



Spiky Rank and Its Applications to Rigidity and Circuits

Lianna Hambardzumyan ^{*} Konstantin Myasnikov [†] Artur Riazanov [‡]
 Morgan Shirley [§] Adi Shraibman [¶]

February 26, 2026

Abstract

We introduce spiky rank, a new matrix parameter that enhances blocky rank by combining the combinatorial structure of the latter with linear-algebraic flexibility. A spiky matrix is block-structured with diagonal blocks that are arbitrary rank-one matrices, and the spiky rank of a matrix is the minimum number of such matrices required to express it as a sum. This measure extends blocky rank to real matrices and is more robust for problems with both combinatorial and algebraic character.

Our conceptual contribution is as follows: we propose spiky rank as a well-behaved candidate matrix complexity measure and demonstrate its potential through applications. We show that large spiky rank implies high matrix rigidity, and that spiky rank lower bounds yield lower bounds for depth-2 ReLU circuits, the basic building blocks of neural networks. On the technical side, we establish tight bounds for random matrices and develop a framework for explicit lower bounds, applying it to Hamming distance matrices and spectral expanders. Finally, we relate spiky rank to other matrix parameters, including blocky rank, sparsity, and the γ_2 -norm.

^{*}The University of Copenhagen, Denmark. lianna.hambardzumyan@gmail.com. This work is funded by the European Research Council (ERC) under grant agreement no. 101125652 (ALBA). Most of the work was done while the author was a postdoctoral researcher at the University of Victoria, Canada, funded by NSERC and at Hebrew University of Jerusalem, Israel, funded by ISF grants 921/22 and 2635/19.

[†]EPFL, Switzerland, kostyamyasso31@gmail.com. Supported by Swiss State Secretariat for Education, Research, and Innovation (SERI) under contract number MB22.00026.

[‡]EPFL, Switzerland, tunyash@gmail.com. Supported by Swiss State Secretariat for Education, Research, and Innovation (SERI) under contract number MB22.00026.

[§]Lund University, Sweden. morgan.shirley@cs.lth.se. Supported by Knut and Alice Wallenberg grant KAW 2023.0116.

[¶]School of Computer Science, The Academic College of Tel Aviv-Yaffo, Israel, adish@mta.ac.il.

1 Introduction

Studying the structural parameters of matrices has proven to be a powerful way to capture the complexity of various computational models and to establish lower bounds. A recurring theme in complexity theory is identifying “well-behaved” matrix parameters: parameters strong enough to capture meaningful lower bounds, structured enough to admit combinatorial or algebraic analysis, and flexible enough to connect to multiple computational models. Rank, rigidity, γ_2 -norm, and blocky rank are all examples of such measures.

In this work we introduce and study *spiky rank*, a new matrix parameter that we argue meets the criteria for a well-behaved complexity measure. Spiky rank is a generalization of the blocky rank defined for communication complexity applications by Hambardzumyan, Hatami and Hatami [HHH23] (Definition 1.1, Definition 1.2) which later found broad range of applications in complexity theory and beyond [AY24, PSS23, Wil24a, GHR25, GH25, AN25].

Spiky rank adds a linear algebraic flavor to the combinatorial nature of blocky rank. This allows us to connect spiky rank to complexity measures that are both algebraic and combinatorial in their nature, such as matrix rigidity and circuits with “algebraic” gates.

Both blocky rank and spiky rank fit into a common paradigm for defining complexity measures: we begin by specifying a class of simple objects that we define to have complexity 1, and then define the complexity of a general object as the minimum number of such simple objects needed to construct it. To define blocky and spiky rank, we define our matrices of complexity 1 as follows:

Definition 1.1: Blocky and Spiky Matrices

A **blocky matrix** is a matrix that can be obtained from the identity matrix by permuting rows or columns, duplicating rows or columns, and adding all-zero rows or columns. Equivalently, a blocky matrix is a matrix whose rows and columns can be permuted so that it consists of diagonal blocks that are all-one submatrices (of possibly different sizes), with all off-diagonal blocks equal to zero. See Figure 1a.

A **spiky matrix** is a blocky matrix in which each diagonal block is an arbitrary rank-one matrix (rather than an all-one matrix). Equivalently, a spiky matrix can be expressed as the entrywise product of a blocky matrix and a rank-one matrix. See Figure 1c.

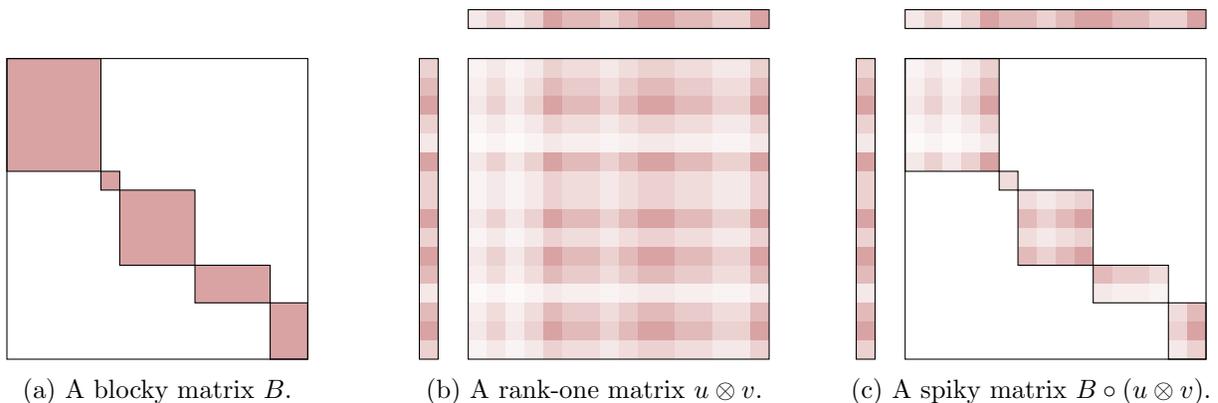


Figure 1: A spiky matrix is the entrywise product of a blocky matrix and a rank-one matrix.

We then define the complexity of general matrices as follows:

Definition 1.2: Blocky and Spiky Rank

A blocky matrix is said to have *blocky rank one*. The **blocky rank** of a matrix M , denoted $\text{br}(M)$, is the minimum number r such that M can be written as linear combination of r blocky rank one matrices.

A spiky matrix is said to have *spiky rank one*. The **spiky rank** of a matrix M , denoted $\text{spr}(M)$, is the minimum number r such that M can be written as a sum of r spiky rank one matrices.

1.1 Motivations for Blocky and Spiky Rank

Blocky rank. The study of blocky matrices and related decompositions has a surprisingly long and diverse history. Although the terminology “blocky matrix” is recent, the same objects have been rediscovered many times under different names across theoretical computer science and mathematics. For instance, fat matchings [Juk06, PR94], equivalence graphs [Fra82, Alo86], adjacency matrices of P_4 -free graphs [BBM⁺21, AN25], and equality matrices [Wil24a]. Various versions of blocky decompositions have been studied; blocky rank, *blocky covering* (union of blocky matrices), *blocky partitioning* (sum of blocky matrices) and more. Each of these variants has turned out to capture useful phenomena in different research areas:

- In *communication complexity* blocky decompositions have been used to study models of communication with access to the equality oracle [HHH23, PSS23, BBM⁺21, GHR25].
- In *operator theory*, it was observed [HHH23] that blocky matrices are the contractive idempotents of Schur multiplier algebras, making blocky rank a natural way to characterize the structure of idempotents in this setting (see also [GH25]).
- In *combinatorics*, blocky coverings of different families of graphs have been studied in [Fra82, Alo86, BBM⁺21, AY24].
- In *cryptography*, Applebaum and Nir [AN25] studied blocky covering to characterize families of graphs that can be realized with linear secret sharing schemes of constant size.
- In *circuit complexity*, blocky decompositions provide lower bounds: Jukna [Juk06] used blocky partitioning to derive bounds against depth-3 circuits with parity gates, and Avraham and Yehudayoff [AY24] showed that blocky partitioning lower bounds depth-2 threshold circuits.
- *Branching program* size lower bounds were obtained via blocky covering in [PR94].

Perhaps the most striking and far-reaching connection is due to Williams [Wil24a], who related blocky rank to the renowned Orthogonal Vectors Conjecture (OVC). In particular, Williams showed that to refute OVC – and consequently the Strong Exponential Time Hypothesis (SETH) – it would suffice to prove that the $2^n \times 2^n$ Disjointness matrix D_n satisfies $\text{br}(D_n) \leq 2^{o(n)}$. On the other hand, if one could show that $\text{br}(D_n) \geq 2^{\omega(n)}$, this would immediately yield exponential lower bounds for depth-2 exact threshold circuits by [AY24, Wil24a], resolving a notorious open problem in circuit complexity. This duality highlights the importance of studying blocky rank: either upper bounds or lower bounds on $\text{br}(D_n)$ would have transformative implications in algorithms, fine-grained complexity, and circuit lower bounds.

Spiky rank. As evident from this wide range of applications, blocky rank is a versatile measure. But it is fundamentally a combinatorial measure and, in particular, not a robust one. Small modifications to the matrix can cause its blocky rank to change drastically.

For example, in [Figure 2](#), the identity matrix has $\text{br}(I_n) = 1$, but if we place distinct weights on the diagonal, the blocky rank jumps to $\text{br}(D_n) = n$. Structurally these matrices are nearly the same – they are both just diagonal matrices – but blocky rank treats them in a completely different way. This lack of robustness makes blocky rank poorly suited for applications involving real-valued matrices or settings where algebraic properties matter. Spiky rank addresses this shortcoming by capturing the intuition that a diagonal matrix is expected to be simple regardless of its entries, in particular, $\text{spr}(D_n) = 1$.

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \quad D_n = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & n \end{bmatrix}$$

Figure 2: The identity matrix I_n (left) and the diagonal matrix D_n (right).

This extra flexibility makes spiky rank well-suited for problems that have both combinatorial and algebraic character. In this work, we demonstrate two such applications:

- Spiky rank lower bounds *matrix rigidity*, a central problem in circuit and communication complexity ([Section 1.2.1](#)),
- Spiky rank lower bounds the size of *depth-2 ReLU circuits*, which sit at the intersection of complexity theory and modern machine learning ([Section 1.2.2](#)).

Finally, note that spiky rank strictly strengthens blocky rank: every blocky decomposition is also spiky, so $\text{spr}(M) \leq \text{br}(M)$ for all M . Thus, lower bounds for spiky rank have the same implications as for blocky rank. On the other hand, upper bounds can be even more significant. As with blocky rank, upper bounds for spiky rank of the Disjointness matrix DISJ_n have far-reaching consequences: if the $2^n \times 2^n$ Disjointness matrix DISJ_n satisfies $\text{spr}(\text{DISJ}_n) \leq 2^{o(n)}$, then Orthogonal Vector Conjecture (and thus SETH) would be false [[Wil24b](#)].

1.2 Results

Our main contribution is conceptual: we introduce spiky rank as a new matrix complexity measure, demonstrate its applicability to problems in complexity theory, and establish its basic properties. On the technical side, we prove lower bounds for both random and explicit matrix families and relate spiky rank to other classical parameters. Our main results are:

1. We demonstrate the natural applicability of spiky rank to two problems in complexity theory: matrix rigidity ([Section 1.2.1](#)) and ReLU circuits ([Section 1.2.2](#)).
2. We show that most Boolean matrices have spiky rank $\Omega(N/\log N)$ (matching blocky rank), while most real matrices have spiky rank $\Omega(N)$ ([Section 1.2.3](#)).
3. We give the following bounds on the spiky rank of specific $N \times N$ matrices of interest:

- $\Omega(\sqrt{\log N})$ for the 1-Hamming Distance matrix (Section 6.2), simultaneously improving the previous $\log \log N$ bound for blocky rank as well [HHH23].
 - $\Omega(\log N)$ for adjacency matrices of certain expander graphs (Section 6.3).
 - $\Omega(\log N / \log \log N)$ for the Inner Product and Disjointness matrix (Section 6.4) (likely far from optimal but it represents a first step toward stronger bounds).
4. For Boolean matrices, we establish a dimension-free relation between blocky and spiky rank, implying in particular that constant spiky rank yields constant blocky rank (Section 1.2.5). We further bound spiky rank in terms of matrix sparsity (Section 7.3), separate it from the γ_2 -norm (Section 7.4), and highlight gaps between the approximate and sign variants of blocky and spiky rank (Appendix A).

1.2.1 Application to Matrix Rigidity

Matrix rigidity, introduced by Valiant [Val77] and Grigoriev [Gri80], asks how far a matrix is from having low rank: the rank- r rigidity $\mathcal{R}_M(r)$ is the minimum number of entry modifications required to reduce the rank of M to r .

Highly rigid matrices are central to long-standing open problems in complexity theory. Valiant showed that if $M \in \mathbb{R}^{N \times N}$ satisfies $\mathcal{R}_M(\varepsilon N) \geq \Omega(N^{1+\delta})$ for some $\varepsilon, \delta > 0$, then any log-depth linear circuit computing the linear transformation $x \mapsto Mx$ must have superlinear size. Additionally, Razborov [Raz89] linked rigidity to communication complexity, showing that matrices in PH^{cc} – the communication complexity analogue of the polynomial hierarchy – cannot be too rigid. In particular, if a $N \times N$ Boolean matrix $M \notin \text{PH}^{cc}$, then its rigidity over \mathbb{F}_2 must satisfy

$$\mathcal{R}_M^{\mathbb{F}_2}(2^{(\log \log N)^{\omega(1)}}) \geq \frac{N^2}{2^{(\log \log N)^{O(1)}}}.$$

Hence, constructing sufficiently rigid matrices would yield lower bounds for PH^{cc} and ultimately separate it from PSPACE^{cc} – both of which remain major open problems in communication complexity.

Despite extensive study, progress has been limited on finding explicit rigid matrices: the best explicit constructions fall well short of the parameters needed for Valiant’s or Razborov’s applications.

A direct link between rigidity and spiky rank is rather straightforward to show: matrices with large spiky rank are also highly rigid.

Theorem 1.1. *Let M be a matrix and $0 < r \leq \text{spr}(M)$. Then,*

$$\mathcal{R}_M(r) \geq \frac{(\text{spr}(M) - r)^2}{4}.$$

This has two key consequences. First, for Valiant’s target rank εN , an explicit matrix with $\text{spr}(M) = \Omega(N)$ would achieve the strongest possible rigidity bound $\Omega(N^2)$. Note that, since Boolean matrices have spiky rank at most $N / \log N$, the only candidate matrices with spiky rank $\Omega(N)$ are the real matrices.

Second, for Razborov’s target rank $2^{(\log \log N)^{\omega(1)}}$, we would need to find an explicit matrix with spiky rank at least (see Corollary 4.7)

$$\frac{N}{2^{(\log \log N)^{O(1)}}}.$$

This shows that spiky rank is not only a structural parameter, but also a potential tool for constructing explicit rigid matrices. With this goal in mind, we highlight the following open challenges to guide future research in this direction.

Major Open Problem 1.2. Find an explicit real matrix $M \in \mathbb{R}^{N \times N}$ with $\text{spr}(M) = \Omega(N)$.

Major Open Problem 1.3. Find an explicit Boolean matrix $M \in \{0, 1\}^{N \times N}$ with

$$\text{spr}(M) = \frac{N}{2^{(\log \log N)^{O(1)}}}.$$

Although most Boolean matrices have spiky rank at least $\Omega(N/\log N)$ and most real matrices have $\Omega(N)$ (see [Section 1.2.3](#)), constructing explicit examples that meet these bounds remains elusive. We expect that solving either problem above will be quite challenging. At present, the best lower bounds we can prove for explicit matrices are of order $\Omega(\log N)$ (see [Section 1.2.4](#)).

Finally, we remark that [Theorem 1.1](#) allows known non-rigidity results to translate into upper bounds on spiky rank. For instance, Alman and Williams [[AW17](#)] proved that for sufficiently small $\varepsilon > 0$, the Inner Product matrix IP_n of size $2^n \times 2^n$ satisfies

$$\mathcal{R}_{\text{IP}_n}(2^{n(1-f(\varepsilon))}) \leq 2^{n(1+\varepsilon)},$$

where $f(\varepsilon) = \Theta(\varepsilon^2/\log(1/\varepsilon))$. By [Theorem 1.1](#), this yields an upper bound $\text{spr}(\text{IP}_n) \leq 2^{(1-\delta)n}$ for some constant $\delta > 0$, improving on the trivial bound of $2^n/n$ (see [Section 1.2.3](#)).

1.2.2 Application to Circuit Complexity

Connections between blocky rank and circuit complexity were first observed by Jukna [[Juk06](#)], and later by Avraham and Yehudayoff [[AY24](#)], who showed that the blocky rank of a matrix lower bounds the size of any $\Sigma \circ \text{LTF}$ circuit computing it up to a certain loss factor (implicit in [[HP10](#)]). Here, $\Sigma \circ \text{LTF}$ denotes linear combinations of *linear threshold functions* (LTFs), where an LTF is a Boolean function of the form $T(x) = \mathbb{1}[\langle w, x \rangle \geq \alpha]$ for some $w \in \{0, 1\}^n$ and $\alpha \in \mathbb{R}$.

Naturally this connection extends to spiky rank and to $\Sigma \circ \text{ReLU}$ circuits, where ReLU functions generalize threshold functions by outputting $\max\{0, \langle w, x \rangle - \alpha\}$. The following proposition admits a simple proof, yet it is conceptually important.

Proposition 1.4. *For $M: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \mathbb{R}$, let s be the size of a $\Sigma \circ \text{ReLU}$ circuit that computes M . Then*

$$s \geq \frac{\text{spr}(M)}{3(n+1)}.$$

Observe that $\Sigma \circ \text{LTF}$ circuits can be simulated by $\Sigma \circ \text{ReLU}$ circuits with only constant factor blowup in size, since

$$\mathbb{1}[\langle w, x \rangle \geq \alpha] = \max\{0, \langle w, x \rangle - \alpha + 1\} - \max\{0, \langle w, x \rangle - \alpha\}.$$

Linear threshold and ReLU circuits are central models in both complexity theory and machine learning. $\Sigma \circ \text{LTF}$ circuits capture linear combinations of halfspaces and have been studied since the early days of threshold logic. On the other hand, ReLU gates have become the dominant activation function in modern neural networks, thus $\Sigma \circ \text{ReLU}$ circuits are the basic building blocks of modern neural networks. However, our understanding of their expressive power from a Boolean complexity perspective remains limited. Understanding the power and limitations of these models is therefore both an important task in circuit complexity and in machine learning.

Despite decades of work, the best known lower bounds for these circuits are very weak: even for $\Sigma \circ \text{LTF}$, essentially no superlinear lower bounds are known for explicit functions [ROS94]. Williams [Wil18] gave the strongest results to date, proving that for every k there exists a function in NP that requires sparsity n^k in both $\Sigma \circ \text{LTF}$ and $\Sigma \circ \text{ReLU}$ circuits. For explicit functions, the strongest known result is due to Kane and Williams [KW16], who showed that even the stronger class $\text{LTF} \circ \text{LTF}$ requires size $\Omega(n^{3/2})$ to compute the n -variate Andreev function. Mukherjee and Basu [MB17] extend this to prove a lower bound of $\Omega(n^{1-\delta})$ for $\text{LTF} \circ \text{ReLU}$ circuits computing the Andreev function.

Our theorem shows that spiky rank serves as a lower bound measure for depth-two neural networks with ReLU activation, just as blocky rank does for $\Sigma \circ \text{LTF}$ circuits. In particular, proving spiky rank lower bounds of $\omega((\log N)^{5/2})$ for explicit $N \times N$ matrices would immediately yield new circuit lower bounds. At present, our lower bounds are of order $\log N$; the first milestone toward this goal would be to establish *any* super-logarithmic spiky rank lower bounds for explicit matrices.

Open Problem 1.5. Find an explicit real matrix $M \in \mathbb{R}^{N \times N}$ with $\text{spr}(M) = \omega((\log N)^{5/2})$.

1.2.3 Bounds for Random Matrices

How large can the spiky rank of a matrix be? Trivially, both the spiky rank and blocky rank of an $N \times N$ Boolean matrix are at most N . For blocky rank, this bound can be improved to $\frac{N}{\log N}$ [PR94, AY24], and therefore the same upper bound applies to spiky rank as well. A natural question is whether spiky rank admits a substantially stronger upper bound. Avraham and Yehudayoff [AY24] showed that most matrices are maximally hard for blocky rank. We prove that the same phenomenon holds for spiky rank.

Theorem 1.6 (Informal). *A random Boolean matrix of size $N \times N$ has spiky rank of order $\Omega(\frac{N}{\log N})$ with high probability.*

For real matrices, the situation is more extreme: the trivial upper bound of N turns out to be essentially tight.

Theorem 1.7 (Informal). *A random real matrix of size $N \times N$ has spiky rank at least $\frac{N}{2}$ with high probability.*

This mirrors what is known for many other natural measures in complexity theory, such as circuit complexity and communication complexity, where random functions or matrices almost always exhibit the largest possible complexity.

1.2.4 Bounds for Specific Matrices.

The lower bounds discussed above, while strong, are for random (non-explicit) matrices. For applications in complexity theory, however, it is essential to establish lower bounds for explicit families of matrices.

For blocky rank, the best known lower bounds for explicit families of matrices are of order $\Omega(\log N)$ (Greater-Than [AY24], Inner-Product, Disjointness [HHH23]).

For spiky rank, we develop a general lower bound framework (Theorem 6.1) and apply it to two examples: adjacency matrices of the Hamming cube and expander graphs. The common properties of these graphs that we use are:

Thinness: every small subgraph has smaller edge-to-node ratio compared to the entire graph;

Induced matchings: there are large induced matchings in all large subgraphs.

This framework can only yield lower bounds up to $\Theta(\log N)$, the bottleneck is in the use of a variation of a lemma from [AY22, Lemma 12].

Hamming distance. Consider the matrix HD_1^n of size $N \times N$ with $N = 2^n$, whose rows and columns are indexed by all n -bit strings, and where

$$\text{HD}_1^n(x, y) = \begin{cases} 1 & \text{if } x \text{ and } y \text{ differ in exactly 1 coordinate,} \\ 0 & \text{otherwise.} \end{cases}$$

Applying our framework to this matrix yields the following bound:

Theorem 1.8. $\text{spr}(\text{HD}_1^n) \geq \Omega(\sqrt{\log N})$.

This bound is not far from the optimal; upper bound on the spiky rank is $\text{spr}(\text{HD}_1^n) \leq \text{br}(\text{HD}_1^n) \leq \log N$, where the last upper bound follows from the following blocky matrix decomposition: Let M_i be the matrix such that $M_i(x, y) = 1$ iff $x_i \neq y_i$ and all the other entries of x and y coincide. Then M_i is a permutation matrix, so $\text{br}(M_i) = 1$ and $\text{HD}_1^n = M_1 + \dots + M_n$.

This lower bound is quite remarkable considering that HD_1^n matrix is very simple with respect to many other matrix complexity measures, as we discuss in Section 1.2.5.

Open Problem 1.9. Is $\text{spr}(\text{HD}_1^n) \geq \Omega(\log N)$?

Expander graphs. As a second application, let G be an (N, d, λ) -spectral expander: a d -regular graph with the second largest eigenvalue of the adjacency matrix not exceeding λ , denote its adjacency matrix with M_G . We prove:

Theorem 1.10. $\text{spr}(M_G) \geq \Omega(\min(\frac{d}{\lambda}, \log \frac{N}{d^2}))$.

This implies (see Corollary 6.4) an explicit $\Omega(\log N)$ lower bound for the spiky rank of adjacency matrices of $(N, \Theta(\log^2 N), \Theta(\log N))$ -spectral expanders.

Inner Product. The Boolean inner product matrix is defined by $\text{IP}_n(x, y) = \sum_{i \in [n]} x_i y_i \pmod 2$ for $x, y \in \{0, 1\}^n$. For IP_n we obtain bounds that are weak but improve over the trivial ones:

$$\frac{n}{\log n} \leq \text{spr}(\text{IP}_n) \leq 2^{(1-\delta)n},$$

for some constant $\delta > 0$. The lower bound is established in Section 6.4, while the upper bound follows from non-rigidity results (Section 1.2.1).

Disjointness. The Disjointness matrix DISJ_n is the $2^n \times 2^n$ Boolean matrix whose rows and columns are indexed by $\{0, 1\}^n$, with $\text{DISJ}_n(x, y) = 1$ iff x and y have no common 1's. Similarly, for the Disjointness matrix we observe the same lower bound as for IP_n (Section 6.4), so combined with Williams' upper bound [Wil24a] we have

$$\frac{n}{\log n} \leq \text{spr}(\text{DISJ}_n) \leq 5^{n/5}.$$

We believe the true lower bounds for DISJ_n and IP_n are much closer to the upper bounds, and in particular ask:

Open Problem 1.11. Is it true that $\text{spr}(\text{IP}_n) = \omega(n)$ or $\text{spr}(\text{DISJ}_n) = \omega(n)$?

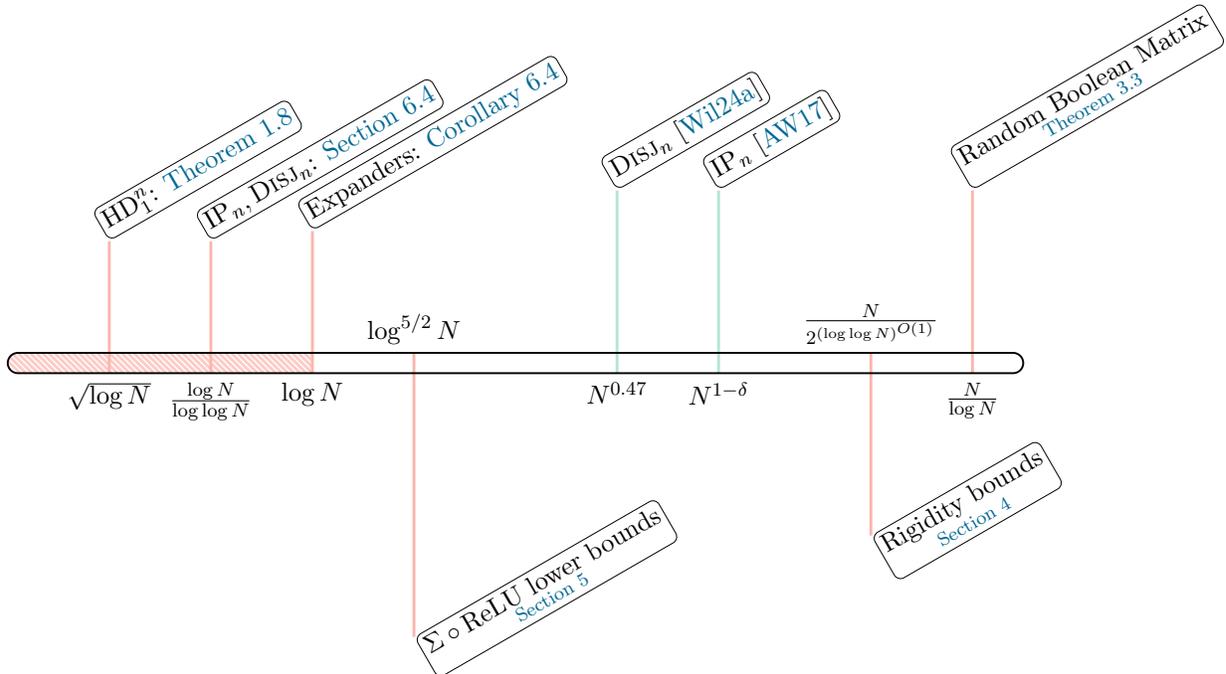


Figure 3: The picture illustrates the upper and lower bounds for spiky rank: above the scale there are known **lower bounds** and **upper bounds**. Below the scale are consequences of the corresponding lower bounds. The dashed part of the scale represents the state of the art of explicit lower bounds. Here N is always the size of the matrix, so for HD_1^n , DISJ_n , and IP_n , $N = 2^n$.

1.2.5 Relations to other matrix parameters

When introducing a new matrix parameter, a natural and useful question to understand is how it compares to existing ones: is it stronger, weaker, or qualitatively different?

Blocky rank. Since spiky rank can be seen as an algebraic strengthening of blocky rank, the first step is to ask whether this additional flexibility actually increases its power. Is blocky rank equivalent to spiky rank?

Question 1.12. Does every Boolean matrix M satisfy $\text{br}(M) \leq \text{poly}(\text{spr}(M))$?

We answer this question qualitatively: there is a dimension-free relation between spiky and blocky rank:

Theorem 1.13. For any Boolean matrix M , $\text{br}(M) \leq \text{spr}(M)^{O(\text{spr}(M))}$.

This situation is reminiscent of tantalizing *log-rank conjecture*. There, too, a combinatorial parameter - the partition number - is vastly strengthened by adding algebraic property via matrix rank. Similar to [Theorem 1.13](#), an exponential upper bound is known, and the challenge is whether this relation can be improved to a quasi-polynomial one.

γ_2 -norm. Since our main goal is to establish lower bounds on spiky rank, it is natural to compare it with matrix parameters that are more amenable to analysis – particularly matrix norms. A

well-studied example, with many applications in complexity theory, is the γ_2 -norm (see [BHT25] and references within). For a matrix M , it is defined as

$$\gamma_2(M) = \min_{U, V: M=UV} \|U\|_{\text{row}} \|V\|_{\text{col}},$$

where $\|U\|_{\text{row}}$ denotes the maximum ℓ_2 -norm of the rows of U , and $\|V\|_{\text{col}}$ denotes the maximum ℓ_2 -norm of the columns of V .

Strong lower bounds are known for γ_2 -norm for functions of interest; for instance, $\gamma_2(\text{IP}_n) \geq N^{\Omega(1)}$, where $\text{IP}_n(a, b) = \sum_{i \in [n]} a_i b_i \bmod 2$ (see e.g. [LS+09]). This motivates comparing γ_2 with rank-based parameters such as blocky rank and spiky rank. For blocky rank, it is known that for every Boolean matrix M , $\text{br}(M) \geq \log \gamma_2(M)$ [HHH23, Proposition 3.1]. For spiky rank, while an analogous inequality currently seems out of reach, we observe a simple non-explicit separation (Section 7.4):

Theorem 1.14. *There exists a Boolean matrix M such that $\text{spr}(M) \geq \Omega(\gamma_2^2(M))$.*

The reverse direction – whether blocky rank is upper-bounded by a function of γ_2 -norm – is a conjecture asked in [HHH23], and is equivalent to a conjecture in operator theory. Very recently, Goh and Hatami [GH25] proved that for $N \times N$ Boolean matrix M , $\text{br}(M) \leq 2^{\gamma_2(M)} (\log N)^2$, and the same bound also applies to spiky rank.

Sparsity. One of the most basic and easily computable parameters of a matrix is its *sparsity*, the number of non-zero entries. We show that sparse matrices necessarily have low spiky rank, a fact that becomes useful when relating spiky rank to matrix rigidity (Lemma 4.3). In particular, for any real matrix M with sparsity m , we prove

$$\text{spr}(M) \leq \sqrt{m}.$$

For Boolean matrices, this bound can be sharpened, and even expressed directly in terms of blocky rank (Theorem 7.6). Specifically, for any Boolean matrix M with sparsity m , it holds

$$\text{br}(M) \leq O\left(\frac{\sqrt{m} \log \log m}{\log m}\right).$$

Sign and approximate measures. Some computational models are naturally characterized not by exact matrix parameters but by their approximate or sign variants. For instance, randomized communication complexity is lower-bounded by approximate rank, while unbounded-error communication complexity is captured by sign-rank. Formally, for a matrix measure $\mu: \{0, 1\}^{N \times N} \rightarrow \mathbb{R}$, we define its approximate version by

$$\mu_\varepsilon(M) := \inf\{\mu(M') \mid \|M - M'\|_\infty \leq \varepsilon, M' \in \mathbb{R}^{N \times N}\}.$$

Similarly, the sign version of μ is $\mu_\pm(M) := \inf\{\mu(M') \mid \text{sign}(M) = \text{sign}(M')\}$.

Approximate measures tend to be stronger than their exact counterparts, and often defy intuition developed for them. For example, the approximate version of the log-rank conjecture is false [CMS20]. In case of spiky rank, it is sharply separated from all related approximate and sign measures of a matrix: $\text{spr}(\text{HD}_1^n) \geq \sqrt{\log N}$ (Theorem 1.8), while $\text{rk}_\pm(\text{HD}_1^n) = O(1)$ [GHIS25], $\gamma_{2, \varepsilon}(\text{HD}_1^n) = O(1)$ [CHZZ24], $\text{br}_\varepsilon(\text{HD}_1^n) = \text{poly}(\log \log N)$ (see Appendix A.3).

This highlights an interesting contrast between the spiky and blocky rank. For blocky rank, there is a dimension-free relation with its sign version: we show in Appendix A.1 that $\text{br}(M) \leq$

$2^{\text{br}_\pm(M)}$. By contrast, for spiky rank we have $\text{spr}(\text{HD}_1^n) \geq \Omega(\sqrt{\log N})$ but $\text{spr}_\pm(\text{HD}_1^n) \leq \text{rk}_\pm(\text{HD}_1^n) = O(1)$, ruling out any dimension-free relation between spr and spr_\pm .

Thus, proving lower bounds for sign spiky rank is an even harder challenge. At the same time, this opens a promising direction to seek explicit versions of the known non-explicit bounds below the logarithmic barrier of our framework. For example, we know that the adjacency matrix of a random degree- d graph has sign spiky rank $\Omega(d)$ (Lemma 7.8). A natural candidate for derandomization is a spectral expander, leading to the following question:

Question 1.15. What is $\text{spr}_\pm(M_G)$ for a spectral expander G ?

Organization. Section 2 introduces preliminaries. In Section 3, we prove the existence of Boolean and real matrices with high spiky rank. The applications to matrix rigidity and circuit complexity are presented in Section 4 and Section 5, respectively. Section 6 establishes spiky rank bounds for specific explicit matrices. In Section 7, we explore connections between spiky rank and other matrix parameters. Finally, Appendix A demonstrates separations between the approximate and sign versions of blocky and spiky rank.

2 Preliminaries

Our convention in this paper is to use capital letters for matrix dimensions, e.g. $A \in \mathbb{R}^{M \times N}$. We use the notation \circ for the entrywise product. The notation $[k]$ represents the set of integers $\{1, \dots, k\}$.

For a field \mathbb{F} , we denote $0_{\mathbb{F}}$ for the additive identity and $1_{\mathbb{F}}$ for the multiplicative identity. Addition, multiplication, and the entrywise product of matrices over \mathbb{F} are denoted $+_{\mathbb{F}}$, $\cdot_{\mathbb{F}}$, and $\circ_{\mathbb{F}}$ respectively.

For a matrix $M \in \mathbb{F}^{N \times N}$ its sparsity $\text{spar}(M)$ is defined as the number of its non- $0_{\mathbb{F}}$ elements.

We write $\mathbb{1}_S$ for $S \in [N] \times [N]$ for the matrix in $\{0, 1\}^{N \times N}$, where $\mathbb{1}_S[i, j] = 1$ iff $(i, j) \in S$.

Blocky and spiky measures. As some of our results generalize to arbitrary fields, we rewrite Definitions 1.1 and 1.2 here with respect to an arbitrary field \mathbb{F} .

Definition 2.1. A matrix $M \in \mathbb{F}^{M \times N}$ is *blocky* if there is an integer k and two collections of disjoint sets $S_1, \dots, S_k \subseteq [M]$ and $T_1, \dots, T_k \subseteq [N]$ such that $M[i, j] = 1_{\mathbb{F}}$ if $(i, j) \in \bigcup_{\ell \in [k]} S_\ell \times T_\ell$ and $M[i, j] = 0_{\mathbb{F}}$ otherwise.

A matrix $M \in \mathbb{F}^{M \times N}$ is *spiky* if there exists a blocky matrix B and a rank-one matrix C (where rank is taken over \mathbb{F}) such that $M = B \circ_{\mathbb{F}} C$.

Definition 2.2. The *blocky rank* of a matrix $M \in \mathbb{F}^{M \times N}$, denoted $\text{br}^{\mathbb{F}}(M)$, is the minimum r such that $M = \sum_{i \in [r]} \alpha_i B_i$ where each $\alpha_i \in \mathbb{F}$ and each B_i is blocky.

The *spiky rank* of a matrix $M \in \mathbb{F}^{M \times N}$, denoted $\text{spr}^{\mathbb{F}}(M)$, is the minimum r such that $M = \sum_{i \in [r]} S_i$ where each S_i is spiky.

When the underlying field is \mathbb{R} , we write simply $\text{br}(M)$ and $\text{spr}(M)$.

Here are some easy facts about spiky rank:

Fact 2.3 (Subadditivity). *For any field \mathbb{F} and matrices $A, B \in \mathbb{F}^{M \times N}$, $\text{spr}^{\mathbb{F}}(A + B) \leq \text{spr}^{\mathbb{F}}(A) + \text{spr}^{\mathbb{F}}(B)$.*

Fact 2.4 (Upper bounds). *For any field \mathbb{F} and matrix $A \in \mathbb{F}^{M \times N}$, $\text{spr}^{\mathbb{F}}(A) \leq \text{rank}^{\mathbb{F}}(A)$ (where $\text{rank}^{\mathbb{F}}$ is the rank over \mathbb{F}) and $\text{spr}^{\mathbb{F}}(A) \leq \text{br}^{\mathbb{F}}(A)$.*

Sign blocky rank and sign spiky rank. We study the sign versions blocky rank and spiky rank over the reals. In the following definition, we assume that the matrices have no 0-entries. For a matrix A , the value $\text{sign}(A)$ is the sign function applied element-wise:

$$\text{sign}(A)[i, j] = \begin{cases} 1 & \text{if } A[i, j] > 0 \\ -1 & \text{if } A[i, j] < 0 \end{cases}.$$

Definition 2.5. Let A be a matrix in $\mathbb{R}^{M \times N}$.

- The *sign blocky rank* of A , denoted $\text{br}_{\pm}(A)$, is defined as

$$\text{br}_{\pm}(A) = \min_{B \in \mathbb{R}^{M \times N}} \{\text{br}(B) : \text{sign}(A) = \text{sign}(B)\}.$$

- The *sign spiky rank* of A , denoted $\text{spr}_{\pm}(A)$, is defined as

$$\text{spr}_{\pm}(A) = \min_{B \in \mathbb{R}^{M \times N}} \{\text{spr}(B) : \text{sign}(A) = \text{sign}(B)\}.$$

Fact 2.6 (Upper bounds for sign measures.). *For any matrix $A \in \mathbb{R}^{M \times N}$, $\text{br}_{\pm}(A) \leq \text{br}(A)$ and $\text{spr}_{\pm}(A) \leq \text{spr}(A)$.*

3 Non-explicit lower bounds for spiky rank

3.1 Lower bound for random Boolean matrices

We first count the number of $N \times N$ Boolean matrices that have spiky rank r for a fixed number $r \leq N$. It is sufficient to count the number of ± 1 -matrices of with spiky rank r . Indeed, if $M \in \{0, 1\}^{N \times N}$, then its signed version $M' \in \{\pm 1\}^{N \times N}$ defined as $M' = J - 2M$ with J being the $N \times N$ all-ones matrix, has $\text{spr}(M') \leq r + 1$.

In fact, we will actually count the number of ± 1 -matrices with sign spiky rank r (Definition 2.5). As this measure is upper-bounded by spiky rank, this gives us an estimate for the number of matrices with spiky rank r as well.

The idea is to use Warren's theorem [War68] to count the number of possible sign patterns. The same idea and argument were used to count the number of matrices with sign rank r in [AMY16]. Before stating the theorem, let us set it up first.

Let P_1, P_2, \dots, P_m be real polynomials in ℓ variables each. For $x \in \mathbb{R}^{\ell}$, such that no polynomial among $\{P_i\}$ vanishes at x , its sign pattern is the vector

$$(\text{sign}(P_1(x)), \text{sign}(P_2(x)), \dots, \text{sign}(P_m(x))) \in \{-1, 1\}^m.$$

Theorem 3.1 ([War68]). *Let P_1, P_2, \dots, P_m real polynomials, each in ℓ variables and of degree at most k . If $m \geq \ell$, then the number of different sign pattern for x ranging over all \mathbb{R}^{ℓ} is at most $(4ekm/\ell)^{\ell}$.*

Lemma 3.2. *The number of $N \times N$ sign matrices with sign spiky rank $r \leq N/2$ is at most N^{6rN} .*

Proof. Let M be a $N \times N$ sign matrix with sign spiky rank r . Let $A := A_1 + \dots + A_r \in \mathbb{R}^{N \times N}$ be the sign representation of M : $M = \text{sign}(A_1 + \dots + A_r)$. Let us represent each $A_i = B_i \circ C_i$ where B_i is a blocky matrix and C_i is a rank-1 matrix. There are at most N^{4rN} collections B_1, \dots, B_r , so let us fix one and count the number of potential matrices M that appear as we change C_1, \dots, C_r .

With fixed B_1, \dots, B_r each entry of A can be written as a degree-2 polynomial in $2Nr$. Let for each $i \in [r]$, $C_i = x_i^T y_i$ for some vectors $x_i, y_i \in \mathbb{R}^N$. Then for $j, k \in [N]$ we have $A_{jk} = \sum_{i \in [r]} (B_i)_{jk} \cdot x_{ij} y_{ik}$. Then by [Theorem 3.1](#) we get that the number of different sign matrices M of form $\text{sign}(B_1 \circ C_1 + \dots + B_r \circ C_r)$ with fixed B_1, \dots, B_r is at most $\left(4e \cdot \frac{2N^2}{2rN}\right)^{2rN} \leq N^{2rN}$, so in total there are at most $N^{4rN} \cdot n^{2rN}$ matrices of sign spiky rank r . \square

Combining the lemma with the observation that $\text{spr}_{\pm}(M) \leq \text{spr}(M)$, we get that a uniformly random Boolean matrix has high spiky rank with high probability. More precisely,

Theorem 3.3. *If M is a $N \times N$ uniformly random Boolean matrix, then*

$$\Pr \left[\text{spr}(M) \geq \frac{N}{12 \log N} \right] \geq \Pr \left[\text{spr}_{\pm}(M) \geq \frac{N}{12 \log N} \right] \geq 1 - 2^{-\frac{N^2}{2}}.$$

Proof. The number of matrices of sign spiky rank less than $N/(12 \log N)$ is at most $N^{6N^2/(12 \log N)} \leq 2^{N^2/2}$ by [Lemma 3.2](#) and the total number of $N \times N$ sign matrices is 2^{N^2} . \square

3.2 Lower bound for real matrices

The spiky rank of any $N \times N$ real matrix is trivially bounded above by N , since the real rank of a matrix is at most N , and every rank-one matrix is also a spiky matrix. For Boolean matrices, even blocky rank can be bounded above by $O(N/\log N)$. However, for general real matrices, we show that the trivial upper bound is essentially tight: there exist matrices whose spiky rank is $\Omega(N)$. The key point in the proof as in [Lemma 3.2](#) is that all spiky matrices with a fixed blocky pattern can be seen as evaluations of a polynomial.

We will need the following standard probabilistic notions: a distribution μ over \mathbb{R}^m is *absolutely continuous* if there exists an integrable density function $f: \mathbb{R}^m \rightarrow \mathbb{R}^+$ such that for every Lebesgue-measurable set $A \subseteq \mathbb{R}^m$ we have $\mu(A) = \int_A f d\lambda$ where λ is the Lebesgue measure. Examples of absolutely continuous distributions include the uniform distribution over $[0, 1]^m$ and the Gaussian distribution.

Theorem 3.4. *Let μ be any absolutely continuous probability distribution over $\mathbb{R}_{N \times N}$. Then*

$$\Pr_{M \sim \mu} [\text{spr}(M) \geq N/2] = 1.$$

In particular this holds for $M \sim [0, 1]^{N \times N}$.

Proof. It suffices to show that the Lebesgue measure of the set of all matrices with spiky rank at most $N/2 - 1$ is zero. This immediately implies the existence of a matrix with spiky rank at least $N/2$.

Let $M \in \mathbb{R}^{N \times N}$ be a matrix with $\text{spr}(M) = r$. Then $M = \sum_{i \in [r]} S_i$, where each S_i is a spiky matrix. Each spiky matrix S_i can be viewed as an entrywise product of a blocky matrix B_i and a rank-1 matrix $a_i^T b_i$ for some vectors $a_i, b_i \in \mathbb{R}^n$. Fix a collection of blocky matrices $B_1, \dots, B_r \in \{0, 1\}^{N \times N}$, and define the corresponding set of spiky rank- r matrices with the blocky pattern B_1, \dots, B_r :

$$\mathcal{S}_{B_1, \dots, B_r} = \left\{ \sum_{i \in [r]} (a_i^T b_i) \circ B_i \mid a_1, \dots, a_r, b_1, \dots, b_r \in \mathbb{R}^N \right\},$$

where \circ is the Hadamard (entrywise) product. Then, the set of all matrices in $\mathbb{R}^{N \times N}$ of spiky rank at most r is equal to the union over all such choices of B_1, \dots, B_r : $\bigcup_{B_1, \dots, B_r} \mathcal{S}_{B_1, \dots, B_r}$. Since the

number of blocky matrices is finite (trivially it does not exceed 2^{N^2}), to show that the union set has measure zero it is sufficient to show that each set $\mathcal{S}_{B_1, \dots, B_r}$ has measure zero whenever $r \leq N/2 - 1$.

To show that $\mathcal{S}_{B_1, \dots, B_r}$ has Lebesgue measure zero, define a smooth function $f: \mathbb{R}^{2r \cdot n} \rightarrow \mathbb{R}^{n \times n}$ as follows: for $a_1, \dots, a_r, b_1, \dots, b_r \in \mathbb{R}^N$ and $i, j \in [N]$, let

$$f(a_1, b_1, \dots, a_r, b_r)[i, j] = \sum_{\ell \in [r]} 1_{B_\ell[i, j]=1} \cdot a_\ell[i] \cdot b_\ell[j].$$

The image of f is exactly $\mathcal{S}_{B_1, \dots, B_r}$.

Fact 3.5. *Suppose $f_1, \dots, f_n: \mathbb{R}^m \rightarrow \mathbb{R}$ are polynomials and $A \subseteq \mathbb{R}^m$ is a set with Lebesgue measure zero. Then $f(A) := \{(f_1(x), \dots, f_n(x)) \mid x \in A\}$ also has Lebesgue measure zero.*

Proof. Since polynomials are C^1 -differentiable (in fact, infinitely differentiable) we apply the more general standard result (see e.g. [GG12, Proposition 1.3]) that states that any C^1 -differentiable transformation maps a measure-zero set to a measure-zero set. \square

Since $\mathbb{R}^{2r \cdot N}$ has measure zero when $2rN < N^2$ and $\mathcal{S}_{B_1, \dots, B_r} = f(\mathbb{R}^{2r \cdot N})$, by Fact 3.5 $\mathcal{S}_{B_1, \dots, B_r}$ has measure zero for all $r < N/2$, and so does their finite union. \square

4 Connection to matrix rigidity

Definition 4.1 (Rigidity). For a matrix $M \in \mathbb{F}^{N \times N}$ over a field \mathbb{F} , the *rank- r rigidity* $\mathcal{R}_M^{\mathbb{F}}(r)$ is the minimum number of entries that must be modified to reduce the rank of M to r . Formally,

$$\mathcal{R}_M^{\mathbb{F}}(r) = \min\{s \mid M = A +_{\mathbb{F}} C, \text{rank}^{\mathbb{F}}(A) \leq r, \text{spar}(C) \leq s\}.$$

We show that matrices with large spiky rank must also be highly rigid. Our main result is the following generalization of Theorem 1.1 to arbitrary fields.

Theorem 4.2. *Let M be a matrix over a field \mathbb{F} and $0 < r \leq \text{spr}^{\mathbb{F}}(M)$. Then,*

$$\mathcal{R}_M^{\mathbb{F}}(r) \geq \frac{(\text{spr}^{\mathbb{F}}(M) - r)^2}{4}.$$

The key ingredient is that sparse matrices have small spiky rank.

Lemma 4.3. *Let $M \in \mathbb{F}^{N \times N}$ be a matrix over an arbitrary field \mathbb{F} with at most m non-zero entries. Then,*

$$\text{spr}^{\mathbb{F}}(M) \leq 2\sqrt{m}.$$

We defer the proof to Section 7.3. This immediately yields the following relationship between spiky rank and matrix rigidity.

Proof of Theorem 4.2. Assume $\mathcal{R}_M^{\mathbb{F}}(r) = s$ for some $r \leq N$. Then M can be written as $M = A +_{\mathbb{F}} C$, where $A \in \mathbb{F}^{N \times N}$ has $\text{rank}^{\mathbb{F}}(A) = r$ and $C \in \mathbb{F}^{N \times N}$ has at most s non-zero entries. By the subadditivity of spiky rank over \mathbb{F} , we have

$$\text{spr}^{\mathbb{F}}(M) = \text{spr}^{\mathbb{F}}(A +_{\mathbb{F}} C) \leq \text{spr}^{\mathbb{F}}(A) + \text{spr}^{\mathbb{F}}(C) \leq r + 2\sqrt{s},$$

where the last inequality follows from Lemma 4.3 and the trivial observation that $\text{spr}^{\mathbb{F}}(M) \leq \text{rank}^{\mathbb{F}}(M)$ for all M . Rearranging gives $s \geq (\text{spr}^{\mathbb{F}}(M) - r)^2/4$, which completes the proof. \square

Remark 4.4. The condition $0 < r \leq \text{spr}^{\mathbb{F}}(M)$ in [Theorem 4.2](#) limits its applicability for proving intermediate rigidity bounds. For instance, for Valiant’s target rank εN , the theorem applies only to matrices already known to have spiky rank $\Omega(N)$.

Remark 4.5. [Lemma 4.3](#), thus also [Theorem 4.2](#), is tight for real matrices: a random real matrix has spiky rank $\Omega(N)$ by [Theorem 3.4](#), and thus sparsity $\Omega(N^2)$ and rigidity $\Omega(N^2)$, the maximum possible. For Boolean matrices, we can slightly improve the sparsity and spiky rank tradeoff (see [Theorem 7.6](#)).

To discuss the implications of [Theorem 4.2](#) for communication complexity, we use the following result of Razborov [[Raz89](#)] and Wunderlich [[Wun12](#)].

Theorem 4.6 ([[Wun12](#), Theorem 5.15]). *Let \mathbb{F} be a finite field and let M be a $N \times N$ matrix. If*

$$\mathcal{R}_M^{\mathbb{F}}\left(2^{(\log \log N)^{\omega(1)}}\right) \geq \Omega\left(\frac{N^2}{2^{(\log \log N)^{O(1)}}}\right).$$

Then $M \notin PH^{cc}$.

As an immediate consequence we obtain:

Corollary 4.7. *Suppose a $N \times N$ Boolean matrix M satisfies*

$$\text{spr}^{\mathbb{F}}(M) \geq \frac{N}{2^{(\log \log N)^{O(1)}},$$

where \mathbb{F} is the finite field. Then $M \notin PH^{cc}$.

Proof. Applying [Theorem 4.2](#) with $r = \text{spr}^{\mathbb{F}}(M)/2$, we obtain

$$\mathcal{R}_M^{\mathbb{F}}\left(\frac{N}{2^{(\log \log N)^{O(1)}}}\right) \geq \Omega\left(\frac{N^2}{2^{(\log \log N)^{O(1)}}}\right).$$

Since

$$\mathcal{R}_M^{\mathbb{F}}\left(2^{(\log \log N)^{\omega(1)}}\right) \geq \mathcal{R}_M^{\mathbb{F}}\left(\frac{N}{2^{(\log \log N)^{O(1)}}}\right),$$

the claim follows directly from [Theorem 4.6](#). □

5 Connection to circuit complexity

Definition 5.1. A ReLU *gate* (rectified linear unit) is a function $\ell: \{0, 1\}^n \rightarrow \mathbb{R}^+$ such that there exist a vector $w \in \mathbb{R}^n$ and a scalar $a \in \mathbb{R}$ such that for all $x \in \{0, 1\}^n$

$$\ell(x) = \max\{0, \langle w, x \rangle + a\}.$$

Intuitively, a ReLU function is a *weighted* linear threshold function. For our purposes it will be convenient to view a ReLU gate as a matrix: a ReLU gate is a matrix $L: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \mathbb{R}^+$ such that there exist vectors $w_1, w_2 \in \mathbb{R}^n$ and a scalar $\alpha \in \mathbb{R}$ for which

$$L(x, y) = \max\{0, \langle w_1, x \rangle + \langle w_2, y \rangle + \alpha\}.$$

We now show that $\Sigma \circ \text{ReLU}$ circuit complexity – the number of ReLU gates – of any function is upper bounded by its spiky rank up to an $O(n)$ factor. As a first step, we bound the spiky rank of a single ReLU gate.

Claim 5.2. Let $L: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \mathbb{R}^+$ be a ReLU gate. Then

$$\text{spr}(L) \leq 3n + 3.$$

Proof. Let $w_1, w_2 \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ be such that $L(x, y) = \max\{0, \langle w_1, x \rangle + \langle w_2, y \rangle + \alpha\}$. Denote the underlying function $L'(x, y) = \langle w_1, x \rangle + \langle w_2, y \rangle + \alpha$. Observe that L' is a sum of three rank-1 matrices $\langle w_1, x \rangle$, $\langle w_2, y \rangle$ and α (each depending only on x , y or being constant), so $\text{rank}(L') \leq 3$.

Define T_L as the threshold function corresponding to L as follows:

$$T_L(x, y) = \begin{cases} 1, & \text{if } L'(x, y) > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Avraham and Yehudayoff [AY24] show that any threshold function on $\{0, 1\}^n \times \{0, 1\}^n$ can be written as a disjoint union of at most $n + 1$ blocky matrices:

$$T_L = B_1 + \dots + B_{n+1},$$

where B_1, \dots, B_{n+1} are disjoint blocky matrices, i.e., their supports are disjoint rectangles in the matrix. We now define weighted matrices R_1, \dots, R_{n+1} corresponding to B_i 's by placing the original values of L on the support of each B_i :

$$R_i(x, y) = \begin{cases} L(x, y), & \text{if } B_i(x, y) = 1 \\ 0, & \text{otherwise.} \end{cases}$$

Now since B_i 's are disjoint, we have $L = R_1 + \dots + R_{n+1}$. The only problem now is that R_i might not be a spiky matrix as it might have blocks of rank greater than 1. However, the blocks of R_i are submatrices of L' , so each block of R_i has rank at most 3. If there is a block R_i of rank greater than 1, then we can decompose R_i into at most three matrices that have rank-1 on that block. If multiple blocks in the same R_i have rank greater than 1, then we can still partition R_i into at most 3 matrices that have rank-1 on all blocks, because the blocks are on disjoint rows and columns. This shows that each R_i can be decomposed into at most three spiky matrices. Hence, $\text{spr}(M) \leq 3(n + 1)$. \square

Proposition 5.3. For $M: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \mathbb{R}$, let s be the size of a $\Sigma \circ \text{ReLU}$ circuit that computes M . Then

$$\frac{\text{spr}(M)}{3(n + 1)} \leq s.$$

Proof. For the lower bound, let $M = \sum_{i=1}^s L_i$, where L_i are ReLU functions. Then,

$$\text{spr}(M) \leq \sum_{i=1}^s \text{spr}(L_i) \leq s \cdot (3n + 3). \quad \square$$

6 Explicit lower bounds for spiky rank

6.1 Lower bound framework

In this section we give a framework for proving spiky ranks lower bounds that applies to thin matrices such as HD_1^n and adjacency matrices of expander graphs.

Theorem 6.1. Let $M \in \mathbb{R}^{N \times N}$, s, k, D be positive integers and $\gamma \in (0, 1)$. If

(P1) For every $S \subseteq [N]$ and $T \subseteq [N]$ with $|S| \leq s$ and $|T| \leq s$ we have $\text{spar}(M|_{S \times T}) \leq k \cdot (|S| + |T|)$.

(P2) For every $M' \leq M$ (entrywise) such that $\text{spar}(M - M') \leq \gamma \cdot \text{spar}(M)$, there exist sets $S \subseteq [N]$ and $T \subseteq [N]$ such that $|S| = |T| = D$ and $M'|_{S \times T}$ is a permutation matrix.

then

$$\text{spr}(M) \geq \Omega \left(\min \left(\frac{\gamma \cdot \text{spar}(M)}{kN}, \log \left(\frac{Ds}{2N} \right) \right) \right).$$

We need the following lemma for the proof. The proof of it is identical to the proof of [AY24, Lemma 12], but we include it here for completeness.

Lemma 6.2. Let I_N be identity matrix of size $N \times N$. If there is a blocky (or spiky) matrix B , such that $B \wedge I_N = 0$ (i.e. there is no entry where both I_N and B are nonzero). Then there exists $V \subset [N]$, s.t. $|V| \geq N/4$ and $B|_{V \times V} = 0$.

Proof. Let blocks of B be $A_1 \times B_1, A_2 \times B_2, \dots, A_k \times B_k$. Define two random subsets of $[N]$ as follows. Let $\xi_1, \xi_2, \dots, \xi_k$ be i.i.d. uniformly distributed on $\{0, 1\}$.

Let S be the complement of $\bigcup_{a:\xi_a=0} S_a$, and T be the complement of $\bigcup_{a:\xi_a=1} T_a$. Let $I = \{i \in [N] \mid (i, i) \in S \times T\}$. Note that $B|_{S \times T}$ is zero matrix. For each $i \in [N]$, the probability that $(i, i) \in S \times T$ is $1/4$, because $B_{ii} = 0$. Thus there is a choice for $S \times T$ so that $|I| \geq N/4$. For such S and T , we define $V = S \cap T$. \square

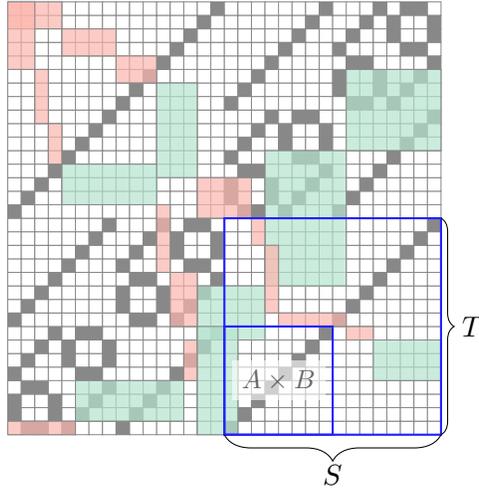


Figure 4: The picture illustrates the proof for the matrix HD_1^5 . The parts T_i are colored red, H_i are colored green. By (P1) not too many 1-cells (black ones) are covered by the small blocks, so by (P2) we find a permutation submatrix $S \times T$ with very few covered 1-cells, and then further shrink it using Lemma 6.2 to $A \times B$ that does not contain any red cells at all. Then the rank of $\text{HD}_1^5|_{A \times B}$ is maximal, but the total rank of large blocks is small, which is a contradiction.

Proof of Theorem 6.1. Let $M = \sum_{i \in [t]} S_i$. We call a block small, if the sum of its dimensions is at most s , and big otherwise. Assume the contrary and let $t \leq \frac{\gamma \cdot \text{spar}(M)}{2kN}$ and $t \leq \frac{1}{4} \log_2 \left(\frac{Ds}{2N} \right)$.

We represent $S_i = T_i + H_i$ where T_i is comprised of the small blocks of the matrix T_i and H_i of the big ones (i.e. the sets of non-zero columns of T_i and H_i are disjoint, same for the non-zero rows).

By (P1), each T_i covers (blocky matrix covers matrix entry, if this entry is non-zero in this matrix) at most $2kN$ 1-entries of M , because the sum of block sizes in a spiky matrix does not exceed $2N$. This implies that all small matrices cover at most $2Ntk \leq \gamma \cdot \text{spar}(M)$ ones.

By (P2) we get a permutation submatrix $S \times T$ and applying Lemma 6.2 to it t times, we obtain a permutation submatrix $A \times B$ of M with $|A| = |B| = D/2^{2t}$, s.t. for each i the intersection of T_i and $A \times B$ is empty. Thus, we obtain

$$\text{rk}\left(M - \sum_{i \in [t]} T_i\right) \geq \frac{D}{2^{2t}} \quad (1)$$

On the other hand, each H_i contains at most $2N/s$ blocks, that means

$$\text{rk}\left(\sum_{i \in [t]} H_i\right) \leq \sum_{i \in [t]} \text{rk}(H_i) \leq t \cdot \frac{2N}{s}. \quad (2)$$

Combining (1) and (2), we get $D \cdot 2^{-2t} \leq 2Nt/s$, which implies $\frac{Ds}{2N} \leq 2^{3t}$, which contradicts our assumption. \square

6.2 Hamming-Distance-1

HD_1^n is the function which determines whether the Hamming distance between the two input strings is 1. Its matrix is the adjacency matrix of the n -dimensional Hamming cube.

$$\text{HD}_1^n(x, y) = \begin{cases} 1 & \text{if } \|x \oplus y\|_1 = 1 \\ 0 & \text{otherwise} \end{cases}$$

Theorem 1.8. $\text{spr}(\text{HD}_1^n) \geq \Omega(\sqrt{\log N})$.

We will make use of the following lemma.

Lemma 6.3 (Lemma 3.10 in [KSP20]). *If $G = (V, E)$ is a subgraph of a Hamming cube, then*

$$|V| \geq 2^{|E|/|V|}.$$

Proof of Theorem 1.8. We use Theorem 6.1 with $N = 2^n$, $M = \text{HD}_1^n$, $s = 2^{\sqrt{n}}$, $k = 2\sqrt{N}$, $\gamma = 0.5$, $D = N/4$. We now prove two conditions of the theorem. (P1) directly from Lemma 6.3 applied to $S \cup T$.

We proceed to prove (P2). The 1-entries of HD_1^n are partitioned into the following $2n$ subsets:

$$K_{2i-j} = \{(x, y) \mid \text{HD}_1^n(x, y) = 1, x_i = j, y_i = 1 - j\} \text{ for } i \in [n]; j \in \{0, 1\}.$$

Thus, each 1-entry in HD_1^n belongs to exactly one of these subsets, and each subset has size 2^{n-1} . The subsets have the following property: if (x, y) and (v, w) are distinct elements of K_t , then $\|x - w\|_1 > 1$ and $\|v - y\|_1 > 1$, because we have $x_i = v_i \neq y_i = w_i$ for some i . Thus, we have at least half of the 1-entries present in M' , which means one of the groups has at least $\frac{1}{2}2^{n-1}$ elements in the intersection with M' . Without loss of generality, we assume that it is the subset K_1 . Then we set

$$S \subseteq \{1\} \times \{0, 1\}^{n-1}, \quad T \subseteq \{0\} \times \{0, 1\}^{n-1},$$

such that $S \times T$ covers exactly 2^{n-2} 1-entries of K_1 , which are present in M' , and they form a permutation submatrix. So, the desired $\Omega(\sqrt{n})$ bound is immediately implied by the theorem. \square

6.3 Expanders

For a graph G , we denote by $M_G \in \{0, 1\}^{N \times N}$ its adjacency matrix, i.e.,

$$(M_G)_{u,v} = \begin{cases} 1 & \text{if } (u, v) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

We recall that an (N, d, λ) -spectral expander is a d -regular graph $G = (V, E)$ on N vertices such that the second largest eigenvalue of its adjacency matrix M_G is at most λ in absolute value (see, for example, [HLW06]).

Theorem 1.10. $\text{spr}(M_G) \geq \Omega(\min(\frac{d}{\lambda}, \log \frac{N}{d^2}))$.

Proof. We use [Theorem 6.1](#) with $M = M_G$, $s = \frac{N\lambda}{10d}$, $k = \lambda$, $D = \frac{N}{4d}$, $\gamma = 1/2$. We check the theorem conditions. (P1) holds by the expander mixing lemma

$$e(S, T) \leq \frac{d|S||T|}{N} + \lambda\sqrt{|S||T|} \leq \frac{ds}{N}\sqrt{|S||T|} + \lambda\sqrt{|S||T|} \leq 2\lambda\sqrt{|S||T|} \leq \lambda(|S| + |T|).$$

Let us prove (P2). We pick edges greedily. At each step, when edge $e = \{u, v\}$ is picked, we throw away all vertices that are connected either to u or v . At each step, at most $2d - 1$ vertices are used, so there will be at least $\frac{\gamma N}{2d} \geq \frac{N}{4d}$ steps and we get at least $D \times D$ permutation submatrix.

Applying the theorem, we get the desired bound. \square

As an implication, we get an $\Omega(\log N)$ lower bound for $(N, \Theta(\log^2 N), \Theta(\log N))$ -expanders:

Corollary 6.4. *There exists an explicit graph G with N nodes and degree $\Theta(\log^2 N)$ such that*

$$\text{spr}(M_G) = \Omega(\log N).$$

Proof. It is shown in [Alo21, Section 2.3] that for every prime p with $p \bmod 4 = 1$ there exist two primes $q_1, q_2 = p^{\Theta(1)}$ such that for every $\ell \geq 1$ there exists a degree $p + 1$ Ramanujan graph G_ℓ with $\Theta((q_1 q_2)^{3\ell})$ nodes, i.e. G_ℓ is a $(\Theta((q_1 q_2)^{3\ell}), p + 1, 2\sqrt{p})$ -spectral expander. Then the number of nodes in G_ℓ is $N_\ell = p^{\Theta(\ell)}$, so taking $\ell = \sqrt{p}/\log p$ we get that $N_\ell = 2^{\Theta(\sqrt{p})}$, so $p = \Theta(\log^2 N_\ell)$. Finally, we apply [Theorem 1.10](#) to get the desired result. \square

We get a stronger corollary for *ultra-lossless vertex expanders*. These are known to exist, but there are currently no explicit constructions.

Theorem 6.5. *Let G be a d -regular graph with N vertices and vertex expansion $d - O(1)$ for at most N^δ vertex subsets, where $\delta \in (0, 1)$ (i.e. for any $S \subseteq V(G)$, it has at least $(d - O(1))|S|$ neighbors). Let M_G be its adjacency matrix. Then $\text{spr}(M_G) \geq \Omega(\min(d, \log N))$.*

Proof. The proof is similar to the proof of [Theorem 1.10](#), except for the proof of (P1). It is proved in the following way.

Let $S \subseteq V$ with $|S| \leq N^\delta$. By the assumption of the claim, it has at least $(d - O(1))|S|$ neighbors. On the other hand the sum of degrees of vertices in $|S|$ is $d|S|$. Thus, for any $T \subseteq V(G)$ with $|T| = |S|$ we have $e(S, T) \leq d|S| - (d - O(1))|S| + |T| = O(|S|)$. \square

6.4 Inner Product and Disjointness

A simple consequence of the lower bound for random Boolean matrices ([Theorem 3.3](#)) is the following. Suppose an explicit $N \times N$ matrix M contains a $k \times 2^k$ submatrix with all columns distinct. Then $\text{spr}_{\pm}(M) \geq \Omega\left(\frac{k}{\log k}\right)$. Indeed, such a submatrix necessarily contains every possible $k \times k$ Boolean matrix (up to column permutations, which do not affect spr_{\pm}), and hence in particular one with maximum sign spiky rank $\Omega\left(\frac{k}{\log k}\right)$.

The largest k for which this holds is exactly the Vapnik–Chervonenkis dimension $\text{VC}(M)$ of the matrix M [[VC71](#)]. Thus we obtain the general bound

$$\text{spr}_{\pm}(M) \geq \Omega\left(\frac{\text{VC}(M)}{\log \text{VC}(M)}\right).$$

This observation immediately implies sign spiky rank lower bounds for Inner Product and Disjointness matrices. Recall that $\text{IP}_n(x, y) = \sum_{i \in [n]} x_i y_i \bmod 2$, where $x, y \in \{0, 1\}^n$, and $\text{DISJ}_n \in \{0, 1\}^{N \times N}$ is defined with $\text{DISJ}_n(x, y) := \bigwedge_{i \in [n]} \neg x_i \vee \neg y_i$ where x and y are identified with their binary representations.

Proposition 6.6. $\text{spr}_{\pm}(\text{IP}_n) \geq \Omega\left(\frac{n}{\log n}\right)$ and $\text{spr}_{\pm}(\text{DISJ}_n) \geq \Omega\left(\frac{n}{\log n}\right)$.

Proof. The VC-dimension of IP_n and DISJ_n is at least n . Consider the $n \times 2^n$ submatrix of DISJ_n (or IP_n) consisting of the n rows indexed by the standard basis vectors

$$e_i := (0^{i-1}, 1, 0^{n-i}), \quad i \in [n].$$

For a column indexed by $x \in \{0, 1\}^n$, its entries in this submatrix are as follows:

- In the inner product matrix IP_n , the column is exactly the vector x .
- In the Disjointness matrix DISJ_n , the column is $1^n - x$, where 1^n is the all-ones vector.

Thus the submatrix captures all distinct column patterns, corresponding either to x (for IP_n) or its complement (for DISJ_n). □

7 Relations to other measures

7.1 Blocky Cover

We start with a very simple connection that we will subsequently use as a lemma. A *blocky cover* $\text{bc}(M)$ of a Boolean matrix M is the minimum number of blocky matrices that cover all 1-entries of M .

Lemma 7.1. For every $M \in \{0, 1\}^{n \times n}$ we have $\text{spr}(M) \leq \text{br}(M) \leq 2^{\text{bc}(M)}$.

Proof. Consider the blocky cover representation of $M = \bigvee_{i \in [\text{bc}(M)]} A_i$. For every set

$$S = \{s_1, \dots, s_{\ell}\} \subseteq [\text{bc}(M)]$$

define $A_S := A_{s_1} \circ \dots \circ A_{s_{\ell}}$ to be the entrywise product of A_i for $i \in S$. Then by the standard inclusion-exclusion principle we have $M = \sum_{S \subseteq [\text{bc}(M)]} (-1)^{|S|} A_S$. Since entrywise-product of blocky matrices is a blocky matrix, this implies the claimed bound on the blocky rank. □

7.2 Blocky rank

In this section, we prove the following theorem.

Theorem 1.13. *For any Boolean matrix M , $\text{br}(M) \leq \text{spr}(M)^{O(\text{spr}(M))}$.*

We call a subset $S \subseteq [N] \times [M]$ of matrix entries *t-blocky complement* if $\text{bc}(\mathbb{1}_{[N] \times [M] \setminus S}) \leq t$.

The proof proceeds as follows. We first partition the matrix entries into $2^{\text{spr}(M)}$ classes, according to the subset of spiky matrices in the representation that influence each entry. Each class can be written as a union of subsets of $[N] \times [M]$ with disjoint rows and columns, where each such subset is a $\text{spr}(M)$ -blocky complement. Finally, observing that each of these subsets can be completed to a real matrix of $\text{spr}(M)$ -rank, we obtain the desired result.

Proof of Theorem 1.13. Let $M = R_1 + R_2 + \dots + R_k$ be the spiky representation of M with $k = \text{spr}(M)$. For each entry (j, ℓ) of M , we assign a vector $v_{j\ell} \in \{0, 1\}^k$ as follows: $(v_{j\ell})_i = 1$ if $(R_i)_{j\ell} \neq 0$. So, all the entries are divided into 2^k types based on the corresponding vector.

We denote $M \uparrow v$ a matrix obtained from M by zeroing out all entries where v is not assigned, so $M = \sum_{v \in \{0,1\}^k} M \uparrow v$. We will prove that $\text{br}(M \uparrow v) = k^{O(k)}$ for $v \in \{0, 1\}^k$. It implies that $\text{br}(M) = 2^k \cdot k^{O(k)} = k^{O(k)}$ by the subadditivity of blocky rank.

Fix some $v \in \{0, 1\}^k$. Let $P \subseteq [k]$ be the set of 1-entries in v , and $Z \subseteq [k]$ be the set of 0-entries. We first observe that the support of $M \uparrow v$ is contained in the intersection of supports of R_i for $i \in P$. Let M' be the result of zeroing-out elements of M outside of this intersection. In order to obtain $M \uparrow v$ from M' one needs to additionally zero-out the entries covered by the union of supports of R_i for $i \in Z$, see Figure 5 for an example. Notice that the set of non-zero elements in M' forms a blocky matrix.

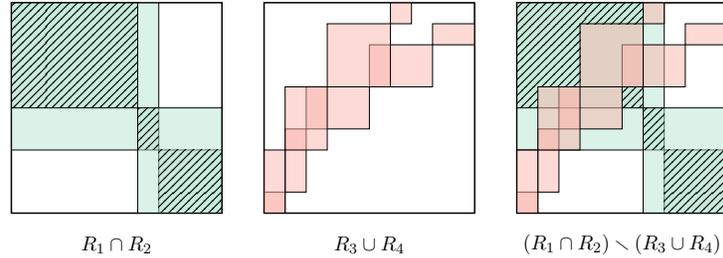


Figure 5: The picture illustrates the process for $v = (1100)$, so $P = \{1, 2\}$ and $Z = \{3, 4\}$. Here the green matrices illustrate R_1 and R_2 , the red matrices are R_3 and R_4 , the hatched area in the right square is the support of the matrix $M \uparrow v$.

Let D be a block in M' . It suffices to upper bound $\text{br}((M \uparrow v)|_D)$, since $\text{br}(M \uparrow v)$ is the maximum of the blocky ranks in the blocks. Let $L = \sum_{i \in P} R_i|_D$, S be the set of entries of D that are non-zero in $M \uparrow v$, by definition S is a k -blocky complement. Then L agrees with $M|_D$ on S . This means $L|_S$ is Boolean. The rank of L can be upper bounded as:

$$\text{rk}(L) = \text{rk}\left(\sum_{i \in P} R_i|_D\right) \leq \sum_{i \in P} \text{rk}(R_i|_D) \leq k$$

where $\text{rk}(R_i|_D) \leq 1$ for each $i \in P$ since D is a subset of a block in R_i . To proceed further, we need the following lemma:

Lemma 7.2. *Let M be a real-valued matrix with $\text{rk}(M) \leq r$ and a subset S of its entries be a t -blocky complement, such that $M|_S$ is Boolean. Then there exist m blocky matrices B_1, \dots, B_m , and coefficients $\alpha_1, \dots, \alpha_m \in \{\pm 1\}$, such that $m = (t+r)^{O(t+r)}$, and $\sum_{i=1}^m \alpha_i B_i$ agrees with M on S : for all $(j, \ell) \in S$*

$$\left(\sum_{i=1}^m \alpha_i B_i \right)_{j\ell} = M_{j\ell}.$$

Now, Lemma 7.2 can be applied to the matrix L , and we get $\sum_{i=1}^m \alpha_i B_i$ agrees with L on S . Observe that this sum is not necessarily Boolean: we only control its entries in S . Now we will get blocky representation of $(M \uparrow v)|_D$ based on $\sum_{i=1}^m \alpha_i B_i$: we need to zero-out the entries outside of S .

By definition $(M \uparrow v)|_D = \sum_{i=1}^m (B_i \circ \mathbb{1}_{\bar{S}})$. By Lemma 7.1 we have $\text{br}(\mathbb{1}_{\bar{S}}) \leq 2^k$ since S is a k -blocky complement. Then since entry-wise product of two blocky matrices is a blocky matrix we get $\text{br}(B_i \circ \mathbb{1}_{\bar{S}}) \leq 2^k$ for each i and by subadditivity of blocky rank we get $\text{br}((M \uparrow v)|_D) \leq m \cdot 2^k \leq k^{O(k)}$ and by combining these across all blocks D in M' we get the same bound for $\text{br}(M \uparrow v)$. \square

Now it remains to prove the lemma:

Proof of Lemma 7.2. For simplicity, we assume $\text{rk}(M) = r$. The case of a smaller $\text{rk}(M)$ follows immediately. We prove that m can be bounded by T_t , where $T_t = rt \cdot T_{t-1} + 2^r$ with $T_0 = 2^r$. We prove it by induction on t .

Base case: $t = 0$. In this case, M is a Boolean matrix of rank at most r . Let c_1, c_2, \dots, c_r be the indices of columns forming the basis. Thus, every row is fully determined by its entries at positions c_1, c_2, \dots, c_r . So, there can be at most 2^r different rows in M and thus $\text{br}(M) \leq 2^r$, as we can represent each type of rows with one rectangle.

Induction step: $t > 0$. Let L_1, \dots, L_t be the blocky matrices, whose union is \bar{S} , and N_{i1}, \dots, N_{it_i} be blocks of L_i for each $i \in [t]$. We denote with R the set of rows in M , and with C the set of columns in M .

Let $c_1, c_2, \dots, c_r \in C$ be the indices of columns forming the basis. We call a row *spoiled* if there exists ℓ , such that its c_ℓ -th entry belongs to N_{ij} for some i, j . We assign each of the spoiled rows to the corresponding class R_{ij} . If one row corresponds to several classes, we choose one of them arbitrarily. Note that there are at most rt non-empty classes, because each column intersects at most one of the blocks $N_{i1}, N_{i2}, \dots, N_{it_i}$ for every i . Now we will solve the problem separately for each (non-empty) R_{ij} and for $R \setminus \bigcup_{i,j} R_{ij}$.

First, we use the induction hypothesis for $M|_{R_{ij} \times (C \setminus C(N_{ij}))}$ and $S \cap (R_{ij} \times (C \setminus C(N_{ij})))$, where $C(N_{ij})$ is set of columns forming the block N_{ij} . Note that $S \cap (R_{ij} \times (C \setminus C(N_{ij})))$ is $(t-1)$ -blocky complement wrt $M|_{R_{ij} \times (C \setminus C(N_{ij}))}$, because L_i is excluded from the complement. Thus, we get exactly the blocky decomposition which agrees with $M|_{R_{ij} \times C}$ on $S \cap (R_{ij} \times C)$.

Second, we use the same argument as in the induction base to prove that $\text{br}(M|_{(R \setminus \bigcup_{i,j} R_{ij}) \times C}) \leq 2^r$.

Overall, we get blocky decomposition, which agrees with M on S with at most $T_t = rt \cdot T_{t-1} + 2^r$ matrices.

It is easy to see using induction that

$$T_t = r^t t! \cdot T_0 + 2^r \sum_{i=0}^{t-1} r^i \prod_{j=0}^{i-1} (t-j).$$

Thus, we can upper bound T_t with $T_t \leq 2^{O(t \log r + t \log t + r)} + t \cdot 2^{O(r + t \log r + t \log t)} \leq (r+t)^{O(r+t)}$. \square

We observe that without the assumption on the structure of S , the lemma is completely false.

Proposition 7.3. *For every $n \in \mathbb{Z}_{>0}$ there exists N and a matrix $M \in \mathbb{R}^{N \times N}$ of rank 2, such that deterministic communication complexity of computing $M(x, y)$ with the promise that $M(x, y) \in \{0, 1\}$ is at least n .*

Proof. Consider the following communication problem \mathcal{M} that embeds all rank-2 matrices: Alice's input is $(x_1, x_2) \in \mathbb{R}^2$, Bob's input is $(y_1, y_2) \in \mathbb{R}^2$. They need to return $x_1x_2 + y_1y_2$ whenever this value is in $\{0, 1\}$. We will show that the communication cost of this problem is $\omega(1)$ which implies that for every n there is a finite $N \times N$ submatrix of \mathcal{M} that satisfies the claim.

If there is a cost- k protocol for a communication problem, then there is a cost- 2^{k+1} one-sided protocol for that problem, since Alice can just send her answers for all potential replies of Bob. Hence, if there exists an $O(1)$ -cost protocol for \mathcal{M} , then there is a partition $A_1 \sqcup A_2 \sqcup \dots \sqcup A_k = \mathbb{R}^2$ such that for each A_i and $(x, y) \in \mathbb{R}^2$ there exists $b \in \{0, 1\}$ such that $\langle z, (x, y) \rangle \neq b$ for every $z \in A_i$ (in that case, Alice can send i and Bob can determine the answer by his input (x, y)).

So let $(\mathbb{R}^2, \{(u, v) \in \binom{\mathbb{R}^2}{2} \mid \exists t \in \mathbb{R}^2: \{\langle u, t \rangle, \langle v, t \rangle\} = \{0, 1\}\})$ be a graph, the partition above is exactly k -coloring of this graph. We now claim that every set of points $u_1, \dots, u_n \in \mathbb{R}^2 \setminus \{0\}$ such that $\{0, u_1, \dots, u_n\}$ are in a general position, forms a clique in this graph contradicting that it is $O(1)$ -colorable. Indeed, for every pair of points u_i, u_j the system $\langle u_i, t \rangle = 1$ and $\langle u_j, t \rangle = 0$ has a solution: take $t' = (-u_j)_1, (u_j)_0$, so the second equation is satisfied by any $\alpha t'$ and then find appropriate α to satisfy the first equation, it fails only in the case if $\{0, u_i, u_j\}$ are on a line, so since $\{0, u_1, \dots, u_n\}$ are in the general position, $\{u_1, \dots, u_n\}$ is indeed a clique. \square

7.3 Sparsity

Lemma 7.4 (Lemma 4.3 restated). *Let $M \in \mathbb{F}^{N \times N}$ be a matrix over an arbitrary field \mathbb{F} ¹ and let $\text{spar}(M) \leq m$. Then*

$$\text{spr}^{\mathbb{F}}(M) \leq 2\sqrt{m}.$$

Proof. Let c_1, \dots, c_N denote the number of non-zero entries in each column of M , so $\sum_{i=1}^N c_i = m$. We show a process to decompose M into spiky rank-1 matrices.

Note that any matrix with a single non-zero column has spiky rank 1. Also, a matrix with at most one non-zero entry in each column has spiky rank 1.

We now describe a strategy that decomposes M into matrices of these two types. Consider the following game on the column counts c_1, \dots, c_N :

Type 1 move: nullify a single c_i (i.e., remove a matrix with one non-zero column).

Type 2 move: subtract 1 from each c_i (i.e., remove a matrix with one non-zero entry per column).

Each such move corresponds to adding a spiky rank-1 matrix to our decomposition, contributing 1 to the spiky rank. Our goal is to make all c_i 's zero using as few moves as possible. Here is a simple strategy that makes only $2\sqrt{m}$ moves:

1. While there exists i such that $c_i > \sqrt{m}$, perform a type 1 move on that column. Since the total sum of c_i 's is at most m , there can be at most \sqrt{m} such i . Thus, this phase requires at most \sqrt{m} moves.
2. Now $c_i \leq \sqrt{m}$ for all i . Perform type 2 moves repeatedly. After at most \sqrt{m} such steps, all c_i become zero.

¹we think of \mathbb{F} either \mathbb{R} or a finite field \mathbb{F}_p .

Altogether, we use at most $2\sqrt{m}$ moves, each corresponding to a spiky rank-1 matrix. Therefore, $\text{spr}^{\mathbb{F}}(M) \leq 2\sqrt{m}$. \square

We will further improve [Lemma 4.3](#) for Boolean matrices using the recipe from [[AY22](#), Theorem 2] (preprint version of [[AY24](#)]). We will need the following simple technical lemma:

Lemma 7.5 (Improved Lemma 11 from [[AY22](#)]). *For large enough N , every $M \in \{0, 1\}^{N \times N}$ with $\text{spar}(M) \geq \frac{N^2}{\log^6 N}$ contains $\frac{\log N}{100 \log \log N} \times \sqrt{N}$ all-1 rectangle.*

Proof. The reference we use is Lemma 5.4 in [[RY20](#)], which says that if $\varepsilon > 0$ and k are such that

$$\varepsilon N \geq 2k \quad \text{and} \quad k \leq \frac{\log N}{2 \log(2e/\varepsilon)},$$

then every matrix M with $\text{spar}(M) \geq \varepsilon N^2$ contains a 1-submatrix with at least k rows and \sqrt{n} columns. Applying this lemma with $\varepsilon = (1000 \log^6 N)^{-1}$ completes the proof. \square

Theorem 7.6. *Any Boolean matrix M with sparsity m satisfies*

$$\text{br}(M) \leq O\left(\frac{\sqrt{m} \log \log m}{\log m}\right).$$

Proof. We are going to eliminate all one-entries of M by subtracting blocky matrices from M and maintain the property that after the subtraction the resulting matrix is still Boolean. Thus, the blocky decomposition we build is a *blocky partition*. At each step, we assume that each row and each column of M contains a nonzero element. Otherwise, the empty rows and columns can be removed from the matrix with no change to the blocky rank.

First, we adopt the same strategy as in [Lemma 4.3](#) for the matrices M , s.t. its number of rows or number of columns is at least $\sqrt{m} \log m$. We can do at most $\frac{m}{\sqrt{m} \log m}$ such steps, using one blocky matrix at each step, thus using at most $\frac{\sqrt{m}}{\log m}$ blocky matrices.

Let $r(A)$ be the number of (non-empty) rows in A , and $c(A)$ be the number of (non-empty) columns in A . So, by the above, we may assume that $r(M) \leq \sqrt{m} \log m$ and $c(M) \leq \sqrt{m} \log m$. We now repeat the proof of [[AY22](#), Theorem 2], ensuring that the same strategy works in our case.

We construct a nested matrix sequence

$$M = M_0 \supseteq M_1 \supseteq M_2 \supseteq \dots \supseteq M_N,$$

with controlled $\text{br}(M_i - M_{i+1})$ and $\text{br}(M_N) \leq \frac{\sqrt{m}}{\log m}$.

Assume we have constructed M_0, M_1, \dots, M_t . If we reach a situation where $c(M_t) \leq \sqrt{m}$ and $r(M_t) \leq \sqrt{m}$, we set $N = t$ and conclude by [[AY24](#), Theorem 1.4]. Otherwise, we use the following algorithm.

Let

$$k = \frac{\log m}{100 \log \log m} \quad \text{and} \quad \ell = \sqrt[4]{m}.$$

Without loss of generality, suppose $c(M_t) \geq \sqrt{m}$. We construct the matrix M_{t+1} with $c(M_{t+1}) \leq c(M_t)/2$. Assume we have already constructed B_0, B_1, \dots, B_q and want to construct B_{q+1} , with $B_0 = 0$. Let

$$M' = M_t - \sum_{i=0}^q B_i.$$

Let us run the following procedure:

Algorithm 1 Extracting a Sequence of Submatrices

- 1: $j \leftarrow 0$
- 2: $N_0 \leftarrow M'$
- 3: **while** $\text{spar}(N_j) \geq m/\log^4 m$ **do** \triangleright by Lemma 7.5 with $n = \sqrt{m} \log m$
- 4: Decompose N_j as

$$N_j = \begin{pmatrix} 1_{k \times \ell} & * \\ * & N_{j+1} \end{pmatrix}$$

- 5: $j \leftarrow j + 1$
-

Let T denote the final value of t . Let B be the blocky matrix with the T one-blocks of size $k \times \ell$ that were found above. It could sometimes be the zero matrix. Define L with

$$L := \begin{pmatrix} 0 & 0 \\ 0 & N_T \end{pmatrix}.$$

The matrices B and L have the same size as M_t .

There are two options to consider.

The first option is that $T\ell \geq \frac{c(M_t)}{2}$. In this case, we set $B_{q+1} = B$ and continue the procedure. There are at least $T\ell$ nonzero columns in B , and each column has k nonzero entries. Hence

$$|B_{q+1}| \geq \frac{c(M_t)k}{2} \geq \frac{\sqrt{mk}}{2}.$$

The second option is that $T\ell < \frac{c(M_t)}{2}$. Denote by B' the matrix obtained from M' by zeroing out every zero row in B . Because the total number of nonzero rows in B is at most $\frac{k \cdot c(M_t)}{\ell}$ we conclude

$$\text{br}(B') \leq \frac{k \cdot c(M_t)}{\ell} \leq k\sqrt[4]{m} \log m.$$

Moreover, because $\text{spar}(L) < \frac{m}{\log^4 m}$, Lemma 4.3 implies $\text{br}(L) \leq \sqrt{m}/\log^2 m$. Then the matrix $M' - L - B'$ satisfies $c(M' - L - B') \leq T\ell < c(M_t)/2$. We have thus reduced the number of columns by a factor of two and we set $M_{t+1} = M' - L - B'$. So, we have

$$\text{br}(M_t - M_{t+1}) \leq \text{br}(L) + \text{br}(B) + q \leq \frac{\sqrt{m}}{\log^2 m} + k\sqrt[4]{m} \log m + \frac{2 \cdot \text{spar}(M_t - M_{t+1})}{\sqrt{mk}}.$$

Each time either $c(M_t)$ or $r(M_t)$ is reduced 2 times, so we will do at most $2 \log \log m$ steps to reach $c(M_t) \leq \sqrt{m}$ and $r(M_t) \leq \sqrt{m}$. Thus, we get

$$\begin{aligned} \text{br}(M) &\leq \log \log m \cdot \left(\frac{\sqrt{m}}{\log^2 m} + k\sqrt[4]{m} \log m \right) + \frac{2 \cdot \text{spar}(M - M_N)}{\sqrt{mk}} \\ &\leq o\left(\frac{\sqrt{m}}{\log m}\right) + \frac{2 \cdot \text{spar}(M)}{\sqrt{mk}} \\ &\leq O\left(\frac{\sqrt{m} \log \log m}{\log m}\right). \end{aligned} \quad \square$$

7.4 γ_2 -norm

In this section, we give a non-explicit quadratic separation between sign spiky rank and γ_2 -norm.

Theorem 7.7 (Theorem 1.14 restated.). *There exists a Boolean matrix M such that*

$$\text{spr}(M) \geq \Omega(\gamma_2^2(M)).$$

Lemma 7.8. *There exists an N -vertex bipartite graph G of left degree at most $d \leq N^{1/4}$ such that its adjacency matrix has $\text{spr}_\pm(M_G) \geq d/18$.*

Proof. The number of left-regular graphs with left degree d is exactly $\binom{N}{d}^N$. On the other hand, the number of boolean matrices with sign spiky rank r is at most N^{6Nr} by Lemma 3.2. It is then sufficient to have

$$\binom{N}{d}^N > N^{6Nr} \tag{3}$$

to conclude that there exists a left-degree- d graph G with adjacency matrix M_G such that

$$\text{spr}_\pm(M_G) > r.$$

Choose $r = d/18$, then $N^{6Nr} = N^{Nd/3}$. On the other hand $\binom{N}{d}^N \geq N^{Nd}/(2d)^{Nd}$, so by rearranging (3) is equivalent to $N^{Nd-Nd/3} > (2d)^{Nd}$, which is satisfied if $d \leq N^{1/4}$. \square

Proof of Theorem 7.7. Since ℓ_2 -norm of the rows of adjacency matrix of a left- d -regular graph is \sqrt{d} we get that for an adjacency matrix M of such a graph we have $\gamma_2(M) \leq \sqrt{d}$. Hence, for a graph given by Lemma 7.8 we have that $\text{br}(M) \geq \text{spr}(M) \geq \text{spr}_\pm(M) \geq \Omega(\gamma_2^2(M))$. \square

Can this separation be made explicit? A variation on Theorem 1.10 may offer a way to achieve it: it says that for an *ultra-lossless* degree- d expander G we have $\text{spr}(M_G) = \Omega(d)$. Since $\gamma_2(M_G) \leq \|M_G\|_{\text{row}} \leq \sqrt{d}$ for every d -regular graph G , it is another way to prove Theorem 1.14 non-explicitly (albeit only for spr and not for spr_\pm). The expanders required in Theorem 6.5 are known to exist, but there are currently no explicit constructions, [HLM⁺25] poses constructing these expanders as an open problem.

Observe that HD_1^n shows a separation between approximate γ_2 -norm and blocky rank, because approximate γ_2 -norm of HD_1^n is $O(1)$ whereas its blocky rank at least \sqrt{n} by Theorem 1.8.

Acknowledgments

We thank Nathan Harms for granting permission to include his proof of the upper bound in Theorem A.2, as well as for helpful feedback that improved the presentation of the paper. We are grateful to Dmitry Sokolov for bringing ultra-lossless expanders to our attention, and to Amir Yehudayoff for pointing us to Warren’s theorem for the proof of Theorem 3.3. Finally, we thank Nati Linial, Shachar Lovett, Ryan Williams, and Amir Yehudayoff for many insightful discussions on spiky rank.

References

- [Alo86] Noga Alon. Covering graphs by the minimum number of equivalence relations. *Combinatorica*, 6(3):201–206, 1986.
- [Alo21] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, 41(4):447–463, 2021.

- [AMY16] Noga Alon, Shay Moran, and Amir Yehudayoff. Sign rank versus vc dimension. In *Conference on Learning Theory*, pages 47–80. PMLR, 2016.
- [AN25] Benny Applebaum and Oded Nir. The meta-complexity of secret sharing. In *Proceedings of the 57th Annual ACM Symposium on Theory of Computing*, pages 965–976, 2025.
- [AW17] Josh Alman and Ryan Williams. Probabilistic rank and matrix rigidity. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2017, page 641–652, New York, NY, USA, 2017. Association for Computing Machinery.
- [AY22] Daniel Avraham and Amir Yehudayoff. On blocky ranks of matrices. Technical Report TR22-137, Electronic Colloquium on Computational Complexity (ECCC), September 2022.
- [AY24] Daniel Avraham and Amir Yehudayoff. On blocky ranks of matrices. *computational complexity*, 33, 03 2024.
- [BBM⁺21] Alexander R. Block, Simina Brânzei, Hemanta K. Maji, Himanshi Mehta, Tamalika Mukherjee, and Hai H. Nguyen. P_4 -free partition and cover numbers & applications. In Stefano Tessaro, editor, *2nd Conference on Information-Theoretic Cryptography (ITC 2021)*, volume 199 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 16:1–16:25, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [BHT25] Igor Balla, Lianna Hambardzumyan, and István Tomon. Factorization norms and an inverse theorem for maxcut. *arXiv preprint arXiv:2506.23989*, 2025.
- [CHZZ24] Tsun Ming Cheung, Hamed Hatami, Rosie Zhao, and Itai Zilberstein. Boolean functions with small approximate spectral norm. *Discrete Analysis*, sep 18 2024.
- [CMS20] Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. *Journal of the ACM*, 67, 2020.
- [DGJ⁺10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010.
- [Fra82] Péter Frankl. Covering graphs by equivalence relations. In *North-Holland Mathematics Studies*, volume 60, pages 125–127. Elsevier, 1982.
- [GG12] Martin Golubitsky and Victor Guillemin. *Stable mappings and their singularities*, volume 14. Springer Science & Business Media, 2012.
- [GH25] Marcel K Goh and Hamed Hatami. Block complexity and idempotent schur multipliers. *International Mathematics Research Notices*, 2025(24):356, 2025.
- [GHIS25] Mika Göös, Nathaniel Harms, Valentin Imbach, and Dmitry Sokolov. Sign-rank of k -hamming distance is constant. *arXiv preprint arXiv:2506.12022*, 2025.
- [GHR25] Mika Göös, Nathaniel Harms, and Artur Riazanov. Equality is far weaker than constant-cost communication. *arXiv preprint arXiv:2507.11162*, 2025.
- [Gri80] Dima Grigoriev. An application of separability and independence notions for proving lower bounds of circuit complexity. *J.Soviet Math.*, 14:1450–1456, 01 1980.

- [HHH23] Lianna Hambardzumyan, Hamed Hatami, and Pooya Hatami. Dimension-free bounds and structural results in communication complexity. *Israel Journal of Mathematics*, 253(2):555–616, 2023.
- [HLM⁺25] Jun-Ting Hsieh, Alexander Lubotzky, Sidhant Mohanty, Assaf Reiner, and Rachel Yun Zhang. Explicit lossless vertex expanders. *arXiv preprint arXiv:2504.15087*, 2025.
- [HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006.
- [HP10] Kristoffer Arnsfelt Hansen and Vladimir V Podolskii. Exact threshold circuits. In *2010 IEEE 25th Annual Conference on Computational Complexity*, pages 270–279. IEEE, 2010.
- [Juk06] Stasys Jukna. On graph complexity. *Combinatorics, Probability and Computing*, 15(6):855–876, 2006.
- [KSP20] Rohan Karthikeyan, Siddharth Sinha, and Vallabh Patil. On the resolution of the sensitivity conjecture. *Bulletin of the American Mathematical Society*, 57:1, 03 2020.
- [KW16] Daniel M Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 633–643, 2016.
- [LS⁺09] Troy Lee, Adi Shraibman, et al. Lower bounds in communication complexity. *Foundations and Trends[®] in Theoretical Computer Science*, 3(4):263–399, 2009.
- [MB17] Anirbit Mukherjee and Amitabh Basu. Lower bounds over boolean inputs for deep neural networks with relu gates. *arXiv preprint arXiv:1711.03073*, 2017.
- [PR94] P. Pudlák and V. Rödl. Some combinatorial-algebraic problems from complexity theory. *Discrete Mathematics*, 136(1):253–279, 1994.
- [PSS23] Toniann Pitassi, Morgan Shirley, and Adi Shraibman. The strength of equality oracles in communication. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 89:1–89:19, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.
- [Raz89] Alexander Razborov. On rigid matrices (in russian). 1989.
- [ROS94] Vwani P Roychowdhury, Alon Orlitsky, and Kai-Yeung Siu. Lower bounds on threshold and related circuits via communication complexity. *IEEE Transactions on Information Theory*, 40(2):467–474, 1994.
- [RY20] Anup Rao and Amir Yehudayoff. *Communication Complexity and Applications*. January 2020. Publisher Copyright: © Anup Rao and Amir Yehudayoff 2020.
- [SS96] M. Sipser and D.A. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996.
- [Val77] Leslie G Valiant. Graph-theoretic arguments in low-level complexity. In *Mathematical Foundations of Computer Science 1977: Proceedings, 6th Symposium, Tatranská Lomnica September 5–9, 1977 6*, pages 162–176. Springer, 1977.

- [VC71] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.
- [War68] Hugh E Warren. Lower bounds for approximation by nonlinear manifolds. *Transactions of the American Mathematical Society*, 133(1):167–178, 1968.
- [Wil18] Ryan Williams. Limits on representing boolean functions by linear combinations of simple functions: Thresholds, relus, and low-degree polynomials. *arXiv preprint arXiv:1802.09121*, 2018.
- [Wil24a] Ryan Williams. The Orthogonal Vectors Conjecture and non-uniform circuit lower bounds. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1372–1387, October 2024.
- [Wil24b] Ryan Williams. Personal communication. 2024.
- [Wun12] Henning Wunderlich. On a theorem of razborov. *computational complexity*, 21:431–477, 2012.

A Sign and Approximate versions of Blocky rank

A.1 Blocky rank and Sign Blocky Rank

We refer to [HHH23] for the definition of D^{Eq} : deterministic communication complexity with EQUALITY oracle.

Theorem A.1. $\text{br}(M) \leq 2^{\text{br}_{\pm}(M)}$.

Proof. As shown in [HHH23] we have $D^{\text{Eq}}(M) \geq \log \text{br}(M)$. On the other hand observe that $D^{\text{Eq}}(M) \leq \text{br}_{\pm}(M)$: let $M = \text{sign}(\alpha_1 B_1 + \dots + \alpha_r B_r)$ for blocky matrices B_1, \dots, B_r . Alice and Bob then compute $B_i(x, y)$ for each $i \in [r]$, where x and y are their inputs, taking r Equality queries. After this computation both of them can compute $M(x, y) = \text{sign}(\sum_{i \in [r]} B_i(x, y) \alpha_i)$ locally. Taking the upper and the lower bound for $D^{\text{Eq}}(M)$ together implies the claimed relation. \square

A.2 Sign Blocky Rank Bound of HD_1^n

We show a tight bound for the sign blocky rank of HD_1^n .

Theorem A.2. $\text{br}_{\pm}(\text{HD}_1^n) = \Theta(\log n)$.

Proof. The lower bound follows directly from Theorem 1.8 and Theorem A.1, so we proceed to prove the upper bound. Without loss of generality, suppose that n is a power of 2, otherwise increase n so it is. Let $M_{i,j}[x, y]$ for $i \in [\log_2 n]$, $j \in \{0, 1\}$, $x, y \in \{0, 1\}^n$ be defined as follows: $M_{i,j}[x, y] = 0$ iff x and y coincide on all indices t with the i th bit in the representation equal to j and 1 otherwise. For example if $M_{i,0}[x, y] = M_{i,1}[x, y] = 0$ then $x = y$ and vice versa. If x and y differ in exactly one position $\ell \in [n]$ then for every $i \in [\log_2 n]$ we have $M_{i,0}[x, y] + M_{i,1}[x, y] = 1$. If x and y differ in positions $\ell_1 \neq \ell_2$, then for $i \in [\log_2 n]$ that corresponds to the difference in bit representations of ℓ_1 and ℓ_2 we get $M_{i,0}[x, y] + M_{i,1}[x, y] = 2$.

Therefore x, y differ in exactly one position iff $\sum_{i \in [\log_2 n]; j \in \{0,1\}} M_{i,j}[x, y] \leq k := \log_2 n$ and $x \neq y$. Let $I \in \{0, 1\}^{2^n \times 2^n}$ be the identity matrix and $J \in \{0, 1\}^{2^n \times 2^n}$ be the all-1 matrix. Then

$$\text{HD}_1^n = \text{sign} \left(- \sum_{i \in [\log_2 n]; j \in \{0,1\}} M_{i,j} - (k+1)I + (k+1/2)J \right). \quad \square$$

A.3 Approximate Blocky Rank Upper Bound for HD_1^n

In this section we adapt [Theorem A.2](#) to the setting of approximate blocky rank.

We are going to use two simple technical claims for the error amplification.

Lemma A.3. *Let $M \in \mathbb{R}^{n \times n}$ be a matrix and p be a univariate polynomial over reals. Let $p(M) \in \mathbb{R}^{n \times n}$ be defined by $p(M)[x, y] := p(M[x, y])$. We then have $\text{br}(p(M)) \leq \text{br}(M)^{\deg(p)} \cdot \deg(p)$.*

Proof. It suffices to prove the lemma for $p = x^i$. Then for $M = \sum_{j \in [\text{br}(M)]} \alpha_j B_j$ we have $p(M) = M^{oi}$ (where M^{oi} is entrywise product of i copies of M). Then we have $M^{oi} = \sum_{j_1, \dots, j_i \in [\text{br}(M)]} \prod_{k \in [i]} \alpha_{j_k} B_{j_k} \circ \dots \circ B_{j_i}$. The proof is finished by observing that entrywise product of blocky matrices is a blocky matrix. \square

The following claim directly follows from the Chernoff-Hoeffding bound.

Lemma A.4 (see e.g. [\[DGJ+10\]](#)). *Let $A_k(x) := \sum_{i=k/2}^k \binom{k}{i} x^i (1-x)^{k-i}$. Then for $x \in [0, 1/2 - \delta)$ we have $A_k(x) \leq e^{-2k/\delta^2}$ and for $x \in (1/2 + \delta, 1]$ we have $A_k(x) \geq 1 - e^{-2k/\delta^2}$.*

Theorem A.5. $\text{br}_\varepsilon(\text{HD}_1^n) = (\log n)^{O(\log(1/\varepsilon))}$.

Proof. As before, suppose that n is a power of 2 without loss of generality, otherwise increase n so it is. Consider an error-correcting code $C \subseteq \{0, 1\}^{K \log_2 n}$ with $|C| = n$ and distance $(K \log_2 n)/5$ (for example, an expander code with appropriate parameters suffices [\[SS96\]](#)) i.e. for every $\alpha \neq \beta \in C$ we have $|\alpha \oplus \beta| \geq (K \log_2 n)/5$. Associate each index $\ell \in [n]$ with a codeword C_ℓ . Now for $i \in [K \log_2 n]$ and $j \in \{0, 1\}$ define $M_{i,j} \in \{0, 1\}^{2^n \times 2^n}$ as follows: $M_{i,j}[x, y] = 0$ iff x and y coincide on all indices $\ell \in [n]$ such that $(C_\ell)_i = j$.

Now suppose x and y differ in exactly one position ℓ : then for every $i \in [K \log_2 n]$ we have $M_{i,0}[x, y] + M_{i,1}[x, y] = 1$.

If x and y differ in at least two positions ℓ_1 and ℓ_2 , then for every $i \in [K \log_2 n]$ we have $M_{i,0}[x, y] + M_{i,1}[x, y] \geq 1$ and for every i such that C_{ℓ_1} and C_{ℓ_2} differ we get $M_{i,0}[x, y] + M_{i,1}[x, y] = 2$.

Thus, assuming $x \neq y$ we get $M'[x, y] := \sum_{i \in [K \log_2 n]; j \in \{0,1\}} M_{i,j}[x, y] \leq K \log_2 n$ if x and y differ in one position and $M'[x, y] \geq K \log_2 n (1 + 1/5)$ if x and y differ in at least two positions. $M'[x, x] = 0$ for every $x \in \{0, 1\}^n$.

Let $I \in \{0, 1\}^{2^n \times 2^n}$ be the identity matrix and $J \in \{0, 1\}^{2^n \times 2^n}$ be the all-1 matrix. Then letting

$$M := \frac{1}{2K \log_2 n} (-M' - (K \log_2 n + 1)I + ((2K - 1/10) \log_2 n)J)$$

we get $\|\text{HD}_1^n - M\|_\infty \leq 1/2 - 1/10K$. Now taking $M_{\text{final}} := A_{200K^2 \log(1/\varepsilon)}(M)$ by [Lemma A.4](#) and [Lemma A.3](#) we get the claim. \square