Lower bounds on the redundancy of linear codes with disjoint repair groups

Sankeerth Rao Karingula
University of California San Diego

Alexander Vardy
University of California San Diego

Mary Wootters
Stanford University

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834487

***Abstract*—An error correcting code exhibits the t-*Disjoint Repair Group Property* (t-DRGP) (for message symbols) if it is possible to recover a single symbol of a codeword (message) in t ways, each from a disjoint set of symbols of the codeword. Codes with the DRGP have found applications in private information retrieval (PIR) and distributed storage, and are related to several notions of locality in coding theory. In this work we prove an impossibility result for codes with the DRGP. We show that the redundancy of any code with the t-DRGP is $\Omega(\sqrt{n})$ for all $t \geq 2$. Our bound is tight, even including the leading constant, for $t = 2$, and is tight up to a constant factor for $t = O(1)$. We also show an analogous result for binary codes with the t-DRGP for message symbols, which has applications to PIR.**

These results first appeared in 2016 and were never published. As our results have not yet been improved upon, and have been referenced by multiple works over the years, we are prompted to publish them now. We hope that publishing these results now will spur more work in the area, and in particular will lead to improved bounds.

***Index Terms*—Information retrieval, Privacy, Coding theory, Distributed storage**

I. Introduction

A. Background and importance

Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a linear code over a finite field $\mathbb{F}$. We are often interested in the *locality* of $\mathcal{C}$. There are many different ways of defining locality, but typically it refers to the ability to obtain a small amount of information about a message or codeword *locally*, by looking at only at a few symbols. One notion of locality that has recently been fruitful in several domains is the *disjoint repair group property* (DRGP), which roughly says that any (codeword or message) symbol should have many disjoint ways of recovering it. While we do not explicitly consider the size of these recovery sets in this work, the requirement that they be disjoint implies that most must be relatively "local;" we discuss the relationship between the DRGP and other notions of locality below.

Below, we formalize the DRGP, first for codeword symbols (Definition 1) and then for message symbols (Definition 2).

Definition 1 (DRGP)**.** *Let $\mathcal{C} \subset \mathbb{F}^n$ be a linear code. We say that $\mathcal{C}$ has the* t-disjoint-repair-group property for s symbols *((t,s)-DRGP) if the following holds. For any given $i \in [s]$, there are vectors $\lambda^{(1)}, \ldots, \lambda^{(t)} \in \mathbb{F}^n$ so that:*

1) *The sets $\mathrm{Supp}(\lambda^{(j)})$ are disjoint from each other for all $j \in [t]$, and are disjoint from $\{i\}$; and*
2) *for all $c \in \mathcal{C}$ and $j \in [t]$ we have $c_i = \sum_{\ell=1}^{n} \lambda_\ell^{(j)} c_\ell$.*

That is, for any $i \in [s]$ and any $c \in \mathcal{C}$, c_i can be recovered in t different ways (other than looking at c_i itself), each of which relies on a disjoint set of indices. The t-DRGP has been studied in many different areas. When t is small, it is related to notions in distributed storage, like *locally repairable codes* (LRCs) with availability.[1] When $t = \Omega(n)$ is large, codes with the t-DRGP are constant-query *locally correctable code* (LCCs). The DRGP is also closely related to *batch codes*. We refer the reader to [11] for a survey of some of these notions.

When we only wish to recover message symbols, rather than any codeword symbol, we can define a version of the t-DRGP for message symbols.

Definition 2 (DRGP for message symbols)**.** *Let $\mathcal{C} \subseteq \mathbb{F}^n$ be a linear code with an encoding map $\mathrm{enc}_\mathcal{C} : \mathbb{F}^k \to \mathbb{F}^n$. We say that $\mathcal{C}$ has the* t-disjoint-repair-group property for s message symbols *((t,s)-DRGP-m) if the following holds. For any given $i \in [s]$, there are vectors $\lambda^{(1)}, \ldots, \lambda^{(t)} \in \mathbb{F}^n$ so that:*

1) *The sets $\mathrm{Supp}(\lambda^{(j)})$ are disjoint for all $j \in [t]$; and*
2) *for all $j \in [t]$ and all messages $m \in \mathbb{F}^k$, we have $m_i = \sum_{\ell=1}^{n} \lambda_\ell^{(j)} c_\ell$, where $c = \mathrm{enc}_\mathcal{C}(m) \in \mathcal{C}$.*

Binary codes with the DRGP for message symbols are also known as *PIR codes* due to their applications to private information retrieval [4]. For large t, ($t = \Omega(n)$), codes with the t-DRGP-m are constant-query *locally decodable codes* (LDCs).

We note that for a systematic linear code of dimension k, the (t,s)-DRGP is the same as the $(t+1,s)$-DRGP-m for any $s \leq k$.[2] However, in general the two notions are different. Similarly, any linear code with the (t,s)-DRGP (for any s) has the $(t+1, \min\{s,k\})$-DRGP-m for some encoding map; thus proving negative results for the DRGP-m is more difficult than proving negative results for the DRGP.

For both the DRGP and the DRGP-m, we are interested in the trade-off between the parameters t and s and the *redundancy* r of the code; if $\mathcal{C} \subset \mathbb{F}^n$ has dimension k, the redundancy is defined as $r = n - k$.

In this paper we present two lower bounds on the redundancy of codes with the DRGP(-m), focusing on the case

MW's work was partially supported by NSF grant DMS-1400558. SRK and AV's work was supported by NSF grants NSF 1405119 and NSF 1719139

[1] We note that typically for LRCs with availability, one is also interested in the *size* of the repair groups, while we are only interested in their being disjoint.

[2] For historical reasons, the use of "t" is off by one between the two definitions.

where t is constant. Our first result applies to codes with the DGRP, and applies to codes over all fields. Our second result applies to codes with the DRGP-m, but only works for binary fields. Both results are tight for constant t. While our bounds hold for any value of t, they are agnostic to t and we suspect that they are not tight for $t = \omega(1)$.

B. Related work

As noted in the abstract, this work first appeared online in 2016 [10], [14], but it was never published. Since then, there has been a great deal of work on codes with the DRGP, which we summarize below. However, almost all of these are positive results (constructions of codes). To our knowledge there have been no major improvements on the negative results presented in this work since it first appeared. We note that there have been new negative results for the related notion of *batch codes*, which we discuss below.

a) Constructions of codes with the DRGP: For $t = 2$, there is a straightforward construction of a code with the $(2, n)$-DRGP with redundancy r, where r is such that $\binom{r+1}{2} = n$ (see Example 1). We will show in Theorem 1 that this is optimal, and that we must have $\binom{r+1}{2} \geq n$. There has been a great deal of work on constructing codes with the DRGP for larger values of t, for example [4]–[8]. Most of these are based on the algebraic notion of *lifting*. The landscape of what is possible is a bit complicated (see [7] for a detailed overview), but over large fields, for $t \leq \sqrt{n}$, the smallest redundancy known is $\Theta(t^{\log_2(3)-1}\sqrt{n})$ [8]. Our bounds apply for $t > 2$, but continue to imply only that $\binom{r+1}{2} \geq n$. Thus, despite much effort, the best possibility results for large t are quite far off from the best impossibility results given by our work, even over large fields.

b) Constructions of codes with the DRGP-m: Motivated by private information retrieval, the work [4] introduced codes with the DRGP-m (calling them *PIR codes*). They showed that, as with the DRGP for $t = 2$, there is a construction (Example 2) of a code with the (t, k)-DRGP-m for $t = 3$ that has with redundancy r for any r so that $\binom{r}{2} \geq k$; our work implies that this is nearly optimal. For larger t, the work [4] gave constructions with redundancy at most $t\sqrt{k}(1 + o(1))$. Since that work, there have been improved constructions [1], [3] constructing good DRGP-m codes, but as with the DRGP, these are not known to be optimal for $t = \omega(1)$.

c) Other impossibility results: To the best of our knowledge, there have been no improvements to the results presented in this paper since they appeared over five years ago. In particular, it has remained open to meaningfully extend our results to larger values of t. However, we mention the recent work [9], which uses similar techniques to one of our proofs (and is in fact inspired by it) to prove stronger impossibility results for the related notion of *batch codes*. We also mention [12], which pre-dates our work and which establishes bounds for the related notion of *LRCs with availability*. These bounds take into account the size of the recovery sets, and thus are not directly comparable to our work. Finally, while the best redundancy for LDCs and LCCs is a major open question, there have been some negative results [2], [13]. However, since the t-DRGP(-m) is related to LDC/LCCs when $t = \Omega(n)$ is very large, while our results are most interesting when $t = O(1)$ is very small, this work is again not comparable to our work (and unfortunately our bounds do not shed any light on the best rate of constant-query LDC/LCCs).

C. Our results

In this section we present our two results. The first result is a general lower bound on the redundancy of (t, s) DRGP correctable codes over any field.

Theorem 1 (Bound for DRGP)**.** *Let $\mathcal{C} \subset \mathbb{F}^n$ be a linear code of length n, dimension k and redundancy $r = n - k$ that has the $(2, s)$-DRGP. Then*

$$\binom{r+1}{2} \geq s.$$

This bound is tight for $s = n$, as the following construction shows.

Example 1 (Optimal construction for $t = 2, s = n$)**.** *Let r be a positive integer and let $n = \binom{r+1}{2}$. Let $H \in \mathbb{F}_2^{(r+1)\times n}$ be the matrix so that $H_{\ell,\{i,j\}} = \mathbf{1}[\ell \in \{i, j\}]$,[3] where we index the columns of H by pairs $\{i, j\}$ so that $i \neq j$. Let $\mathcal{C} = \{c \in \mathbb{F}_2^n : Hc = 0\}$. The $\mathcal{C}$ has the $(2, n)$-DRGP. Notice that in this construction the redundancy is r (it is not $r + 1$ because the rows of H sum to zero and hence are linearly dependent), and the length of the code is $n = \binom{r+1}{2}$.*

To see that the code in Example 1 indeed has the $(2, n)$-DRGP, suppose that we wish to recover a coordinate ℓ, which we associate with a pair $\{i, j\}$. The i'th and j'th row of H have support intersecting only in ℓ, and thus give two disjoint repair groups for ℓ.

Theorem 1 is also nearly tight for $s = k$: the following example gives a construction (due to [4]) with $s = \binom{r}{2}$.

Example 2 (Near-optimal construction for $t = 2, s = k$ [4])**.** *Let r be a positive integer and choose $n = r + \binom{r}{2}$. Let $G \in \mathbb{F}_2^{\binom{r}{2}\times n}$ be the block matrix given by $G = [I_{\binom{r}{2}}|P]$, where $P \in \binom{r}{2} \times r$ is the matrix whose rows are indexed by pairs $\{i, j\}$ for $i, j \in [r]$ so that $P_{\{i,j\},\ell} = \mathbf{1}[\ell \in \{i, j\}]$. Then G is the (systematic) generator matrix of a code with the $(2, k)$-DRGP and the $(3, k)$-DRGP-m. Notice that in this construction the redundancy is r and the dimension of the code is $k = \binom{r}{2}$.*

To see that the code in Example 2 indeed has the $(2, k)$-DRGP (or equivalently the $(3, k)$-DRGP-m, as the code is systematic), suppose that we wish to recover a symbol indexed by some $\ell \in [k]$; as $k = \binom{r}{2}$, we identify ℓ with a tuple $\{i, j\}$. Consider the two parity-checks given by columns i and j of the parity part P of the generator matrix G; call them $p^{(i)}$ and

[3]Here, $\mathbf{1}[\ell \in \{i, j\}]$ is the indicator function that is 1 if $\ell \in \{i, j\}$ and 0 otherwise.

$p^{(j)}$. The supports of these columns intersect in location ℓ and are otherwise disjoint. Thus one can write $e^{(\ell)}$ in two ways as

$$e^{(\ell)} = p^{(i)} - \sum_{h \in \mathrm{Supp}(p^{(i)}) \setminus \{\ell\}} g^{(h)}$$

and as

$$e^{(\ell)} = p^{(j)} - \sum_{h \in \mathrm{Supp}(p^{(j)}) \setminus \{\ell\}} g^{(h)}.$$

where $g^{(h)} = e^{(h)}$ is the h'th column of G. We observe that the sets of column indices used in these two different ways are disjoint (and are disjoint from $\{\ell\}$), and so these provide our two disjoint repair groups for ℓ (in addition to $\{\ell\}$ itself).

As mentioned above, proving negative results for the DRGP-m is harder than for the DRGP. In the following theorem, we provide a similar result that holds for the DRGP-m; however, it only applies to binary codes.

Theorem 2 (Bound for DRGP-m)**.** *Let $\mathcal{C} \subset \mathbb{F}_2^n$ be a binary linear code with length n, dimension k, and redundancy $r = n - k$ that has the $(3, k)$-DRGP-m. Then*

$$\binom{r+1}{2} \geq k.$$

Again, this bound is nearly tight, as Example 2 shows.

Remark 1 (Comparison between Theorems 1 and 2)**.** *Theorem 2 is not comparable to Theorem 1. While Theorem 1 does apply for $s = k$, if the code in question is not systematic, it does not yield an impossibility result for the DRGP-m; thus Theorem 2 is stronger in this respect. On the other hand, Theorem 1 applies to codes over general fields, while Theorem 2 applies only over $\mathbb{F}_2$.*

Next, we observe that both Theorems 1 and 2 hold for any $t \geq 2$. More precisely, as was observed in [4], it is clear that the best redundancy possible for codes with the (t, s)-DRGP(-m) is non-decreasing in t. (Indeed, if a code $\mathcal{C}$ has the (t, s)-DRGP(-m), then it also trivially has the (t', s)-DRGP(-m) for any $t' \leq t$). Thus, we have the following corollary.

Corollary 1. *Theorem 1 holds when "$(2, s)$-DRGP" is replaced by "(t, s)-DRGP" for any $t \geq 2$. Theorem 2 holds when "$(3, k)$-DRGP-m" is replaced by "(t, k)-DRGP-m" for any $t \geq 3$.*

As noted above, Theorems 1 and 2 first appeared online over five years ago [10], [14], to the best of our knowledge they have not been improved; in particular we are not aware of any results that obtain improved bounds for larger t. We hope that one contribution of the current work will be to highlight the open problem of improving our bounds.

Question 1. *Is there some constant $c > 0$ so that the following holds? For any $2 \leq t \leq \sqrt{n}$, for any code $\mathcal{C}$ with the (t, n)-DRGP (or (t, k)-DRGP-m) with redundancy r and length n, we must have $r = \Omega(t^c \sqrt{n})$.*

Current constructions imply that such a c must satisfy $c \leq \log_2(3) - 1$.

D. Overview of the paper

In Section II we will introduce preliminary notation and definitions. In Section III we prove Theorem 1 about the t-DRGP, and in Section IV we prove Theorem 2 on the t-DRGP-m for binary codes.

II. Preliminaries

In this section we define notions we will need going forward. For vectors $x, y \in \mathbb{F}^n$, we use the notion $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$ to denote the dot product between x and y. For an integer t, we use $[t]$ to mean the set $\{1, 2, \ldots, t\}$. We use $e^{(i)}$ to refer to the i'th standard basis vector. (The dimension of $e^{(i)}$ depends on the context, and we will make it clear when we use it).

The following product operation $\circ$ will be useful in the proof of Theorem 2.

Definition 3 (Product operation $\circ$)**.** *Given two binary vectors $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$, we define their product component wise,*

$$u \circ v := (u_1 v_1, u_2 v_2, \ldots, u_n v_n)$$

where $u_1 v_1, u_2 v_2, \ldots, u_n v_n$ are computed in $\mathbb{F}_2$.

Next, we define the square of a set as its square under the $\circ$ operation.

Definition 4 (Square of a set)**.** *Given a set $X \subseteq \mathbb{F}_2^n$, we define the square of X as the set of component-wise products of the elements in X. That is,*

$$X^2 = \{x \circ y : x, y \in X\}.$$

Definition 5 (Span of a set)**.** *We let $\mathrm{Span}(X)$ denote the linear span over $\mathbb{F}_2$ of a set $X \subseteq \mathbb{F}_2^n$.*

The following propositions follow straightforwardly from Definitions 3 and 4 of the operation $\circ$ and of X^2 respectively.

Proposition 1. *For any set $X \subseteq \mathbb{F}_2^n$, $|X^2| \leq \binom{|X|+1}{2}$.*

Proof. If $|X| = r$, then X^2 consists of the $\binom{r}{2}$ vectors $u \circ v = v \circ u$ for some $u \neq v$ in X, along with the r vectors $u \circ u = u$ for some $u \in X$. Some of these vectors may coincide. □

Proposition 2. *If $a, b \in \mathrm{Span}(X)$, then $a \circ b \in \mathrm{Span}(X^2)$.*

Proof. Let $X = \{x_1, x_2, \ldots, x_r\}$. Since $a, b \in \mathrm{Span}(X)$ we can write $a = \sum_i \alpha_i x_i$ and $b = \sum_i \beta_i x_i$ for some binary coefficients $\alpha_1, \alpha_2, \ldots \alpha_r, \beta_1 \beta_2, \ldots, \beta_r$.

Then

$$\begin{aligned} a \circ b &= \left(\sum_{i=1}^r \alpha_i x_i \right) \circ \left(\sum_{j=1}^r \beta_j x_j \right) \\ &= \sum_{i=1}^r \sum_{j=1}^r \alpha_i \beta_j (x_i \circ x_j) \in \mathrm{Span}(X^2). \end{aligned}$$

□

Proposition 3. *Let $u, v_1, v_2, v_3 \in \mathbb{F}_2^n$ such that $v_1 \circ v_2 + v_1 \circ v_3 + v_2 \circ v_3 = 0$, then*

$$(u+v_1)\circ(u+v_2)+(u+v_2)\circ(u+v_3)+(u+v_3)\circ(u+v_1) = u$$

Proof. The proof follows by straightforward verification using distributivity and commutativity of $\circ$. □

III. Proof of Theorem 1: DRGP

Proof. Let $\mathcal{C} \subset \mathbb{F}^n$ be a code of dimension k and redundancy $r = n-k$, as in the statement of Theorem 1. Consider the dual code $\mathcal{C}^\perp$, which is a linear code of dimension r and length n. Let $\omega^{(1)}, \ldots, \omega^{(n)} \in \mathbb{F}^r$ be vectors so that

$$\mathcal{C}^\perp = \left\{ \left(\langle \alpha, \omega^{(1)} \rangle, \langle \alpha, \omega^{(1)} \rangle, \ldots, \langle \alpha, \omega^{(n)} \rangle \right) : \alpha \in \mathbb{F}^r \right\}.$$

(That is, for $i \in [n]$, the $\omega^{(i)}$ are the columns of a parity-check matrix for $\mathcal{C}$). In this language, the $(2, s)$-DRGP correctable can be restated as follows:

For all $i \in [s]$, there exist some $\alpha^{(i)}, \beta^{(i)} \in \mathbb{F}^r$ so that

1. $\langle \alpha^{(i)}, \omega^{(i)} \rangle \cdot \langle \beta^{(i)}, \omega^{(i)} \rangle \neq 0$, and
2. for all $j \neq i$, $\langle \alpha^{(i)}, \omega^{(j)} \rangle \cdot \langle \beta^{(i)}, \omega^{(j)} \rangle = 0$.

Indeed, suppose that $\lambda^{(1)}$ and $\lambda^{(2)}$ are the vectors guaranteed by Definition 1 for $t = 2$. Then for $i \in [s]$ and $j \in \{1, 2\}$, we have $\lambda^{(j)} - e^{(i)} \in \mathcal{C}^\perp$, where $e^{(i)} \in \mathbb{F}^n$ is the i'th standard basis vector. Let $H \in \mathbb{F}^{r \times n}$ be the parity-check matrix for $\mathcal{C}$ with the vectors $\{\omega^{(j)} : j \in [n]\}$ as columns. Then we see the correspondence between the above condition and Definition 1 as follows. For $i \in [s]$, set $\alpha^{(i)} \in \mathbb{F}^r$ so that $\alpha^{(i)} H = \lambda^{(1)} - e^{(i)}$ and set $\beta^{(i)}$ so that $\beta^{(i)} H = \lambda^{(2)} - e^{(i)}$.

Now for $i \in [s]$, construct polynomials $P_i : \mathbb{F}^r \to \mathbb{F}$ as

$$P_i(X_1, \ldots, X_r) = \left(\sum_{j=1}^{r} \alpha_j^{(i)} X_j \right) \cdot \left(\sum_{j=1}^{r} \beta_j^{(i)} X_j \right).$$

Note that the conditions 1 and 2 above implies that

$$P_i(\omega^{(j)}) = \begin{cases} \langle \alpha^{(i)}, \omega^{(i)} \rangle \cdot \langle \beta^{(j)}, \omega^{(j)} \rangle \neq 0, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Thus, the P_i's are linearly independent over $\mathbb{F}$. However, they are spanned by the monomials of degree exactly two in $X_1, \ldots, X_r$. But there are $\binom{r+1}{2}$ of these, and so $s \leq \binom{r+1}{2}$, as claimed. □

IV. Proof of Theorem 2: DRGP for message symbols

We now prove Theorem 2 using Propositions 1, 2, 3. In particular we establish a lower bound on the redundancy of binary linear codes with the (t, k)-DRGP-m for $t \geq 3$.

Proof. Let $\mathcal{C} \subseteq \mathbb{F}_2^n$ be a binary code of dimension k and redundancy $r = n - k$, with the (t, k)-DRGP-m, as in the theorem statement. Let $G \in \mathbb{F}_2^{k \times n}$ be a generator matrix for $\mathcal{C}$ so that for all $i \in [k]$, there are three disjoint sets of columns of G that add up to $e^{(i)} \in \mathbb{F}_2^k$. Notice that such a matrix exists by the definition of the DRGP for message symbols (Definition 2). Let $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ denote the columns of G. Fix $i \in [k]$, and let $R_1, R_2, R_3 \subset [n]$ denote the disjoint sets so

$$e^{(i)} = \sum_{j \in R_1} x^{(j)} = \sum_{j \in R_2} x^{(j)} = \sum_{j \in R_3} x^{(j)}$$

Since G has full column rank, some k columns of G are linearly independent, and we assume without loss of generality that these are the first k columns. Consequently, there exists a non-singular matrix $A \in \mathbb{F}_2^{k \times k}$ such that

$$G' = AG = [I_k | P]$$

where I_k is the $k \times k$ identity matrix and $P \in \mathbb{F}_2^{k \times r}$. Let $y^{(1)}, y^{(2)}, \ldots, y^{(n)}$ denote the columns of G', with $y^{(j)} = e^{(j)}$ for $j \in [k]$.

Then it follows that

$$a^{(i)} = \sum_{j \in R_1} y^{(j)} = \sum_{j \in R_2} y^{(j)} = \sum_{j \in R_3} y^{(j)} \tag{1}$$

where $a^{(1)}, a^{(2)}, \ldots, a^{(k)} \in \mathbb{F}_2^k$ are the columns of A.

Note that $\dim \mathrm{Span}(a^{(1)}, \ldots, a^{(k)}) = k$, since the matrix A is non-singular. Let us now further define for $\ell \in [3]$:

$$\begin{aligned} S_\ell &= R_\ell \cap [k] \\ T_\ell &= R_\ell \cap ([n] \backslash [k]), \\ v^{(\ell)} &= \sum_{j \in S_\ell} y^{(j)} = \sum_{j \in S_\ell} e^{(j)} \end{aligned}$$

With this notation, we can rewrite (1) as follows:

$$a^{(i)} + v^{(\ell)} = \sum_{j \in T_\ell} y^{(j)} \ \forall \ell \in [3]$$

Finally, let us define $X = \left\{ y^{(k+1)}, y^{(k+2)}, \ldots, y^{(n)} \right\}$. It follows that $a^{(i)} + v^{(1)}$, $a^{(i)} + v^{(2)}, a^{(i)} + v^{(3)} \in \mathrm{Span}(X)$. We are now ready to use Propositions 1, 2, and 3 in order to complete the proof. Since the sets S_1, S_2, S_3 are disjoint, it follows that the supports of $v^{(1)}, v^{(2)}, v^{(3)}$ are also disjoint. In other words, $v^{(1)} \circ v^{(2)} = v^{(1)} \circ v^{(3)} = v^{(2)} \circ v^{(3)} = 0$. Using Proposition 3, we conclude that

$$\begin{aligned} a^{(i)} = &(a^{(i)} + v^{(1)}) \circ (a^{(i)} + v^{(2)}) \\ &+ (a^{(i)} + v^{(2)}) \circ (a^{(i)} + v^{(3)}) \\ &+ (a^{(i)} + v^{(3)}) \circ (a^{(i)} + v^{(1)}). \end{aligned}$$

Since the vectors $a^{(i)} + v^{(1)}, a^{(i)} + v^{(2)}, a^{(i)} + v^{(3)} \in \mathrm{Span}(X)$, it follows from Proposition 2 that the products of these vectors are in $\mathrm{Span}(X^2)$. This implies that $a^{(i)} \in \mathrm{Span}(X^2)$ for all i. Hence

$$\dim \mathrm{Span}(X^2) \geq \dim \mathrm{Span}(a^{(1)}, a^{(2)}, \ldots, a^{(k)}) = k$$

Using Proposition 1, we have $\dim \mathrm{Span}(X^2) \leq |X^2| \leq \binom{r+1}{2}$. Thus $\binom{r+1}{2} \geq k$, which completes the proof of the theorem. □

Acknowledgement

We would like to thank Eitan Yaakobi for helpful discussions.

REFERENCES

[1] H. Asi and E. Yaakobi. Nearly optimal constructions of pir and batch codes. *IEEE Transactions on Information Theory*, 65(2):947–964, 2018.

[2] A. Bhattacharyya, L. S. Chandran, and S. Ghoshal. Combinatorial lower bounds for 3-query ldcs. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[3] S. R. Blackburn and T. Etzion. Pir array codes with optimal pir rates. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2658–2662. IEEE, 2017.

[4] A. Fazeli, A. Vardy, and E. Yaakobi. Codes for distributed PIR with low storage overhead. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 2852–2856, 2015.

[5] S. L. Frank-Fischer, V. Guruswami, and M. Wootters. Locality via partially lifted codes. *CoRR*, abs/1704.08627, 2017.

[6] J. Hastings, A. Kanne, R. Li, and M. Wootters. Wedge-lifted codes. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 2990–2995. IEEE, 2021.

[7] L. Holzbaur, R. Polyanskaya, N. Polyanskii, I. Vorobyev, and E. Yaakobi. Lifted reed-solomon codes and lifted multiplicity codes. *IEEE Transactions on Information Theory*, 67(12):8051–8069, 2021.

[8] R. Li and M. Wootters. Lifted multiplicity codes and the disjoint repair group property. *IEEE Transactions on Information Theory*, 67(2):716–725, 2020.

[9] R. Li and M. Wootters. Improved batch code lower bounds. *arXiv preprint arXiv:2106.02163*, 2021.

[10] S. Rao and A. Vardy. Lower bound on the redundancy of pir codes. *arXiv preprint arXiv:1605.01869*, 2016.

[11] V. Skachek. Batch and pir codes and their connections to locally repairable codes. In *Network Coding and Subspace Designs*, pages 427–442. Springer, 2018.

[12] I. Tamo and A. Barg. Bounds on locally recoverable codes with multiple recovering sets. In *2014 IEEE International Symposium on Information Theory*, pages 691–695. IEEE, 2014.

[13] D. P. Woodruff. A quadratic lower bound for three-query linear locally decodable codes over any field. *Journal of Computer Science and Technology*, 27(4):678–686, 2012.

[14] M. Wootters. Linear codes with disjoint repair groups. Unpublished manuscript, available at https://web.stanford.edu/∼marykw/files/disjoint_repair_groups.pdf, 2016.