(\log^6 n)),\end{aligned}$$

where the bound holds uniformly over all possible deletion patterns $\mathcal{D}$. ◀

3.6 Alternate Analysis: Threshold Functions and Perfect Matchings

We consider the Edmonds' graph of the matrix as a random bipartite graph, and show using the Erdős-Rényi threshold function on perfect matchings that the individual marginal edge probabilities, both before and after vertex deletion, clear the required threshold of $\frac{\ln n}{n}$, thus guaranteeing a perfect matching with high probability.

▶ **Definition 30** (Random Bipartite Graph Model). *Define the random bipartite graph model* $G_{n_1,n_2,p}$ *as the probability space of bipartite graphs with* $V = V_1 \sqcup V_2$, $|V_1| = n_1, |V_2| = n_2$, *where each edge between partitions is chosen independently with probability* p.

▶ **Definition 31** (Threshold Function). *A graph property* P *is said to have a threshold function* $f(n)$ *if:*
1. *For all* $p \ll f(n)$, $\Pr[G \in G_{n,n,p}$ *has* $P] \to 0$ *as* $n \to \infty$,
2. *For all* $p \gg f(n)$, $\Pr[G \in G_{n,n,p}$ *has* $P] \to 1$ *as* $n \to \infty$.

The property of having a perfect matching is monotone (adding edges preserves the property), and by classical results in random graph theory [7], every non-trivial monotone property has a threshold function.

▶ **Theorem 32** (Erdős-Rényi Threshold for Perfect Matchings [7]). *The function* $f(n) = \frac{\ln n}{n}$ *is a threshold function for the property of* $G_{n,n,p}$ *having a perfect matching. Specifically, for* $p = \frac{\ln n + c(n)}{n}$ *where* $c(n) \to \infty$, *the probability of a perfect matching is* $O(\exp(-2\exp(-c(n))))$, *which goes to* 1 *as* $n \to \infty$.

3.6.1 Correlating to the Erdős-Rényi Model

Our rectangle sampling model can be shown to be dominating the standard $G_{n,n,p}$ model for a specific p by analyzing the individual edge survival probabilities as shown below:

▶ **Theorem 33** (Perfect Matching via Threshold). *The graph* G *obtained from rectangle sampling with* $p = \frac{\log^6 n}{n^3}$ *has a perfect matching with probability* $O(\exp(-2\exp(-\Omega(\log^6 n))))$.

Proof. By monotonicity, if the individual edge marginals in G are greater than that of G_{n,n,p_0} for some $p_0 = \Theta\left(\frac{\log^6 n}{n}\right) < p_{\text{edge}}$, then

$$\Pr[G \text{ has perfect matching}] \geq \Pr[G_{n,n,p_0} \text{ has perfect matching}]$$

As $p_0 = \Theta(\frac{\log^6 n}{n}) = \Theta(\frac{\ln n + (\log^6 n - \ln n)}{n})$, we have $c(n) = \Theta(\log^6 n - \ln n) = \Theta(\log^6 n)$. Applying Theorem 32 yields the stated bound. ◀

3.6.2 Robustness Under Vertex Deletion

▶ **Proposition 34** (Distribution Invariance Under Deletion). *Let* $S_L \subseteq V_1$ *and* $S_R \subseteq V_2$ *be arbitrary subsets with* $|S_L| = |S_R| = 2m$. *The induced graph* $G_{rest} := G|_{V_1 \setminus S_L, V_2 \setminus S_R}$ *obtained from our rectangle sampling process has the same distribution as a fresh rectangle sampling on the* $(n-2m) \times (n-2m)$ *complete bipartite graph with parameter* p.

Proof. The rectangle sampling process selects each rectangle independently with probability p. Rectangles are defined by pairs of rows and pairs of columns. For the induced subgraph G_{rest}, we condition on rectangles that use only rows from $V_1 \setminus S_L$ and columns from $V_2 \setminus S_R$.

Since the original rectangle selections are independent Bernoulli(p) random variables indexed by all possible rectangles, and we restrict to rectangles on $(V_1 \setminus S_L) \times (V_2 \setminus S_R)$, these restricted selections remain independent Bernoulli(p) random variables. This gives precisely the distribution of our random process of selecting rectangles from $K_{n-2m,n-2m}$ with probability p. ◀

▶ **Corollary 35** (Perfect Matching After Deletion). *After removing any* $2m$ *vertices from each partition where* $m = \Theta(n^{1/3})$, *the residual graph* G_{rest} *has a perfect matching with probability* $O(\exp(-2\exp(-\Omega(\log^6 n))))$. *Specifically, this probability goes to* 1 *as* $n \to \infty$. *So,* G_{rest} *has a perfect matching with high probability.*

Proof. By Proposition 34, the residual graph consists of the exact same distribution as if our random process was performed on $K_{n-2m,n-2m}$. Since $m = \Theta(n^{1/3}) = o(n)$, we have $\frac{\log^6(n-2m)}{n-2m} \sim \frac{\log^6 n}{n}$, so the edge survival probability p_{edge} remains $\Theta\left(\frac{\log^6 n}{n}\right) \gg \frac{\ln(n-2m)}{n-2m}$. Applying Theorem 32 to the resulting $(n-2m) \times (n-2m)$ graph yields the result. ◀

▶ Remark 36. This threshold-based proof of perfect matching survival complements the explicit Hall condition verification in Theorem 29. While Proposition 34 assumes arbitrary (possibly adversarial) deletion, our actual application in Section 4 deletes vertices based on sampled rectangles, which creates dependencies. Theorem 29 handles this carefully by conditioning on the deletion pattern. Both approaches confirm the robustness of our construction.

3.7 Sparse matrix

We can now formally define the matrix whose determinant complexity we analyze. Given the bipartite graph $G = (V_1 \sqcup V_2, E_G)$ constructed via rectangle sampling, where $|E_G| = \Theta(n \log^6 n)$, define the $n \times n$ *sparse symbolic matrix* X_G by setting $(X_G)_{i,j} = x_{i,j}$ (a formal variable) if $(i,j) \in E_G$, and $(X_G)_{i,j} = 0$ otherwise.

Our construction ensures X_G satisfies three critical properties:

1. **Sparsity:** $\Theta(\log^6 n)$ nonzero entries per row/column (Proposition 19),
2. **Non-vanishing:** $\mathsf{Det}(X_G) \neq 0$ (Corollary 28 or Corollary 33),
3. **Diagonal imbalance:** For large variable sets S with $|S| \geq |E_G|^{0.75}$, under random vertex permutations σ, at least $(1-\delta)|S|/2$ rectangles exhibit diagonal imbalance relative to S with high probability (Theorem 21).

Property (3) provides the geometric foundation for our lower bound: the rectangle sampling creates an asymmetric structure that, combined with random permutations, guarantees factor imbalance in the analysis of Section 4. The polynomial $\mathsf{Det}(X_G)$ over $\Theta(n \log^6 n)$ variables is our explicit candidate.

4 Lower Bounds Against Multilinear Formulas

▶ Remark 37. Our sparse determinant polynomial $\mathsf{Det}(X_G)$ is defined over the edge variables in E_G. Therefore, the total number of variables is $N = |E_G| = \Theta(n \log^6 n)$.

▶ **Lemma 38** (Count of Large Factors). *In the decomposition of Theorem 12 with $N = |E_G| = \Theta(n \log^6 n)$ total variables, the number of levels j for which $|X_{ij}| \geq N^{0.75} = \Theta(n^{0.75} \log^{4.5} n)$ is at least $\Theta(\log n)$.*

Proof. We require $(1/3)^j \cdot N \geq N^{0.75}$, which gives $(1/3)^j \geq N^{-0.25}$. Taking logarithms yields $j \log(1/3) \geq -0.25 \log N$, so $j \leq \frac{0.25 \log N}{\log 3}$. With $N = \Theta(n \log^6 n)$, we have $\log N = \Theta(\log n + \log\log n) = \Theta(\log n)$, so $j \leq \Theta(\log n)$. Since the total number of levels is $t = \Theta(\log N) = \Theta(\log n)$, and all levels $j \leq \Theta(\log n)$ satisfy $|X_{ij}| \geq N^{0.75}$, we have $\Theta(\log n)$ large factors per product. ◀

4.1 Sampling Rectangles and Concentration

We begin the second-round analysis by studying what happens when we sample rectangles uniformly at random from $\mathcal{R}$. This analysis connects the first-round construction to the factor imbalance needed for the lower bound.

▶ **Definition 39** (Rectangle Sampling Distributions). *We consider two probability distributions over sets of m rectangles from $\mathcal{R}$:*

- *Distribution μ^*: Sample m rectangles uniformly and independently from $\mathcal{R}$ with replacement. Let $\mathbf{R} = (R_1, \dots, R_m)$ denote the sampled sequence.*
- *Distribution μ: Sample m rectangles uniformly from $\mathcal{R}$ such that no two rectangles share a row or a column (i.e., the rectangles are vertex-disjoint). Let $\mathbf{R} = \{R_1, \dots, R_m\}$ denote the sampled set.*

For a set $S \subseteq E_G$ and permutation σ, define $Y_S(\sigma, \mathbf{R}) = |\{i \in [m] : R_i \in \mathcal{R}_{S,diag}(\sigma)\}|$ as the number of sampled rectangles that belong to $\mathcal{R}_{S,diag}(\sigma)$.

4.1.1 Choice of Distribution

In our lower bound proof[3], we apply the random restriction to both the determinant polynomial $\mathsf{Det}(X_G)$ and any multilinear formula computing it using the same distribution μ (rectangles with no shared rows or columns). The vertex-disjoint constraint in μ ensures that the random restriction creates a clean partition of variables: each variable in E_G belongs to at most one sampled rectangle, and the remaining graph G_{rest} after removing sampled vertices has a perfect matching with high probability (Theorem 29). This property is crucial for arguing that $\mathsf{Det}(X_G)$ maintains full rank after restriction. While our analysis uses distribution μ (vertex-disjoint rectangles), many of our concentration bounds are more naturally analyzed under the simpler distribution μ^* (sampling with replacement). Fortunately, the statistical distance between μ and μ^* is $o(1)$ for our parameter regime. With $m = \Theta(n^{1/3})$ rectangles sampled from $|\mathcal{R}| = \Theta(n \log^6 n)$ rectangles, the probability that any two sampled rectangles share a vertex under μ^* is at most $o(1)$ when $m^2 = o(n)$. Therefore, events that hold with probability q under μ^* also hold with probability $q - o(1)$ under μ. This allows us to freely use concentration results established for μ^* in our proofs while applying the restriction under μ.

4.2 Random Restriction and Variable Assignment

We now apply the random restriction, which operates on the variables from X_G. After sampling $m = \Theta(n^{1/3})$ rectangles uniformly at random from $\mathcal{R}$, we define a random restriction that assigns variables to create the $Y \sqcup Z$ partition (and rest of the variables to $\{0, 1\}$), which proceeds as follows:

- **Sampling rectangles:** Let $\mathbf{R} = (R_1, \dots, R_m)$ denote the $m = \Theta(n^{1/3})$ rectangles sampled uniformly and independently from $\mathcal{R}$ with replacement. Each rectangle R_i involves four vertices: two rows $\{r_1^i, r_2^i\}$ and two columns $\{c_1^i, c_2^i\}$. Let $S_L = \bigcup_{i=1}^m \{r_1^i, r_2^i\}$ and $S_R = \bigcup_{i=1}^m \{c_1^i, c_2^i\}$ denote the sets of rows and columns involved in the sampled rectangles. We have $|S_L|, |S_R| \le 2m$, with equality when rectangles are disjoint on vertices.
- **Rectangle variable assignment:** For each rectangle R_i defined by rows $\{r_1^i, r_2^i\}$ and columns $\{c_1^i, c_2^i\}$, we independently choose one of two configurations with equal probability. The configurations are as follows:
  - Configuration 1: Set $x_{r_1^i, c_1^i} \to y_i$ and $x_{r_2^i, c_2^i} \to z_i$, while setting $x_{r_1^i, c_2^i} = x_{r_2^i, c_1^i} = 1$ as constants.

[3] We also borrow inspiration for this line of argument from [14, Section 6].

  - Configuration 2: Set $x_{r_1^i,c_1^i} \to z_i$ and $x_{r_2^i,c_2^i} \to y_i$, while setting $x_{r_1^i,c_2^i} = x_{r_2^i,c_1^i} = 1$ as constants.

  This creates two diagonal corners as active variables assigned to y_i and z_i, and sets the other two diagonal corners to constants.
- **Remaining variable assignment:** Consider the submatrix obtained by removing the rows in S_L and columns in S_R from X_G. Let G_{rest} denote the induced bipartite graph on the remaining vertices $V_1 \setminus S_L$ and $V_2 \setminus S_R$. We established earlier that with high probability, regardless of which specific rows and columns are deleted, the graph G_{rest} has a perfect matching. Fix an arbitrary perfect matching M_{rest} in G_{rest}. For all remaining variables (those not in the m rectangles), if edge e belongs to M_{rest}, we set the corresponding variable to 1, and if edge e does not belong to M_{rest}, we set the corresponding variable to 0.
- **Final partition:** Let $Y = \{y_1, \dots, y_m\}$ and $Z = \{z_1, \dots, z_m\}$ denote the sets of active variables after restriction. The restriction creates a partition of the original $N = |E_G|$ variables into $|Y| = m$ variables in the Y class, $|Z| = m$ variables in the Z class, and $N - 2m$ variables set to constants.

  For a factor f_{ij} in the formula decomposition over variable set X_{ij}, let $Y_{ij} = Y \cap X_{ij}$ and $Z_{ij} = Z \cap X_{ij}$ denote the y and z variables in this factor respectively.

4.3 Balance Analysis via Binomial Distribution

▶ **Lemma 40** (Binomial Balance Distribution). *Consider a factor f_{ij} with $|X_{ij}| = M \geq N^{0.75}$ variables. Under the random restriction with $m = \Theta(n^{1/3})$ rectangles sampled from $\mathcal{R}$ according to distribution μ (or μ^*), the random variable $|Y_{ij}|$ (the number of y-variables in factor f_{ij}) is distributed as Binomial$(Y_S, 1/2)$ where $Y_S = \Omega(n^{1/12}\log^{-1.5} n)$ with probability at least $1 - \exp(-\Omega(n^{1/12}\log^{-1.5} n))$.*

Proof. By Theorem 23, when $|X_{ij}| \geq N^{0.75}$, with probability at least $1-\exp(-\Omega(n^{1/12}\log^{-1.5} n))$ over the rectangle sampling and permutation σ, the number of sampled rectangles that contribute to X_{ij} with diagonal imbalance is at least $Y_S = \Omega(n^{1/12}\log^{-1.5} n)$.

Each such rectangle R_ℓ with diagonal imbalance relative to X_{ij} has exactly one corner in X_{ij} that will be assigned to a y-variable and one corner that will be assigned to a z-variable. The assignment depends on whether Configuration 1 or Configuration 2 is chosen for rectangle R_ℓ, and this choice is made independently and uniformly at random for each rectangle.

Therefore, conditioned on having Y_S rectangles with diagonal imbalance relative to X_{ij}, the number of these rectangles that contribute a y-variable to X_{ij} (as opposed to a z-variable) is the sum of Y_S independent Bernoulli random variables, each with success probability 1/2. This gives $|Y_{ij}| \sim \text{Binomial}(Y_S, 1/2)$ with expectation $Y_S/2$ and variance $Y_S/4$. ◀

▶ **Proposition 41** (Single Factor Balance Probability). *For a factor f_{ij} with $|X_{ij}| = M \geq N^{0.75} = \Theta(n^{0.75}\log^{4.5} n)$ variables and $k = n^{1/32}$, we have $\Pr[\text{factor } f_{ij} \text{ is } k\text{-balanced}] = O(n^{-1/96}\log^{0.75} n)$.*

Proof. Let $\mathcal{G}_{ij}$ denote the event that the number of sampled rectangles with diagonal imbalance relative to X_{ij} is at least $Y_S = \Omega(n^{1/12}\log^{-1.5} n)$. By Theorem 23, we have $\Pr[\mathcal{G}_{ij}] \geq 1 - \exp(-\Omega(n^{1/12}\log^{-1.5} n))$.

Conditioned on $\mathcal{G}_{ij}$, by Lemma 40, $|Y_{ij}|$ is distributed as Binomial$(Y_S, 1/2)$ with expectation $Y_S/2$ and variance $Y_S/4$. The factor is k-balanced if $||Y_{ij}| - |Z_{ij}|| < 2k$, which is equivalent to $||Y_{ij}| - Y_S/2| < k$ since the Y_S rectangles contribute exactly Y_S variables split between Y_{ij} and Z_{ij}.

By the normal approximation to the binomial distribution, we have

$$\Pr[k\text{-balanced} \mid \mathcal{G}_{ij}] = O\left(\frac{k}{\sqrt{Y_S}}\right) = O\left(\frac{n^{1/32}}{\sqrt{n^{1/12}\log^{-1.5} n}}\right) = O(n^{-1/96}\log^{0.75} n).$$

Therefore, by the law of total probability,

$$\begin{aligned}\Pr[k\text{-balanced}] &= \Pr[k\text{-balanced} \mid \mathcal{G}_{ij}] \cdot \Pr[\mathcal{G}_{ij}] + \Pr[k\text{-balanced} \mid \mathcal{G}_{ij}^c] \cdot \Pr[\mathcal{G}_{ij}^c] \\ &\le O(n^{-1/96}\log^{0.75} n) \cdot 1 + 1 \cdot \exp(-\Omega(n^{1/12}\log^{-1.5} n)) \\ &= O(n^{-1/96}\log^{0.75} n),\end{aligned}$$

where the dominant term is the first summand since $\exp(-\Omega(n^{1/12}\log^{-1.5} n)) = o(n^{-1/96}\log^{0.75} n)$. ◀

4.4 Read-2 Structure and Dependent Factors

The balance events for different factors depend on two sources of randomness: which rectangles are sampled, and how each sampled rectangle is configured. We analyze these dependencies carefully to establish the read-2 structure.

▶ **Lemma 42** (Read-2 Structure of Balance Events). *Conditioned on the set of sampled rectangles $\mathbf{R} = (R_1, \ldots, R_m)$, the balance events for factors in a formula decomposition form a read-2 family with respect to the configuration choices.*

Proof. Given a fixed set of sampled rectangles $\mathbf{R} = (R_1, \ldots, R_m)$, the remaining randomness consists of m independent random variables $C_1, \ldots, C_m$, where $C_i \in \{1, 2\}$ indicates whether rectangle R_i uses Configuration 1 or Configuration 2. Each configuration choice C_i is made independently with equal probability.

The configuration choice C_i for rectangle R_i determines which two of the four edges in R_i become active variables y_i and z_i, and consequently determines how these two variables are assigned to the Y and Z classes. In the multilinear formula decomposition, within a single product (summand), the variable sets of different factors are disjoint. Therefore, each active variable belongs to exactly one factor within the product.

Since rectangle R_i contributes exactly two active variables (one assigned to y_i and one to z_i), and each active variable belongs to at most one factor in the product, the configuration choice C_i can influence the balance of at most two factors. Therefore, conditioned on $\mathbf{R}$, the balance events form a read-2 family with respect to the independent random variables $C_1, \ldots, C_m$. ◀

4.5 Imbalance in Multiple Factors

▶ **Theorem 43** (Multiple Factor Imbalance with High Probability). *Let $\mathcal{L}$ denote the set of $r = \Theta(\log n)$ levels j where $|X_{ij}| \ge N^{0.75}$. Consider a fixed summand $f_i = \prod_{j=1}^{t} f_{ij}$ in the formula decomposition. Under the random restriction, with probability at least $1 - \exp(-\Omega(\log^2 n))$ over both the choice of sampled rectangles $\mathbf{R}$ and the configuration choices, at least half of the factors in $\mathcal{L}$ exhibit k-imbalance with $k = n^{1/32}$.*

Proof. The random restriction involves two stages of randomness: sampling rectangles $\mathbf{R} = (R_1, \ldots, R_m)$ uniformly from $\mathcal{R}$ with replacement, then independently choosing configurations for each sampled rectangle. We condition on the first stage and apply concentration bounds for the second.

For each factor $\ell \in \mathcal{L}$, let $B_\ell = \mathbf{1}[\text{factor } \ell \text{ is } k\text{-balanced}]$. By Theorem 23, with probability at least $1 - \exp(-\Omega(n^{1/12} \log^{-1.5} n))$ over the rectangle sampling and permutation, every factor $\ell \in \mathcal{L}$ receives at least $\Omega(n^{1/12} \log^{-1.5} n)$ sampled rectangles with diagonal imbalance. Call this the good event $\mathcal{G}$.

Conditioned on any outcome $\mathbf{R} \in \mathcal{G}$, by Lemma 42, the balance events form a read-2 family with respect to the configuration choices. Moreover, by the binomial analysis from Proposition 41, each factor satisfies $\Pr[B_\ell = 1 \mid \mathbf{R}] = O(n^{-1/96} \log^{0.75} n)$. Therefore, the average conditional probability is $p = O(n^{-1/96} \log^{0.75} n)$.

We apply Theorem 16 with $k = 2$ to bound the probability that at least half the factors are balanced. For $q = 1/2$ and $p = O(n^{-1/96}) \ll 1/2$, the Kullback-Leibler divergence satisfies $D\,(1/2 \,\|\, p) \approx \frac{1}{2} \log \frac{1}{2p} = \Theta(\log n)$. Therefore, for any $\mathbf{R} \in \mathcal{G}$,

$$\Pr\left[\sum_{\ell \in \mathcal{L}} B_\ell \geq \frac{r}{2} \;\middle|\; \mathbf{R}\right] \leq \exp\left(-\frac{r}{k} \cdot D\,(1/2 \,\|\, p)\right) = \exp(-\Omega(\log^2 n)).$$

Since this bound holds uniformly for all $\mathbf{R} \in \mathcal{G}$, and the complement event $\mathcal{G}^c$ has probability at most $\exp(-\Omega(n^{1/12} \log^{-1.5} n))$, the total probability satisfies

$$\Pr\left[\sum_{\ell \in \mathcal{L}} B_\ell \geq \frac{r}{2}\right] \leq \exp(-\Omega(\log^2 n)) + \exp(-\Omega(n^{1/12} \log^{-1.5} n)) = \exp(-\Omega(\log^2 n)).$$

Therefore, with probability at least $1 - \exp(-\Omega(\log^2 n))$, at least half the factors in $\mathcal{L}$ exhibit k-imbalance. ◀

4.6 The Lower Bound via Union Bound

▶ **Theorem 44** (Lower Bound for Sparse Determinants). *Let X_G be the sparse matrix constructed with all the requisite combinatorial properties as constructed above (in 3.7). Then any multilinear formula computing $\mathsf{Det}(X_G)$ must have size $s = \exp(\Omega(\log^2 n))$.*

Proof. Suppose a multilinear formula F of size $s = n^{O(\log n)}$ computes $\mathsf{Det}(X_G)$. By Theorem 12, the formula F admits a decomposition $\sum_{i=1}^{s+1} f_i$ where each $f_i = \prod_{j=1}^{t} f_{ij}$ is a product of $t = \Theta(\log N) = \Theta(\log n)$ factors (since $N = \Theta(n \log^6 n)$). By Lemma 38, each product contains $r = \Theta(\log n)$ factors with at least $N^{0.75}$ variables.

Consider a fixed summand f_i. By Theorem 43, under the random restriction of Section 4.2, with probability at least $1 - \exp(-\Omega(\log^2 n))$, at least $\ell = r/2 = \Theta(\log n)$ factors in this summand exhibit k-imbalance with $k = n^{1/32}$.

Applying Proposition 11, on this high-probability event, the summand f_i satisfies

$$\mu_\sigma(f_i) \leq 2^m \cdot 2^{-\ell k} = 2^m \cdot 2^{-\Theta(\log n) \cdot n^{1/32}} = 2^m \cdot 2^{-\Omega(n^{1/32} \log n)},$$

where $m = \Theta(n^{1/3})$ is the number of active variables after the restriction.

Taking a union bound over all $s+1$ summands, if $s = n^{O(\log n)}$, then with probability at least $1 - (s+1) \cdot \exp(-\Omega(\log^2 n)) = 1 - o(1)$, all summands simultaneously exhibit $\ell = \Theta(\log n)$ imbalanced factors. By subadditivity of rank,

$$\mu_\sigma\left(\sum_{i=1}^{s+1} f_i\right) \leq \sum_{i=1}^{s+1} \mu_\sigma(f_i) \leq (s+1) \cdot 2^m \cdot 2^{-\Omega(n^{1/32} \log n)}.$$

For $s = n^{O(\log n)}$ and $m = \Theta(n^{1/3})$, since $2^{\Omega(n^{1/32} \log n)} \gg n^{O(\log n)}$ for sufficiently large n, we have with probability $1 - o(1)$,

$$\mu_\sigma(F) \leq n^{O(\log n)} \cdot 2^m \cdot 2^{-\Omega(n^{1/32} \log n)} \ll 2^m$$

On the other hand, the restricted determinant $\mathsf{Det}(X_G)$ maintains full rank after the restriction. By Theorem 29, with probability at least $1-\exp(-\Omega(\log^6 n))$, the residual graph G_{rest} obtained after removing the $2m$ rows and columns involved in sampled rectangles retains a perfect matching. The random restriction assigns variables as follows: for each of the m sampled rectangles, two opposite diagonal corners become active variables (y_i, z_i) while the other diagonal is set to constants; the remaining $|E_G| - 2m$ edges are assigned values according to the perfect matching M_{rest} in G_{rest} (matched edges set to 1, unmatched edges set to 0). Under this assignment, the restricted determinant expands as

$$\mathsf{Det}(X_G)|_{\text{restricted}} = \prod_{i=1}^{m}(y_i z_i - 1)\,.$$

This polynomial is defined over m variables in $Y = \{y_1, \ldots, y_m\}$ and m variables in $Z = \{z_1, \ldots, z_m\}$, yielding a coefficient matrix M_{Det} of dimension $2^m \times 2^m$ with full rank 2^m (since the determinant expansion contains all 2^m monomials over the Y variables paired with all 2^m monomials over the Z variables).

Therefore, $\mu_\sigma(\mathsf{Det}(X_G)) = 2^m$, contradicting the bound $\mu_\sigma(F) \ll 2^m$ established above. This contradiction shows that any multilinear formula computing $\mathsf{Det}(X_G)$ requires size $s = n^{\Omega(\log n)}$. ◀

References

1 Stuart J. Berkowitz. On computing the determinant in small parallel time using a small number of processors. *Inf. Process. Lett.*, 18(3):147–150, 1984. doi:10.1016/0020-0190(84)90018-8.

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

3 Richard P. Brent, David J. Kuck, and Koichi Maruyama. The parallel evaluation of arithmetic expressions without division. *IEEE Trans. Computers*, 22(6):532–534, 1973. doi:10.1109/T-C.1973.223745.

4 Suryajith Chillara, Christian Engels, Nutan Limaye, and Srikanth Srinivasan. A near-optimal depth-hierarchy theorem for small-depth multilinear circuits. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 934–945. IEEE Computer Society, 2018. doi:10.1109/FOCS.2018.00092.

5 Suryajith Chillara, Nutan Limaye, and Srikanth Srinivasan. Small-depth multilinear formula lower bounds for iterated matrix multiplication with applications. *SIAM J. Comput.*, 48(1):70–92, 2019.

6 Zeev Dvir, Guillaume Malod, Sylvain Perifel, and Amir Yehudayoff. Separating multilinear branching programs and formulas. *SIAM J. Comput.*, 45(4):1974–2002, 2016. doi:10.1137/130950215.

7 Alan Frieze and Michał Karoński. *Introduction to Random Graphs.* Cambridge University Press, Cambridge, 2016. doi:10.1017/CBO9781316339831.

8 Dmitry Gavinsky, Shachar Lovett, Michael E. Saks, and Srikanth Srinivasan. A tail bound for read-k families of functions. *Random Struct. Algorithms*, 47(1):99–108, 2015. URL: https://doi.org/10.1002/rsa.20532, doi:10.1002/RSA.20532.

9 Meena Mahajan and V. Vinay. Determinant: Combinatorics, algorithms, and complexity. *Chicago J. Theor. Comput. Sci.*, 1997, 1997. URL: http://cjtcs.cs.uchicago.edu/articles/1997/5/contents.html.

10 Colin McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, 141:148–188, 1989.

11 Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis.* Cambridge University Press, 2nd edition, 2017. doi:10.1017/9781316822135.

12 Ran Raz. Multi-linear formulas for permanent and determinant are of super-polynomial size. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 633–641. ACM, 2004. `doi:10.1145/1007352.1007353`.

13 Ran Raz. Separation of multilinear circuit and formula size. *Theory Comput.*, 2(1):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.

15 Ran Raz and Amir Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, July 15-18, 2009*, pages 128–139. IEEE Computer Society, 2009. `doi:10.1109/CCC.2009.16`.

16 Ramprasad Saptharishi. A survey of lower bounds in arithmetic circuit complexity. `https://github.com/dasarpmar/lowerbounds-survey/`, 2021. Manuscript.

17 Amir Shpilka and Amir Yehudayoff. Arithmetic circuits: A survey of recent results and open questions. *Found. Trends Theor. Comput. Sci.*, 5(3-4):207–388, 2010.

18 Leslie G. Valiant. Completeness classes in algebra. In *STOC*, pages 249–261. ACM, 1979.

19 Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201, 1979.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il