

Visible Rank and Codes with Locality*

Omar Alrabiah[†]Venkatesan Guruswami[‡]

Abstract

We propose a framework to study the effect of local recovery requirements of codeword symbols on the dimension of linear codes, based on a combinatorial proxy that we call *visible rank*. The locality constraints of a linear code are stipulated by a matrix H of \star 's and 0's (which we call a "stencil"), whose rows correspond to the local parity checks (with the \star 's indicating the support of the check). The visible rank of H is the largest r for which there is a $r \times r$ submatrix in H with a unique generalized diagonal of \star 's. The visible rank yields a field-independent combinatorial lower bound on the rank of H and thus the co-dimension of the code.

We point out connections of the visible rank to other notions in the literature such as unique restricted graph matchings, matroids, spanoids, and min-rank. In particular, we prove a rank-nullity type theorem relating visible rank to the rank of an associated construct called *symmetric spanoid*, which was introduced by Dvir, Gopi, Gu, and Wigderson [DGGW20]. Using this connection and a construction of appropriate stencils, we answer a question posed in [DGGW20] and demonstrate that symmetric spanoid rank cannot improve the currently best known $\tilde{O}(n^{(q-2)/(q-1)})$ upper bound on the dimension of q -query locally correctable codes (LCCs) of length n . This also pins down the efficacy of visible rank as a proxy for the dimension of LCCs.

We also study the t -Disjoint Repair Group Property (t -DRGP) of codes where each codeword symbol must belong to t disjoint check equations. It is known that linear codes with 2-DRGP must have co-dimension $\Omega(\sqrt{n})$ (which is matched by a simple product code construction). We show that there are stencils corresponding to 2-DRGP with visible rank as small as $O(\log n)$. However, we show the second tensor of any 2-DRGP stencil has visible rank $\Omega(n)$, thus recovering the $\Omega(\sqrt{n})$ lower bound for 2-DRGP. For q -LCC, however, the k 'th tensor power for $k \leq n^{o(1)}$ is unable to improve the $\tilde{O}(n^{(q-2)/(q-1)})$ upper bound on the dimension of q -LCCs by a polynomial factor. Inspired by this and as a notion of intrinsic interest, we define the notion of *visible capacity* of a stencil as the limiting visible rank of high tensor powers, analogous to Shannon capacity, and pose the question whether there can be large gaps between visible capacity and algebraic rank.

*Research supported in part by NSF grants CCF-1814603 and CCF-1908125 and a Simons Investigator award.

[†]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA. Email: oalrabia@andrew.cmu.edu

[‡]Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213. Email: venkatg@cs.cmu.edu.

1 Introduction

The notion of *locality* in error-correcting codes refers to the concept of recovering codeword symbols as a function of a small number of other codeword symbols. Local decoding requirements of various kinds have received a lot of attention in coding theory, due to both their theoretical and practical interest. For instance, q -query locally correctable codes (LCCs) aim to recover any codeword symbol as a function of q other codeword symbols in a manner robust to a constant fraction of errors. On the other hand, locally recoverable codes (LRCs), in their simplest incarnation, require each codeword symbol to be a function of some ℓ other codeword symbols, allowing local recovery from any single erasure.¹

LCCs have been extensively studied in theoretical computer science, and have connections beyond coding theory to topics such as probabilistically checkable proofs and private information retrieval. We refer the reader to [Yek12] and the introduction of [Gop18] for excellent surveys on LCCs and their connections. LRCs were motivated by the need to balance global fault tolerance with extremely efficient repair of a small number of failed storage nodes in modern large-scale distributed storage systems [GHSY12]. They have led to intriguing new theoretical questions, and have also had significant practical impact with adoption in large scale systems such as Microsoft Azur [HSX⁺12] and Hadoop [SAP⁺13].

Let us define the above notions formally, in a convenient form that sets up this work. We will restrict attention to linear codes in this work, i.e., subspaces C of \mathbb{F}_q^n for some finite field \mathbb{F}_q . In this case, the i 'th symbol c_i of every codeword $c = (c_1, \dots, c_n) \in C$ can be recovered as a function of the symbols c_j , for indices j in a (minimal) subset $R_i \subset [n] \setminus \{i\}$, iff c_i and $\{c_j \mid j \in R_i\}$ satisfy a linear check equation, or in other words, there is a dual codeword whose support equals $\{i\} \cup R_i$. The set R_i is called a *repair group* for the i 'th codeword symbol (other terminology used in the literature includes regenerating sets and recovery sets).

The q -LCC property, for a fixed number of queries q and growing n , corresponds to having $\Omega(n)$ disjoint groups of size $\leq q$ for each position $i \in [n]$, or equivalently $\Omega(n)$ dual codewords of Hamming weight at most $(q + 1)$ whose support includes i and are otherwise disjoint. The ℓ -LRC corresponds to having a dual codeword of Hamming weight at most $(\ell + 1)$ whose support includes i , for each $i \in [n]$. A property that interpolates between these extremes of a single repair group and $\Omega(n)$ disjoint repair groups is the Disjoint Repair Group Property (t -DRGP) where we require t disjoint repair groups for each position $i \in [n]$ (equivalently t dual codewords whose support includes i but are otherwise disjoint).

There is an exponentially large gap between upper and lower bounds on the trade-off between code dimension and code length for q -LCCs. The best known code constructions have dimension only $O((\log n)^{q-1})$ (achieved by generalized Reed-Muller codes or certain lifted codes [GKS13]), whereas the best known upper bound on the dimension of q -LCCs is much larger and equals $\tilde{O}(n^{(q-2)/(q-1)})$ [KT00, Woo12, IS20]². Narrowing this huge gap has remained open for over two decades.

In contrast the best possible dimension of a ℓ -LRC is easily determined to be $\lfloor \frac{\ell n}{\ell+1} \rfloor$.³ However, for t -DRGP, there are again some intriguing mysteries. For 2-DRGP, we have tight bounds—the

¹There is also a distance requirement on LRCs to provide more global error/erasure resilience.

²The $\tilde{O}(\cdot)$ and $\tilde{\Omega}(\cdot)$ are used to suppress factors poly-logarithmic in n .

³In this case, a more interesting trade-off is a Singleton-type bound that also factors in the distance of the code [GHSY12].

minimum possible redundancy (co-dimension) equals $\sqrt{2n} \pm \Theta(1)$. The lower bound is established via very elegant proofs based on the polynomial method [Woo16] or rank arguments [RV16]. However, for fixed $t > 2$, we do not know better lower bounds, and the best known constructions have co-dimension $\approx t\sqrt{n}$ [FVY15]. There are better constructions known for some values of $t = n^{\Theta(1)}$ [FGW17, LW21]. A lower bound on the co-dimension of $c(t)\sqrt{n}$ for some function $c(t)$ that grows with t seems likely, but has been elusive despite various attempts, and so far for any fixed t , the bound for $t = 2$ is the best known.

This work was motivated in part by these major gaps in our knowledge concerning q -LCCs and t -DRGPs. Our investigation follows a new perspective based on *visible rank* (to be defined soon), which is a combinatorial proxy for (linear-algebraic) rank that we believe is of broader interest. This is similar in spirit to a thought-provoking recent work [DGGW20] that introduced a combinatorial abstraction of spanning structures called *spanoids*⁴ to shed light on the limitations of current techniques to prove better upper bounds on the dimension of q -LCCs. They noted that current techniques to bound LCC dimension apply more generally to the associated spanoids, which they showed could have rank as large as $\tilde{\Omega}(n^{(q-2)/(q-1)})$. Therefore to improve the LCC bound one needs techniques that are more specific than spanoids and better tailored to the LCC setting. One such possibility mentioned in [DGGW20] is to restrict attention to *symmetric spanoids*, which have a natural symmetry property that linear LCCs imply.

Our visible rank notion turns out to be intimately related to symmetric spanoids via a rank-nullity type theorem (Theorem 9). While technically simple in hindsight, it offers a powerful viewpoint on symmetric spanoids which in particular resolves a question posed in [DGGW20]—we show that symmetric spanoids are also too coarse a technique to beat the $\tilde{O}(n^{(q-2)/(q-1)})$ upper bound on q -LCC dimension.

1.1 Stencils and visible rank

With the above backdrop, we now proceed to describe the setup we use to study these questions, based on the rank of certain matrix templates which we call “stencils.” We can represent the support structure of the check equations (i.e., dual codewords) governing a locality property by an n -column matrix of 0’s and \star ’s. For each check equation involving the i ’th symbol and a repair group $R_i \subset [n] \setminus \{i\}$, we place a row in the stencil with \star ’s precisely at $R_i \cup \{i\}$ (i.e., with \star ’s at the support of the associated dual codeword). For the ℓ -LRC property for instance, an associated stencil would be an $n \times n$ matrix with \star ’s on the diagonal and ℓ other \star ’s in each row. For q -LCC, we would have a $\delta n^2 \times n$ matrix whose rows are split into n groups with the rows in the i ’th group having a \star in the i ’th column and q other \star ’s in disjoint columns.

The smallest co-dimension of linear codes over a field \mathbb{F} with certain locality property is, by design, the minimum rank $\text{rk}_{\mathbb{F}}(H)$ of the associated stencil H when the \star ’s are replaced by arbitrary nonzero entries from \mathbb{F} . In this work, our goal is to understand this quantity via field oblivious methods based only on the combinatorial structure of the stencil of \star ’s.

The tool we put forth for this purpose is the *visible rank* of H , denoted $\text{vrk}(H)$ and defined to be the largest r for which there is a $r \times r$ submatrix of H that has exactly one general diagonal whose entries are all \star ’s. By the Leibniz formula, the determinant of such a submatrix is nonzero for any substitution of nonzero entries for the \star ’s. Thus $\text{rk}_{\mathbb{F}}(H) \leq \text{vrk}(H)$ for every field \mathbb{F} .

Our goal in this work is to understand the interrelationship between visible rank and the co-

⁴We defer a precise description of spanoids, along with their strong connection to visible rank, to Section 2.4.

dimension of linear codes under various locality requirements. This can shed further light on the bottleneck in known techniques to study trade-offs between locality and code dimension, and optimistically could also lead to better constructions.

1.2 Visible rank and Locality

For ℓ -LRCs, a simple greedy argument shows that its associated parity-check stencil H satisfies $\text{vrk}(H) \geq n/(\ell + 1)$. Thus visible rank captures the optimal trade-off between code dimension and locality ℓ .

For q -LCCs with $q \geq 3$, an argument similar to (in fact a bit simpler than and implied by) the one for spanoids in [DGGW20] shows that the stencil corresponding to q -LCCs has visible rank at least $n - \tilde{O}(n^{(q-2)/(q-1)})$, showing an upper bound of $\tilde{O}(n^{(q-2)/(q-1)})$ on the dimension of q -LCCs. We show that visible rank suffers the same bottleneck as spanoids in terms of bounding the dimension of q -LCCs.

Theorem 1. *For $q \geq 3$, there exist n -column stencils H with \star 's structure compatible with q -LCCs for which $\text{vrk}(H) \leq n - \tilde{\Omega}(n^{(q-2)/(q-1)})$.*

Through the precise connection we establish between visible rank and symmetric spanoids, this shows the same limitation for symmetric spanoids, thus answering a question posed in [DGGW20].

For the t -DRGP property, we focus on the $t = 2$ case, with the goal of finding a combinatorial substitute for the currently known $\Omega(\sqrt{n})$ lower bounds on co-dimension [Woo16, FGW17] which are algebraic. Unfortunately, we show that visible rank, in its basic form, is too weak in this context.

Theorem 2. *There exist $2n \times n$ stencils H with \star 's structure compatible with 2-DRGP for which $\text{vrk}(H) \leq O(\log n)$.*

1.3 Visible rank and tensor powers

In view of Theorem 2, we investigate avenues to get better bounds out of the visible rank approach. Specifically, we study the visible rank of tensor powers of the matrix. It turns out that the visible rank is super-multiplicative: $\text{vrk}(H \otimes H) \geq \text{vrk}(H)^2$, while on the other hand algebraic rank is sub-multiplicative, so higher tensor powers could yield better lower bounds on the rank. Indeed, we are able to show precisely this for 2-DRGP:

Theorem 3. *For every $2n \times n$ stencil H with \star 's structure compatible with 2-DRGP, we have $\text{vrk}(H \otimes H) \geq \Omega(n)$, and thus $\text{rk}_{\mathbb{F}}(H) \geq \Omega(\sqrt{n})$ for every field \mathbb{F} .*

On the other hand, for q -LCCs with $q \geq 3$, we show that higher tensor powers suffer the same bottleneck as Theorem 1.

Theorem 4. *For $q \geq 3$, there exist n -column stencils H with \star 's structure compatible with q -LCCs for which $\text{vrk}(H^{\otimes k})^{1/k} \leq n - \tilde{\Omega}(n^{(q-2)/(q-1)})/k$ for any integer k . In particular even for $k = n^{o(1)}$, we get no polynomial improvements to the current upper bounds on dimension of q -LCCs.*

1.4 Visible capacity

Given the super-multiplicativity of visible rank under tensor powers, and drawing inspiration from the Shannon capacity of graphs, we put forth the notion of *visual capacity* of a matrix H of 0's and \star 's, defined as

$$\Upsilon(H) := \sup_k \text{vrk}(H^{\otimes k})^{1/k}.$$

The visual capacity is also a field oblivious lower bound on algebraic rank $\text{rk}_{\mathbb{F}}(H)$ for any field \mathbb{F} . It is not known whether there are stencils that exhibit a gap between visible capacity and its minimum possible $\text{rk}_{\mathbb{F}}(H)$ over all fields \mathbb{F} .

The proofs of our results are technically simple, once the framework is set up. Our contributions are more on the conceptual side, via the introduction and initial systematic study of visible rank and its diverse connections. Our inquiry also raises interesting questions and directions for future work, some of which are outlined in Section 7, including the relationship between visible capacity and algebraic rank.

1.5 Connections and related work

Studying the interplay between the combinatorial structure of a matrix and its rank is a natural quest that arises in several contexts. See Chapter 3 of [Tré16] for a survey of works on lower bounding the algebraic rank. For works specific to codes with locality, the work of [BDYW11] analyzed the combinatorial properties of design matrices over the reals to improve bounds on LCCs over the real numbers, although the methods used are particular to the field of reals and do not carry over to any field.

Visible rank in particular turns out to have a diverse array of connections, some of which we briefly discuss here. The connection to spanoids, that we already mentioned, is described in more detail in Section 2.4.

Uniquely restricted matchings. Given a stencil $H \in \{0, \star\}^{m \times n}$, there is a *canonical bipartite graph* G between the rows and columns of H , where a row connects to a column if and only if their shared entry has a star. Visual rank has a nice graph-theoretic formulation: it turns out (see Section 2.3) that a submatrix of H has a unique general diagonal of \star 's iff the corresponding induced subgraph of G has a unique perfect matching. Such induced bipartite graphs are known in the literature as *Uniquely Restricted Matchings (URMs)* and have been extensively studied [GHL01, GKT01, HMT06, Mis11, TD13, FJJ18]. They were first introduced in [GHL01], wherein they proved that computing the maximum URM of a bipartite graph is NP-complete. It was later shown in [Mis11] that $n^{1/3-o(1)}$ approximations of the maximum URM is also NP-hard unless $\text{NP} = \text{ZPP}$ and additionally that the problem of finding the maximum URM is APX-complete.

Matroids. One can encode any matroid into a stencil. Recall that a circuit of a matroid is a minimal dependent set—that is, a dependent set whose proper subsets are all independent (the terminology reflects the fact that in a graphic matroid, the circuits are cycles of the graph). Given a matroid \mathcal{M} on universe $[n]$ and a set $\mathcal{C} = \{C_1, \dots, C_m\}$ of circuits of \mathcal{M} , we consider a $m \times n$ stencil H where the entry at (i, j) is a \star if and only if $j \in C_i$. For this matrix, one can show that a collection of visibly independent columns (see Section 2.2 for the definition of visible independence) is an independent set in the dual matroid. Therefore, we have $\text{rk}(\mathcal{M}) + \text{vrk}(H) \leq n$ —this also follows from our rank-nullity theorem for symmetric spanoids as one can associate a symmetric spanoid with any matroid (the collection of sets in Definition 5 will just be the circuits of the matroid).

Min-rank. The minimum possible rank of a square $0\text{-}\star$ stencil over assignments to the \star 's from some field has been well studied in combinatorics.⁵ For example, we have Haemers' classic bound on independent set of a graph and its applications to Shannon capacity [Hae79]. Note that in this case we are using a linear-algebraic tool to understand a combinatorial quantity, whereas visible rank goes the other way, serving as a combinatorial proxy for a linear-algebraic quantity. Recent interest in minrank has included their characterization of the most efficient linear index codes [BBJK11]. The minrank of stencils corresponding to n -vertex random Erdős-Rényi graphs was recently shown to be $\Theta(n/\log n)$ over any field that is polynomially bounded [GRW18].

Matrix Rigidity. Given a square matrix $A \in \mathbb{F}^{n \times n}$ and a natural number $r \leq n$, the *rigidity* of A is the minimal number of entries that one can perturb in A so that its rank becomes at most r . Matrix rigidity was introduced in the seminal work [Val77] and since then had expansive research on constructing explicit rigid matrices. See [Ram20] for a recent survey on matrix rigidity and related connections. The visible rank provides a combinatorial guarantee on the rank of a matrix, and that conjures up the possibility of constructing explicit rigid matrices by finding explicit stencils whose visible rank is robust to small amounts of corruptions of its entries. We pose this approach as Question 6 in Section 7.

Incidence Theorems. Given an $m \times n$ matrix A over the field \mathbb{F} with rank r , one can decompose $A = MN$ where M and N are $m \times r$ and $r \times n$ matrices. If we consider the rows of M as hyperplanes over the projective plane $\mathbb{P}\mathbb{F}^{r-1}$ of dimension $(r-1)$ and the columns of N as points in $\mathbb{P}\mathbb{F}^{r-1}$, then the stencil of A defines a point-hyperplane incidence over $\mathbb{P}\mathbb{F}^{r-1}$. In particular, when $r = 3$, the stencil of A defines a point-line incidence over the field \mathbb{F} . Thus studying the combinatorial properties of a stencil whose \mathbb{F} -rank (see Definition 2) is at most 3 is equivalent to studying the combinatorics of point-line incidences over the field \mathbb{F} . For more on incidence theorems, see [Dvi12] for an excellent survey in the area.

Communication complexity. The visible rank provides a connection between deterministic and nondeterministic communication complexity [Lov89]. For a communication problem $f : X \times Y \rightarrow \{0, 1\}$, define the stencil $H_f \in \{0, \star\}^{X \times Y}$ by $H_f(x, y) = \star$ if $f(x, y) = 0$ and $M_f(x, y) = 0$ if $f(x, y) = 1$. Then it is known that $D(f) \leq (\log_2 \text{vrk}(H_f)) \cdot (N(f) + 1)$ where $D(f)$ and $N(f)$ are respectively the deterministic and nondeterministic communication complexity of f [Lov89, Thm 3.5].

Distance of expander codes. We can associate a stencil with the parity-check matrix H of a code C in the natural way, by placing \star 's at the positions of the nonzero field elements. Suppose that every subset of d columns of the stencil is visibly independent. Then the minimum distance of C is at least d . This is in fact the lower bound on distance of expander codes established in [SS96]. The visible independence of subsets of d columns is argued via *unique expansion* of the canonical bipartite graph associated with H (also called the factor graph in coding theory). Namely, for every subset S of d vertices on the left, there is a vertex on the right that's adjacent to exactly one vertex in S . Thus, linear independence is really argued via visible independence.

1.6 Organization

We begin in Section 2 by formally introducing the notations and terminology for stencils, and establishing some simple but very useful combinatorial facts about visible rank. We use these to

⁵There is a slight difference in the minrank setup, in that the \star 's can take any value including 0, except the \star 's on the diagonal which must take nonzero values.

show that there are q -LCC stencils for $q \geq 3$ with visible rank at most $n - \tilde{\Omega}(n^{(q-2)/(q-1)})$ (Section 3), and the existence of a 2-DRGP stencil with visible rank of at most $O(\log n)$ (Section 4). In Section 5, we introduce a tensor product operation on stencils and prove various properties about them. In Section 6, we utilize tensor powers to show that the rank of a 2-DRGP over any field \mathbb{F} is at least \sqrt{n} , which asymptotically matches the current best lower bounds on t -DRGP codes. We also show that for q -LCC stencils, the tensor powers at the k 'th level for $k \leq \text{polylog}(n)$ do not yield better lower bounds on the rank than the ones obtained from the visible rank. Finally, in Section 7, we discuss further directions and questions inspired by this work.

2 Stencils and their visible rank

In this section, we will be formally setting up the model of stencils and all the associated definitions and notations.

We denote $[n]$ to be the set $\{1, 2, \dots, n\}$. For any matrix $H \in \{0, \star\}^{m \times n}$, we denote it as a *stencil*. For an $m \times n$ stencil H , we denote its entry in the i 'th row and j 'th column by $H[i, j]$. Any restriction to the specific sub-collection of the rows and columns of H is said to be a *sub-stencil* of H . For given sets A and B , a stencil H is said to be an $A \times B$ if it is an $|A| \times |B|$ stencil along with an associated indexing of the rows by A and the columns by B . Given a square stencil $M \in \{0, \star\}^{n \times n}$, a *general diagonal* of M , is a collection of entries $\{M[1, \pi(1)], \dots, M[n, \pi(n)]\}$ where π is a permutation on $[n]$. We say that a general diagonal is a *star diagonal* if all its n entries are \star 's.

2.1 Algebraic witnesses of stencils

Instantiating a code with the locality properties stipulated by a stencil amounts to filling its \star 's with field entries, or realizing an algebraic witness as defined below.

Definition 1 (Algebraic witness). *For field \mathbb{F} and stencil $H \in \{0, \star\}^{m \times n}$, a matrix $W \in \mathbb{F}^{m \times n}$ is said to be an \mathbb{F} -witness of H if it satisfies the property that $W[i, j] \neq 0$ if and only if $H[i, j] = \star$. More generally, any \mathbb{F} -witness of H is said to be an algebraic witness of H .*

We stress that every \star in the stencil H must be replaced by a *nonzero* entry from \mathbb{F} and cannot be zero. Of the possible algebraic witnesses for H , we will be primarily focused in this paper on the algebraic witnesses that attain the smallest feasible rank, which leads us to the following definition.

Definition 2 (Rank). *Given an $m \times n$ stencil H , the \mathbb{F} -rank of H is the smallest natural number r such that there exists a field \mathbb{F} and an \mathbb{F} -witness $W \in \mathbb{F}^{m \times n}$ whose rank is equal to r . We denote the value r by $\text{rk}_{\mathbb{F}}(H)$.*

2.2 Visible Rank

In this section, we introduce our notion of the *visible rank* of a stencil. The main motivation of introducing the visible is to be able to determine the most optimal lower bound on the rank of a matrix with only the knowledge of knowing the support of a matrix and nothing else about the values of that support.

Consider a square matrix $A \in \mathbb{F}^{n \times n}$, and suppose we are interested in determining if it is full rank. A natural approach would be to inspect its determinant. From the Leibniz formula, we know

that $\det(A) = \sum_{\pi \in S_n} (-1)^{\text{sgn}(\pi)} \prod_{i=1}^n A_{i,\pi(i)}$, where S_n denotes the symmetric group of order n and $\text{sgn}(\pi)$ denotes the sign of a permutation π . From Leibniz formula, notice that $\det(A)$ is a linear combination of the nonzero general diagonals of A . If our hope is to obtain $\det(A) \neq 0$ without any knowledge of the values of the support of A , one way to guarantee it is to say that A has exactly one nonzero general diagonal. In such a case, we can guarantee that $\det(A) \neq 0$. As when A has more than one general diagonal, there is no guarantee if $\det(A) \neq 0$ without inspecting the values of the support of A .

From the previous discussion, it seems natural to define the notion of a rank on stencils as follows.

Definition 3 (Visibly Full Rank). *For a square stencil $M \in \{0, \star\}^{n \times n}$, we say that M is visibly full rank if M has exactly one star diagonal. That is, a general diagonal whose entries are all \star 's.*

Of course, in most cases, when we are given a matrix $A \in \mathbb{F}^{m \times n}$, we would be interested in determining its rank. One way to define the rank of the matrix A is to say that $\text{rank}(A)$ is the size of the largest square submatrix in A that is full-rank. From this viewpoint, it seems clear to define the rank of a stencil in a similar fashion.

Definition 4 (Visible Rank). *For a stencil $H \in \{0, \star\}^{m \times n}$, the visible rank of H , denoted $\text{vrk}(H)$, is the largest square sub-stencil in H that is visibly full rank.*

We also say that a set of k columns in H is *visibly independent* if there exists a $k \times k$ sub-stencil within these k columns that is visibly full rank.

Of course, not all full-rank square matrices $A \in \mathbb{F}^{n \times n}$ necessarily have exactly one nonzero general diagonal, but all squares that have exactly one nonzero general diagonal are necessarily full-rank. Thus if we are interested in determining $\text{rank}(A)$ by finding the size of the largest square submatrix in A that is full-rank, we can instead search for the largest square submatrix in A that has exactly one nonzero general diagonal, and that will lead us to lower bounds on the rank of A .

Proposition 5. *Given a field \mathbb{F} and stencil $H \in \{0, \star\}^{m \times n}$, we have the inequality $\text{rk}_{\mathbb{F}}(H) \geq \text{vrk}(H)$.*

Proof. Let $W \in \mathbb{F}^{m \times n}$ be algebraic witness of H . Let H_0 be a $k \times k$ square sub-stencil of H that is visibly independent. Let W_0 be the corresponding $k \times k$ submatrix of H_0 in W . Because H_0 has exactly one general diagonal of \star 's, then W_0 has exactly one general diagonal of nonzero entries, which implies $\det(W_0) \neq 0$. This proves that $\text{rank}(W) \geq \text{rank}(W_0) = k$. By picking the largest such H_0 , we deduce that $\text{rank}(H) \geq \text{vrk}(H)$. Since W was an arbitrary algebraic, this proves that $\text{rk}_{\mathbb{F}}(H) \geq \text{vrk}(H)$. \square

2.3 Combinatorial properties of visible rank

In this subsection, we will be proving some properties about visible rank. In particular, we will show that any visibly independent stencil M is permutationally equivalent to an upper triangular stencil, and from this observation, we will be able to upper bound the visible rank by the largest rectangle of zeros in the stencil, which will be our main tool in our constructions of q -LCC and t -DRGP stencils. We also show an upper bound on the rank of a stencil by the maximum number of zeros in each row.

Given two stencils $H_1, H_2 \in \{0, \star\}^{m \times n}$, we say that H_1 is *permutationally equivalent* to H_2 if there are permutations $\pi : [m] \rightarrow [m]$ and $\sigma : [n] \rightarrow [n]$ such that $H_1[i, j] = H_2[\pi(i), \sigma(j)]$ for all $i \in [m]$

and $j \in [n]$. For such H_1 and H_2 , we introduce the notation $H_2 = (H_1)_{\pi, \sigma}$ to say that H_2 is obtained from H_1 by permuting the rows with the permutation π and the columns by the permutation σ (We remark that row permutations commute with column permutations).

Lemma 6. *Let $M \in \{0, \star\}^{n \times n}$ be visibly full rank. Then there exists permutations π and σ on $[n]$ such that $N := M_{\pi, \sigma}$ is an upper triangular stencil. That is $N[i, i] = \star$ and $N[i, j] = 0$ for all $i, j \in [n]$ with $i > j$.*

Proof. We first prove the following claim.

Claim 1. *Let $M \in \{0, \star\}^{n \times n}$ be visibly full rank. Then there exists a row in M with exactly one star.*

Proof: Assume (for the sake of a contradiction) that such a row doesn't exist. Index the rows of M by $R = \{r_1, \dots, r_n\}$ and the columns by $C = \{c_1, \dots, c_n\}$. Let $G = (R, C, S)$ be a bipartite graph on the rows and columns of M with edges S , where S is the set of \star 's in M . Because M is visibly independent, G has a unique perfect matching. Moreover, by our initial assumption, $d_G(v) \geq 2$ for all $v \in L$.

Color the edges of the unique matching of G red and all remaining edges of G blue. Let $P \subseteq S$ be the maximum alternating path in G . Since every vertex in G is incident to a red edge, then P has red edges at both of its ends. Let v_1 be a vertex at one of the ends of P . Since $d_G(v_1) \geq 2$ and v_1 is incident to only one red edge, then v_1 is incident to a blue edge, but because our path is maximal, then this blue edge (v_1, v_2) must satisfy $v_2 \in P$. This implies that our graph G has an alternating cycle C with red edges R_C and blue edges B_C . Notice that both R_C and B_C match the same vertex sets. Since M is a matching, so is $(M \setminus R_C) \cup B_C$, but that's a contradiction by uniqueness of M . ■

We proceed by induction on n . The base case $n = 1$ is immediate to see. As for the induction step, index the rows and columns of M as $\{1, \dots, n\}$. We know by Claim 1 that there exists a row i in M which has exactly a single \star in column j . Let M' be the $n \times n$ matrix produced from M by transposing row i with row n and column j with column n . Because M is visibly independent, so is M' . Moreover, the n 'th row of M' has exactly one \star at column n , meaning that any general diagonal in M' must contain $M[n, n]$. Thus the $(n - 1) \times (n - 1)$ minor M'_0 , which is obtained by deleting row n and column n of M' , has the same number of general diagonals as M' . This proves that M'_0 is visibly independent. By our induction hypothesis, we can permute the rows and columns of M'_0 to make it upper triangular, and thus the same holds for M' . Since M' is a permutation of M , we conclude that M 's rows and columns can be permuted to make it upper triangular. This proves our induction step. □

Thus we can characterize all visibly full rank matrices, and that help us obtain the following upper bound on the visible rank.

Lemma 7. *Given an $m \times n$ stencil H , if there are natural numbers a, b such that H has no $a \times b$ sub-stencil of zeros, then we have $\text{vrk}(H) < a + b$.*

Proof. Assume (for the sake of a contradiction) that H has a $(a + b) \times (a + b)$ square sub-stencil H_0 that is visibly independent. By Lemma 6, we know that H_0 is permutationally equivalent to an $(a + b) \times (a + b)$ upper triangle. Since such triangle has a $a \times b$ sub-stencil of zeros, we arrive to a contradiction. □

We also provide an upper bound on the rank of a stencil by the maximum number of zeros in each rows.

Proposition 8. *For any $m \times n$ stencil H . If each row of H has at most d zeros, then we have that $\text{rk}_{\mathbb{F}}(H) \leq d + 1$ for all fields \mathbb{F} such that $|\mathbb{F}| \geq n$.*

Proof. Pick a field $|\mathbb{F}| \geq n$. Label the columns of H by pairwise distinct entries $a_1, \dots, a_n \in \mathbb{F}$. For row i , let the columns that are zero along row i be $Z_i \subseteq \{a_1, \dots, a_n\}$. Consider the polynomial $p_i(x) := \prod_{a \in Z_i} (x - a)$. Notice that p_i evaluates to zero on Z_i . On everywhere else, it evaluates to a nonzero value. Thus the matrix $E \in \mathbb{F}^{m \times n}$ defined by $E_{ij} = p_i(a_j)$ is an \mathbb{F} -witness of H . Moreover, since $|Z_i| \leq d$ for each $i \in [m]$, then we know that the monomials $\{1, x, \dots, x^d\}$ span the polynomials $\{p_1, \dots, p_m\}$. This shows that $\text{rank}(E) \leq d + 1$ and thus $\text{rk}_{\mathbb{F}}(H) \leq d + 1$. \square

We remark that the bound $|\mathbb{F}| \geq n$ is crucial for Proposition 8. Consider the stencil $D \in \{0, \star\}^{n \times n}$ that has \star 's everywhere except on the diagonal. Such a stencil has a visible rank of 2, but one can show that its rank over \mathbb{F}_2 is at least $n - 1$.

2.4 A Rank-Nullity Type Theorem Between Stencils and Symmetric Spanoids

In this subsection, we formally setup spanoids and prove a rank-nullity type theorem between symmetric spanoids and stencils.

A *spanoid* \mathcal{S} is a collection of inference rules in the form of pairs (S, i) , which are written in the form $S \rightarrow i$, where $S \subseteq [n]$ and $i \in [n]$. The objective in spanoids is to determine the size of the smallest subset $B \subseteq [n]$ such that one can use the inference rules of \mathcal{S} to obtain all of $[n]$. Spanoids were introduced in [DGGW20] as an abstraction of LCCs, wherein they proved that the spanoid analog of q -LCCs satisfy the upper bound $\tilde{O}(n^{(q-2)/(q-1)})$ on the rank. Moreover, they also showed that there are q -LCC spanoids for which their rank is $\tilde{\Omega}(n^{(q-2)/(q-1)})$.

Let us setup the definitions needed for spanoids. A derivation in \mathcal{S} of $i \in [n]$ from a set $T \subseteq [n]$ is a sequence of sets $T_0 = T, T_1, \dots, T_r$ satisfying $T_j = T_{j-1} \cup \{i_j\}$ for some $i_j \in [n]$, $j \in [r]$, and with $i_r = i$. Further, for every $j \in [r]$, there is a rule (S_{j-1}, i_j) in \mathcal{S} for some $S_{j-1} \subseteq T_{j-1}$. The *span* of a set $T \subseteq [n]$, denoted $\text{span}_{\mathcal{S}}(T)$, is the set of all $i \in [n]$ for which there is a derivation of i from T . The rank of a spanoid, denoted $\text{rank}(\mathcal{S})$, is the size of the smallest set $T \subseteq [n]$ such that $\text{span}_{\mathcal{S}}(T) = [n]$. Finally, we define symmetric spanoids below.

Definition 5 (Symmetric Spanoids). *A spanoid \mathcal{S} over $[n]$ is a symmetric spanoid if there are a collection of sets $\{S_1, \dots, S_m\}$ so that the inference rules of \mathcal{S} are of the form $S_j \setminus \{i\} \rightarrow \{i\}$ for any $i \in S_j$ and $j \in [m]$.*

Now we may proceed to prove our theorem that relates the rank of symmetric spanoids with the visible rank of an associated stencil.

Theorem 9. *For any symmetric spanoid \mathcal{S} over $[n]$ with m sets, there exists a canonical stencil H of size $m \times n$ such that for any collection of columns $C \subseteq [n]$ in H , they are visibly independent if and only if $\text{span}_{\mathcal{S}}([n] \setminus C) = [n]$. Moreover, we have $\text{vrk}(H) + \text{rank}(\mathcal{S}) = n$.*

Proof. Define $H[i, j] = \star$ if $j \in S_i$ and zero otherwise. We claim that such H satisfies the conditions. Indeed, suppose that the columns $C = \{c_1, \dots, c_k\}$ are visibly independent. Then that means there are rows r_1, \dots, r_k so that the $k \times k$ sub-stencil formed by these columns and rows is visibly full rank. Denote this matrix as H_C . By Lemma 6, we can find permutations π, σ over $[k]$ such that the matrix $H'_C := H_{\pi, \sigma}$ is upper triangular. In terms of spanoids, that means $c_{\sigma(i)} \in S_{r_{\pi(i)}}$

and $S_{r_{\pi(i)}} \subseteq [n] \setminus \{c_{\sigma(1)}, \dots, c_{\sigma(i-1)}\}$ for all $i \in [k]$. We can rewrite the last set containment as $S_{r_{\pi(i)}} \subseteq ([n] \setminus C) \cup \{c_{\sigma(i)}, \dots, c_{\sigma(k)}\}$. Thus we apply the inference rules $S_{r_{\pi(k)}} \setminus c_{\sigma(k)} \rightarrow c_{\sigma(k)}$, $S_{r_{\pi(k-1)}} \setminus c_{\sigma(k-1)} \rightarrow c_{\sigma(k-1)}$, \dots , $S_{r_{\pi(1)}} \setminus c_{\sigma(k)} \rightarrow c_{\sigma(1)}$ in that order with the set $[n] \setminus C$, to deduce that the set $[n] \setminus C$ spans the set $[n]$ in \mathcal{S} . Thus $\text{span}_{\mathcal{S}}([n] \setminus C) = [n]$.

Now, suppose that $\text{span}_{\mathcal{S}}([n] \setminus C) = [n]$. Then that means that we can find sets S_{i_1}, \dots, S_{i_k} and a permutation σ over $[k]$ such that $c_{\sigma(j)} \in S_{i_j}$ and we can apply the inference rules $S_{i_1} \setminus \{c_{\sigma(1)}\} \rightarrow c_{\sigma(1)}$, \dots , $S_{i_k} \setminus \{c_{\sigma(k)}\} \rightarrow c_{\sigma(k)}$ in that order. That implies then that $S_{i_j} \subseteq ([n] \setminus C) \cup \{c_{\sigma(1)}, \dots, c_{\sigma(j)}\}$. In terms of the stencil H , that means $H[i_j, \sigma(j)] = \star$ and $H[i_\ell, \sigma(j)] = 0$ for $\ell < j$. Thus the $k \times k$ sub-stencil H' that is restricted to the columns $c_{\sigma(1)}, \dots, c_{\sigma(k)}$ and rows i_1, \dots, i_k in that order forms a lower triangular stencil, which is permutationally equivalent to an upper triangular stencil. Thus we deduce that the set of columns C is visibly independent.

Now, for any set $S \subseteq [n]$ such that $\text{span}_{\mathcal{S}}(S) = [n]$, we know that $[n] \setminus S$ is visibly independent in H . Thus $n - |S| \leq \text{vrk}(H)$. Since this holds for any such set S , then we deduce that $\text{rank}(\mathcal{S}) + \text{vrk}(H) \geq n$. On the other hand, for any collection of columns C in H that is visibly independent, we know that $\text{span}_{\mathcal{S}}([n] \setminus C) = [n]$. This implies $n - |C| \geq \text{rank}(\mathcal{S})$. Since this holds for any visibly independent set of columns C in H , then we find that $n \geq \text{rank}(\mathcal{S}) + \text{vrk}(H)$. Hence $\text{rank}(\mathcal{S}) + \text{vrk}(H) = n$. \square

3 Constructing q -LCC Stencils

In this section, we define q -LCC stencils and construct q -LCC stencils whose visible rank achieves the known lower bounds up to polylog factors.

Definition 6 (q -LCC Stencils). For $\delta > 0$, a $\delta n^2 \times n$ stencil H whose rows are labelled by $[n] \times [\delta n]$ is said to be a q -LCC stencil if $H[(i, j), i] = \star$ for all $(i, j) \in [n] \times [\delta n]$. Moreover, for every $k \in [n] \setminus \{i\}$, the collection of entries $\{M[(i, 1), k], \dots, M[(i, t), k]\}$ has at most one star, and the number of \star 's in each row of H is at most $q + 1$.

Now we proceed to prove our main theorem for this section.

Theorem 10. For $q \geq 3$, there exists a q -LCC stencil M whose visible rank is at most $n - \tilde{\Omega}(n^{(q-2)/(q-1)})$.

Proof. Enumerate the columns by $\{1, \dots, n\}$. For $(i, j) \in [n] \times [\delta n]$. Define $r_j^i := \{k \in [n] : H[(i, j), k] = \star\}$ to be the support of row (i, j) , and let $G_i := \{r_j^i : j \in [\delta n]\}$ be the δn groups for column i . We shall show that by picking the groups G_i uniformly at random, the visible rank will at most be $n - n^{\frac{q-2}{q-1}/\log n}$ with high probability.

Consider natural numbers $s > k$ smaller than n that we shall pick later. We shall show that there aren't any $(s - k) \times (n - s)$ sub-stencils in H that are all zeros with high probability, which will yield an upper bound of $n - k$ on the visible rank via Lemma 7. Consider a collection $C \subseteq [n]$ of the columns of size s . Consider a collection of $s - k$ rows $R \subseteq \cup_{i=1}^n G_i$. We will compute the probability that all the rows in R belong to C . Of course, if $R \cap G_i \neq \emptyset$ for $i \notin C$, then C cannot contain all the row in R as every row in G_i contains i . Thus suppose that $R \subseteq \cup_{i \in C} G_i$. Enumerate

$R \cap G_i = \{r_1^i, \dots, r_{a_i}^i\}$. Then by the chain rule and the definition of G_i , we have that

$$\begin{aligned} \Pr \left[\bigwedge_{r \in R \cap G_i} r \subseteq C \right] &= \prod_{j=1}^{a_i} \Pr \left[r_j^i \subseteq C \mid r_k^i \subseteq C \text{ for } k = 1, \dots, j-1 \right] \\ &= \prod_{j=1}^{a_i} \Pr \left[r_j^i \setminus \{i\} \subseteq C \setminus \{r_1^i, \dots, r_{j-1}^i\} \mid r_k^i \subseteq C \text{ for } k = 1, \dots, j-1 \right] \\ &= \prod_{j=1}^{a_i} \frac{\binom{s-(j-1)q-1}{q}}{\binom{n-(j-1)q-1}{q}} \leq \prod_{j=1}^{a_i} \frac{\binom{s-1}{q}}{\binom{n-1}{q}} = \prod_{r \in R \cap G_i} \Pr[r \subseteq C] \end{aligned}$$

And so by the independence of the G_i 's, we find that

$$\Pr \left[\bigwedge_{r \in R} r \subseteq C \right] = \prod_{i=1}^n \Pr \left[\bigwedge_{r \in R \cap G_i} r \subseteq C \right] \leq \prod_{i=1}^n \prod_{r \in R \cap G_i} \Pr[r \subseteq C] = \prod_{r \in R} \Pr[r \subseteq C] = \left(\frac{\binom{s-1}{q}}{\binom{n-1}{q}} \right)^{s-k} \leq \left(\frac{s}{n} \right)^{q(s-k)}$$

Thus by a Union Bound over all such R , we find that

$$\Pr \left[\exists R \subseteq \cup_{i \in C} G_i, |R| = s-k : \bigwedge_{r \in R} r \subseteq C \right] \leq \binom{\delta n s}{s-k} \left(\frac{s}{n} \right)^{q(s-k)}$$

Thus by applying another Union Bound over all possible collections of columns C of size s and use the bound $\binom{a}{b} \leq \left(\frac{ea}{b} \right)^b$, we deduce that

$$\Pr \left[\exists C \subseteq [n], R \subseteq \cup_{i \in C} G_i, |C| = s, |R| = s-k : \bigwedge_{r \in R} r \subseteq C \right]$$

is at most

$$\begin{aligned} \binom{n}{s} \binom{\delta n s}{s-k} \left(\frac{s}{n} \right)^{q(s-k)} &\leq \left(\frac{en}{s} \right)^s \left(\frac{e\delta n s}{s-k} \right)^{s-k} \left(\frac{s}{n} \right)^{q(s-k)} \\ &= \left(\frac{en}{s} \right)^k \left(\frac{en}{s} \cdot \frac{e\delta s n}{s-k} \cdot \frac{s^q}{n^q} \right)^{s-k} = \left(\frac{en}{s} \right)^k \left(\frac{e^2 \delta s^{q-1}}{\left(1 - \frac{k}{s}\right) n^{q-2}} \right)^{s-k} \end{aligned}$$

Making the choice $s = n^{\frac{q-2}{q-1}}$, we find that

$$\left(\frac{en}{s} \right)^k \left(\frac{e^2 \delta s^{q-1}}{\left(1 - \frac{k}{s}\right) n^{q-2}} \right)^{s-k} = \left(en^{1/(q-1)} \right)^k \left(\frac{e^2 \delta}{\left(1 - \frac{k}{s}\right)} \right)^{s-k}$$

Thus if $k = s/\log n = n^{\frac{q-2}{q-1}}/\log n$, then for small enough δ , the quantity above becomes $\exp(-\Omega(n))$. Thus we can find a q -LCC stencil M such that no $s-k$ rows whose support is entirely contained within s columns. That is equivalent to saying that there is no $(s-k) \times (n-s)$ sub-stencil that is all zeros. By Lemma 7, we therefore conclude that $\text{vrk}(M) < n-k = n - \Omega\left(n^{\frac{q-2}{q-1}}/\log n\right)$. \square

4 t -DRGP Stencils

We now define the stencils that capture the requirement of each codeword symbol having t disjoint recovery groups.

Definition 7 (t -DRGP Stencils). A $tn \times n$ stencil H whose rows are labelled by $[n] \times [t]$ is said to be a t -DRGP stencil if $H[(i, j), i] = \star$ for all $(i, j) \in [n] \times [t]$. Moreover, for every $k \in [n] \setminus \{i\}$, the collection of entries $\{M[(i, 1), k], \dots, M[(i, t), k]\}$ has at most one star.

Now we proceed to prove our main theorem for this section.

Theorem 11. For any fixed natural number $t \geq 2$, there exists a t -DRGP stencil H satisfying $\text{vrk}(H) \leq O(t^2 \log n)$.

Proof. Consider a random t -DRGP stencil H as follows: define the set of entries $S_{i,j} := \{(i, s), j \mid s \in [t]\}$. Set for each $i \neq j \in [n]$, set all of the entries $S_{i,i}$ to be \star 's, and uniformly sample an entry from $S_{i,j}$ to be a \star while everything else in $S_{i,j}$ is set to be zero.

We will show that $\text{vrk}(H) \leq c_1 \log n$ occurs with high probability. Indeed, fix $k \in \mathbb{N}$. Given any square sub-stencil H_0 of H of size k , we have by Lemma 6 that if H_0 is visibly independent, then we must have at least $\binom{k}{2}$ zeros. Let this set of entries be $Z \subseteq ([n] \times [t]) \times [n]$. Since Z must all be zeros, then we deduce that $Z \subseteq \cup_{i \neq j} S_{i,j}$. Since each $S_{i,j}$ has size t , then Z has at least $\binom{k}{2}/t$ entries, each of which belongs to an $S_{i,j}$ that is different than the other. Let That is, for each $i \neq j \in [n]$ such that $Z \cap S_{i,j} \neq \emptyset$, arbitrarily pick an entry $e \in Z \cap S_{i,j}$, and let T be those set of entries. Then we know that $T \geq \binom{k}{2}/t$. Moreover, the events that the entries in T are zero are all independent, with each having a chance of at most $1 - 1/t$ of being zero. Thus the chance that Z is all-zeros is at most $(1 - 1/t)^{\binom{k}{2}/t}$. By a Union Bound over all possible such Z 's, which is enumerated over all $(k!)^2$ different permutations of the rows and columns, we deduce that

$$\Pr [H_0 \text{ is visibly independent}] \leq (k!)^2 \left(1 - \frac{1}{t}\right)^{\frac{\binom{k}{2}}{t}} \leq \left(k^2 \left(1 - \frac{1}{t}\right)^{\frac{k-1}{2t}}\right)^k.$$

And by applying another Union Bound over all such H_0 , we find that

$$\Pr [\text{vrk}(H) \geq k] \leq \binom{tn}{k} \binom{n}{k} \left(k^2 \left(1 - \frac{1}{t}\right)^{\frac{k-1}{2t}}\right)^k \leq \left(k^2 t n^2 \left(1 - \frac{1}{t}\right)^{\frac{k-1}{2t}}\right)^k.$$

Picking $k = 6t^2 \ln n + 1$ makes the right hand side less than 1 for large enough n . □

5 Tensor Products

In this section, we will be introducing a tensor product operation on stencils and explore its properties. Given the natural tensor product $A \otimes B$ for matrices A and B , notice that the support of $A \otimes B$ is determined completely by the support of the matrices A and B . As a consequence of this observation, we will be able to define the stencil of $A \otimes B$ based solely on the stencils of A and B . This leads us to our definition of a tensor product over stencils.

Definition 8 (Tensor product). *Given an $A_1 \times B_1$ stencil H_1 and an $A_2 \times B_2$ stencil H_2 , let $H_1 \otimes H_2$ be a $(A_1 \times A_2) \times (B_1 \times B_2)$ stencil such that*

$$(H_1 \otimes H_2)[(a_1, a_2), (b_1, b_2)] = \begin{cases} \star & \text{if } H_1[a_1, b_1] \text{ and } H_2[a_2, b_2] \text{ both equal } \star, \\ 0 & \text{if at least one of } H_1[a_1, b_1] \text{ and } H_2[a_2, b_2] \text{ equals } 0. \end{cases}$$

We remark that our tensor product follows similar properties as the natural tensor product for matrices, such as associativity and non-commutativity.

5.1 Algebraic witnesses of tensor products

In this subsection, we will be proving that any algebraic witnesses of the stencils H_1 and H_2 is also an algebraic witness of $H_1 \otimes H_2$. This will therefore show us that the \mathbb{F} -rank is a *sub-multiplicative* function with respect to the tensor product.

Proposition 12. *Let M and N be matrices over a field \mathbb{F} who are \mathbb{F} -witnesses to stencils H_1, H_2 , respectively. Then $M \otimes N$ is an \mathbb{F} -witness of $H_1 \otimes H_2$.*

Proof. For every entry in $H_1 \otimes H_2$, we know that $(H_1 \otimes H_2)[(i_1, i_2), (j_1, j_2)]$ is a \star if and only if $H_1[i_1, j_1]$ and $H_2[i_2, j_2]$ are both \star 's. This holds if and only if $M_{i_1 j_1}$ and $N_{i_2 j_2}$ are both nonzero. Because $(M \otimes N)_{(i_1, i_2), (j_1, j_2)} = M_{i_1 j_1} N_{i_2 j_2}$, then the entry $(M \otimes N)(i_1, i_2), (j_1, j_2)$ is nonzero if and only if $M_{i_1 j_1}$ and $N_{i_2 j_2}$ are both nonzero. Thus $M \otimes N$ is an \mathbb{F} -witness of $H_1 \otimes H_2$. \square

By applying Proposition 12 on the \mathbb{F} -witnesses of H_1 and H_2 with the smallest ranks, we deduce the following corollary.

Corollary 13. *For a field \mathbb{F} , we have the inequality $rk_{\mathbb{F}}(H_1)rk_{\mathbb{F}}(H_2) \geq rk_{\mathbb{F}}(H_1 \otimes H_2)$*

5.2 Visible rank and tensor products

In this subsection, we will show that the tensor product of two visibly full rank stencils is also visibly full rank. This will therefore show us that the visible rank is *super-multiplicative* with respect to the tensor product. We will also show an upper bound on the visible rank of the tensor product with respect to the visible rank of one of the stencils.

Proposition 14. *Given visibly independent matrices A and B of size n , their tensor $A \otimes B$ is also visibly independent.*

Proof. By Lemma 6, we know that there are permutations $\pi_A, \sigma_A, \pi_B, \sigma_B$ on $[n]$ such that the stencils $A_0 = (A)_{\pi_A, \sigma_A}$ and $B_0 = (B)_{\pi_B, \sigma_B}$ are both upper triangular stencils. Moreover, we see that $A_0 \otimes B_0 = (A \otimes B)_{(\pi_A, \pi_B), (\sigma_A, \sigma_B)}$. Therefore, it suffices for us to show that $A_0 \otimes B_0$ is an upper triangular stencil.

Consider the lexicographical ordering on $[n] \times [n]$. When $(i_1, i_2) > (j_1, j_2)$, then we know that one of the inequalities $i_1 > j_1$ and $i_2 > j_2$ must hold, which means one of $A_0[i_1, j_1]$ or $B_0[i_2, j_2]$ must be a zero. This proves that $(A_0 \otimes B_0)[(i_1, i_2), (j_1, j_2)] = 0$ whenever $(i_1, i_2) > (j_1, j_2)$. As for when $(i_1, i_2) = (j_1, j_2)$, then we immediately know that $(A_0 \otimes B_0)[(i_1, i_2), (i_1, i_2)] = \star$ as $A_0[i_1, i_1] = B_0[i_2, i_2] = \star$. Hence $A_0 \otimes B_0$ is an upper triangular stencil with respect to the lexicographical ordering. \square

Given stencils H_1 and H_2 , we know by Proposition 14 that the tensor product of any of their visibly full rank sub-stencils will also be visibly full rank in $H_1 \otimes H_2$. This yields us the following corollary.

Corollary 15. *For stencils H_1 and H_2 , We have the inequality $\text{vrk}(H_1 \otimes H_2) \geq \text{vrk}(H_1)\text{vrk}(H_2)$.*

Lastly, we end this subsection with an upper bound on the visible rank of $H_1 \otimes H_2$.

Proposition 16. *For stencils H_1 and H_2 of sizes $m_1 \times n_1$ and $m_2 \times n_2$, respectively, We have the inequality $\text{vrk}(H_1 \otimes H_2) \leq \text{vrk}(H_1)n_2$.*

Proof. Consider a visibly full rank sub-stencil M in $H_1 \otimes H_2$ of size $k \times k$. By Lemma 6, we can find permutations π and σ on the rows and columns of M , respectively, so that the stencil $M_0 := M_{\pi, \sigma}$ is an upper triangular matrix. Let the columns of M_0 be indexed left to right as $(a_1, b_1), \dots, (a_k, b_k)$. Let the rows of M_0 be index from top to bottom by $(c_1, d_1), \dots, (c_k, d_k)$. For $b \in [n_2]$, define $I_b := \{\ell \in [k] : b_\ell = b\}$. Define $b_{max} := \arg \max_{b \in [n_2]} \{|I_b|\}$, and let $s := |I_{b_{max}}|$. Then we have that $s \geq k/n_2$. Write $I_{b_{max}} = \{i_1 < \dots < i_s\}$.

Now, consider the $s \times s$ sub-stencil M_1 in M_0 with columns $(a_{i_1}, b_{i_1}), \dots, (a_{i_s}, b_{i_s})$ and rows $(c_{i_1}, d_{i_1}), \dots, (c_{i_s}, d_{i_s})$. Since M_0 is an upper triangular stencil, then so is M_1 . Moreover, since $b_{max} = b_{i_1} = \dots = b_{i_s}$, then the columns of M_1 are $(a_{i_1}, b_{max}), \dots, (a_{i_s}, b_{max})$. We claim that the $s \times s$ sub-stencil H'_1 in H_1 formed by the rows c_{i_1}, \dots, c_{i_s} and the columns a_{i_1}, \dots, a_{i_s} is upper triangular, which will give us $\text{vrk}(H_1) \geq s \geq k/n_2$. Because this holds for every full rank sub-stencil M of $H_1 \otimes H_2$, we can then conclude $\text{vrk}(H_1) \geq \text{vrk}(H_1 \otimes H_2)/n_2$. So it suffices to show that the sub-stencil H'_1 of H_1 is upper triangular.

Now, since M_1 is upper triangular, we know that $M_1[(a_i, b), (c_i, d_i)] = \star$, which implies $H_1[a_i, c_i] = H_2[b, d_i] = \star$. Moreover, we also know for $i > j$ that $M_1[(a_j, b), (c_i, d_i)] = 0$, which means that one of $H_1[a_j, c_i]$ or $H_2[b, d_i]$ is a zero, but since $H_2[b, d_i] = \star$, then we deduce that $H_1[a_j, c_i] = 0$. Thus we conclude that the $s \times s$ sub-stencil H'_1 in H_1 formed by the rows c_{i_1}, \dots, c_{i_s} and the columns a_{i_1}, \dots, a_{i_s} is upper triangular. \square

5.3 Visible rank of the tensor powers

Naturally, one would be interested in tensoring a stencil H with itself several times and examine such a stencil.

Definition 9 (Tensor power). *Given an $m \times n$ stencil H , the k 'th tensor of H is the $m^k \times n^k$ stencil $H^{\otimes k}$ defined as*

$$H^{\otimes k} := \underbrace{H \otimes H \otimes \dots \otimes H}_{k \text{ times}}$$

By combining all the results from the previous subsections, we obtain the following corollary.

Corollary 17. *For a natural number k and an $m \times n$ stencil H , we have the inequality*

$$\text{rk}_{\mathbb{F}}(H) \geq \text{rk}_{\mathbb{F}}(H^{\otimes k})^{1/k} \geq \text{vrk}(H^{\otimes k})^{1/k} \geq \text{vrk}(H)$$

Moreover, we also have the inequality $\text{vrk}(H^{\otimes k}) \leq n^{k-1}\text{vrk}(H)$.

From the previous corollary, we can see that by consider the visible rank of the higher tensor powers of a stencil H , one might hope to attain better lower bounds on the \mathbb{F} -rank. Naturally, one would define the highest possible bound achieved through this vein.

Definition 10 (Visible Capacity). *The visible capacity of a stencil H , denoted as $\Upsilon(H)$, is defined as $\Upsilon(H) := \sup_k \text{vrk}(H^{\otimes k})^{1/k}$.*

By Corollary 17, we deduce that $\text{rk}_{\mathbb{F}}(H) \geq \Upsilon(H)$ over any field \mathbb{F} . It is not known to us if there are stencils for which there is a gap between its visible capacity and all its \mathbb{F} -ranks. We leave the discussion of this point to Question 1 in Section 7.

6 Tensor Powers of Stencils for 2-DRGP Codes and q -LCCs

In this section, we will be considering the tensor product in the previous section and use it to analyze the visible rank of the tensor powers of 2-DRGP and q -LCC stencils to see if they might yield better lower bounds on the rank via Corollary 17.

6.1 2-DRGP stencils

In this subsection, we prove that the second tensor power of an arbitrary 2-DRGP stencil has a large visible rank.

Theorem 18. *For any 2-DRGP stencil H , we have $\text{vrk}(H \otimes H) \geq n$.*

Proof. We cite [RV16, Woo16] for the proof of this part. We will follow the notations given in [Woo16] closely. While both proofs show that $\text{rk}_{\mathbb{F}}(H) \geq \sqrt{2n} - O(1)$, we will prove that $\text{rk}_{\mathbb{F}}(H) \geq \sqrt{n}$ by showing that $\text{vrk}(H \otimes H) \geq n$ and then applying Corollary 17. We rewrite their proofs in terms of tensor powers.

Consider the $n \times n$ sub-stencil D in $H \otimes H$ whose columns are $(1, 1), \dots, (n, n)$ and whose rows are $((1, 1), (1, 2)), \dots, ((n, 1), (n, 2))$. We claim that D has \star 's along the diagonal and zero everywhere else, which implies that it is visibly full rank. Indeed, the entry $D[((i, 1), (i, 2)), (j, j)]$ equals \star if and only if both $H[(i, 1), j]$ and $H[(i, 2), j]$ are \star 's. Since H is a 2-DRGP stencil, this happens precisely when $i = j$. Thus D is a diagonal stencil. \square

Thus by Corollary 17, we obtain a lower bound $\text{rk}_{\mathbb{F}}(H) \geq \sqrt{n}$ for any field \mathbb{F} . On the other hand, the best known lower bounds yield $\text{rk}_{\mathbb{F}}(H) \geq \sqrt{2n} - O(1)$, and so one might be interested in achieving this lower bound through the viewpoint of tensor products. In order to improve the lower bound that we have, we first have to translate our proof into linear-algebraic terms.

Given a field \mathbb{F} , suppose that we have an \mathbb{F} -witness A of the 2-DRGP stencil H whose rank is r . Decompose $A = MN$ where M is an $2n \times r$ matrix and N is an $r \times n$ matrix. Denote the i 'th column of N by w_i . Then the proof of Theorem 18 is equivalent to saying that the tensors $\{w_i \otimes w_i\}_{i=1}^n$ are linearly independent. Since they live in a space $\mathbb{F}^r \otimes \mathbb{F}^r$, then we obtain the inequality $r^2 \geq n$, which gives us the same lower bound as we obtained in Theorem 18. Now, if one is more careful about the vector space, one can notice that the tensors $\{w_i \otimes w_i\}_{i=1}^n$ belong to the space of *symmetric* tensors, which has a dimension of $\binom{r+1}{2}$. Thus we obtain the inequality $\binom{r+1}{2} \geq n$, which gives us $r \geq \sqrt{2n} - O(1)$.

6.2 q -LCC stencils

In this subsection, we will show that the visible ranks of the k 'th tensor power of a q -LCC stencil would not improve the current bound of $n - \tilde{\Omega}(n^{(q-2)/(q-1)})$ in Theorem 10 for $k \leq \text{polylog}(n)$. More generally, we will show that the visible rank of small tensor powers could be not significantly bigger than the visible rank for the regime of high-rate stencils.

Proposition 19. *Let H be an $m \times n$ stencil whose visible rank is at most $n - s$. For any fixed natural number k , we have $\text{vrk}(H^{\otimes k})^{1/k} \leq n - \frac{s}{k}$.*

Proof. By Corollary 17 and the inequality $(1 - x)^{1/k} \leq 1 - \frac{x}{k}$ for $x \geq 0$, we have that

$$\text{vrk}(H^{\otimes k})^{1/k} \leq \left(n^{k-1} \text{vrk}(H)\right)^{1/k} \leq \left(n^{k-1}(n - s)\right)^{1/k} = n \left(1 - \frac{s}{n}\right)^{1/k} \leq n \left(1 - \frac{s}{kn}\right) = n - \frac{s}{k}. \quad \square$$

From Proposition 19, we notice that looking at the visible rank of the $n^{o(1)}$ level tensor powers of a q -LCC stencil would not improve the current bounds on q -LCCS by a polynomial factor. We state it more formally in the following corollary.

Corollary 20. *Let H be a q -LCC stencil whose visible rank is at most $n - \tilde{\Omega}(n^{(q-2)/(q-1)})$. For any natural number k , we have $\text{vrk}(H^{\otimes k})^{1/k} \leq n - \tilde{\Omega}(n^{(q-2)/(q-1)})/k$.*

7 Further Directions and Discussion

Stencils provide an initial framework toward combinatorial methods for effectively lower bounding the rank of a matrix. However, we have seen the limitations of the visible rank with 2-DRGP stencils as well as small tensor powers of q -LCC stencils. We leave the reader with questions that remain open about the current framework and possibilities of imposing further restrictions on the model to obtain sharper lower bounds on the rank for restricted cases.

1. While we may have shown that the k 'th tensor power of a q -LCC does not yield better lower bounds for $k \leq n^{o(1)}$, this does not rule out the possibility that the visible capacity might yield better lower bounds. In fact, we do not know if there are any stencils for which the visible capacity does not match the lowest possible rank for the stencil. In other words, does there exist a stencil H such that $\text{rk}_{\mathbb{F}}(H) > \Upsilon(H)$ for every field \mathbb{F} ?
2. Random 2-DRGP patterns have shown an exponential gap between the visible ranks of the first and second tensor powers, but one might be curious to see an exponential separation between the visible ranks of the $(t - 1)$ 'th and t 'th tensor powers. Formally speaking, for every natural number t greater than 1, does there exist a $m \times n$ stencil H and a constant $c > 0$ such that $\text{vrk}(H^{\otimes i}) = O(\log^{ci} n)$ for $i = 1, 2, \dots, t - 1$ while $\text{vrk}(H^{\otimes t}) = \Omega(n)$? Such a phenomena holds with Shannon capacity [AL06].
3. In this paper, we shown a polynomial gap between $\text{rk}_{\mathbb{F}}(H)$ and $\text{vrk}(H)$ by proving that there are 2-DRGP stencils H with $\text{vrk}(H) = O(\log n)$ and $\text{rk}_{\mathbb{F}}(H) = \Omega(\sqrt{n})$. On the other hand, can there also be a similar polynomial gap with the quantities $n - \text{rk}_{\mathbb{F}}(H)$ and $n - \text{vrk}(H)$? From Proposition 19, we have seen that the visible ranks of the k 'th tensor power for $k \leq n^{o(1)}$ would not suffice to show this polynomial gap. Nonetheless, it still leaves the possibility of using

the visible capacity to showing this polynomial gap, but we do not know of any methods that can lower bound the visible capacity other than the visible ranks of finite tensor powers. Note that this question is the symmetric spanoid version of Question 2 posed in [DGGW20].

4. One natural restriction is to consider attaining lower bounds for $\text{rk}_{\mathbb{F}}(H)$ over all fields \mathbb{F} such that $|\mathbb{F}| \leq M$ for some $M \in \mathbb{N}$. In such cases, methods such as those used in Proposition 8 would not work. More generally, suppose that we are interested in attaining lower bounds for k -color stencils, a stencil wherein any algebraic witness of it can only use at most k distinct nonzero entries. Can one consider better combinatorial proxies for the algebraic rank of k -color stencils (particularly constant-colored stencils)?

It is possible to attain asymptotically better lower bounds on the algebraic rank for colored stencils. Indeed, consider a k -color stencil $H \in \{0, \star\}^{n \times n}$ where $H[i, i] = 0$ and $H[i, j] = 1$ for $i \neq j$. We know that $\text{vrk}(H) = 2$, but since H is a k -color stencil, then one can show by a polynomial argument (which we leave for the reader to try) that any algebraic witness of H has a rank of at least $\Omega_k(n^{1/k})$.

5. One component that stencils do not capture is the distance of a code, but that's partly due to the distance of a code being dependent on the field size. If we consider k -color stencils, can one find combinatorial proxies that upper and lower bound the distance of any algebraic witness of the stencil?
6. Can one explicitly construct an $n \times n$ stencil H whose visible rank is robust to small entry corruptions? Any algebraic witness of H would automatically be a rigid matrix.

Finally, our aim in introducing the visible rank is to study combinatorial proxies for the rank of a matrix. Are there other combinatorial quantities associated with a matrix that could yield more effective lower bounds on its rank?

References

- [AL06] Noga Alon and Eyal Lubetzky. The shannon capacity of a graph and the independence numbers of its powers. *IEEE Transactions on Information Theory*, 52(5):2172–2176, 2006.
- [BBJK11] Ziv Bar-Yossef, Yitzhak Birk, T. S. Jayram, and Tomer Kol. Index coding with side information. *IEEE Trans. Inf. Theory*, 57(3):1479–1494, 2011.
- [BDYW11] Boaz Barak, Zeev Dvir, Amir Yehudayoff, and Avi Wigderson. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 519–528, 2011.
- [DGGW20] Zeev Dvir, Sivakanth Gopi, Yuzhou Gu, and Avi Wigderson. Spanoids - an abstraction of spanning structures, and a barrier for LCCs. *SIAM J. Comput.*, 49(3):465–496, 2020.
- [Dvi12] Zeev Dvir. Incidence theorems and their applications. *arXiv preprint arXiv:1208.5073*, 2012.

- [FGW17] S. Luna Frank-Fischer, Venkatesan Guruswami, and Mary Wootters. Locality via partially lifted codes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, volume 81 of *LIPICs*, pages 43:1–43:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.
- [FJJ18] Mathew C Francis, Dalu Jacob, and Satyabrata Jana. Uniquely restricted matchings in interval graphs. *SIAM Journal on Discrete Mathematics*, 32(1):148–172, 2018.
- [FVY15] Arman Fazeli, Alexander Vardy, and Eitan Yaakobi. Codes for distributed PIR with low storage overhead. In *IEEE International Symposium on Information Theory*, pages 2852–2856, 2015.
- [GHL01] Martin Charles Golumbic, Tirza Hirst, and Moshe Lewenstein. Uniquely restricted matchings. *Algorithmica*, 31(2):139–154, 2001.
- [GHSY12] Parikshit Gopalan, Cheng Huang, Huseyin Simitci, and Sergey Yekhanin. On the locality of codeword symbols. *IEEE Transactions on Information theory*, 58(11):6925–6934, 2012.
- [GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the Innovations in Theoretical Computer Science Conference*, pages 529–540, 2013.
- [GKT01] Harold N Gabow, Haim Kaplan, and Robert E Tarjan. Unique maximum matching algorithms. *Journal of Algorithms*, 40(2):159–183, 2001.
- [Gop18] Sivakanth Gopi. *Locality in coding theory*. PhD thesis, Princeton University, 2018.
- [GRW18] Alexander Golovnev, Oded Regev, and Omri Weinstein. The minrank of random graphs. *IEEE Trans. Inf. Theory*, 64(11):6990–6995, 2018.
- [Hae79] Willem H. Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Trans. Inf. Theory*, 25(2):231–232, 1979.
- [HMT06] Thanh Minh Hoang, Meena Mahajan, and Thomas Thierauf. On the bipartite unique perfect matching problem. In *International Colloquium on Automata, Languages, and Programming*, pages 453–464. Springer, 2006.
- [HSX⁺12] Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li, and Sergey Yekhanin. Erasure coding in windows azure storage. In *USENIX Annual Technical Conference*, 2012.
- [IS20] Eran Iceland and Alex Samorodnitsky. On coset leader graphs of structured linear codes. *Discrete and Computational Geometry*, 63:560–576, 2020.
- [KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, pages 80–86, 2000.
- [Lov89] László Lovász. Communication complexity: A survey. Technical report, Princeton University TR-204-89, February 1989. University of Zurich, Department of Informatics.

- [LW21] Ray Li and Mary Wootters. Lifted multiplicity codes and the disjoint repair group property. *IEEE Trans. Inf. Theory*, 67(2):716–725, 2021.
- [Mis11] Sounaka Mishra. On the maximum uniquely restricted matching for bipartite graphs. *Electronic Notes in Discrete Mathematics*, 37:345–350, 2011.
- [Ram20] C. Ramya. Recent progress on matrix rigidity—a survey. *arXiv preprint arXiv:2009.09460*, 2020.
- [RV16] Sankeerth Rao and Alexander Vardy. Lower bound on the redundancy of PIR codes. *arXiv preprint arXiv:1605.01869*, 2016.
- [SAP⁺13] Maheswaran Sathiamoorthy, Megasthenis Asteris, Dimitris Papailiopoulos, Alexandros Dimakis, Ramkumar Vadali, Scott Chen, and Dhruva Borthakur. XORing elephants: Novel erasure codes for big data. *Proc. VLDB Endow.*, 6:325–336, 2013.
- [SS96] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.
- [TD13] Maguy Trefois and Jean-Charles Delvenne. Zero forcing sets, constrained matchings and minimum rank. *Linear and Multilinear Algebra*, 2013.
- [Tré16] Maguy Tréfois. *Topics in combinatorial matrix theory*. PhD thesis, PhD thesis, UCL, 2016.
- [Val77] Leslie G Valiant. Graph-theoretic arguments in low-level complexity. In *International Symposium on Mathematical Foundations of Computer Science*, pages 162–176. Springer, 1977.
- [Woo12] David P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *J. Comput. Sci. Technol.*, 27(4):678–686, 2012.
- [Woo16] Mary Wootters. Linear codes with disjoint repair groups. *unpublished manuscript*, 2016.
- [Yek12] Sergey Yekhanin. Locally decodable codes. *Foundations and Trends in Theoretical Computer Science*, 6(3):139–255, 2012.