Westerbäck, Thomas; Freij-Hollanti, Ragnar; Ernvall, Toni; Hollanti, Camilla

On the Combinatorics of Locally Repairable Codes via Matroid Theory

Published in:
IEEE Transactions on Information Theory

DOI:
10.1109/TIT.2016.2598149

Published: 01/01/2016

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Westerbäck, T., Freij-Hollanti, R., Ernvall, T., & Hollanti, C. (2016). On the Combinatorics of Locally Repairable Codes via Matroid Theory. *IEEE Transactions on Information Theory*, *62*(10), 5296-5315. Article 7555340. https://doi.org/10.1109/TIT.2016.2598149

On the Combinatorics of Locally Repairable Codes via Matroid Theory

Thomas Westerbäck, Ragnar Freij-Hollanti, Toni Ernvall, and Camilla Hollanti

Abstract

This paper provides a link between matroid theory and locally repairable codes (LRCs) that are either linear or more generally almost affine. Using this link, new results on both LRCs and matroid theory are derived. The parameters (n, k, d, r, δ) of LRCs are generalized to matroids, and the matroid analogue of the generalized Singleton bound in [P. Gopalan *et al.*, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*] for linear LRCs is given for matroids. It is shown that the given bound is not tight for certain classes of parameters, implying a nonexistence result for the corresponding locally repairable almost affine codes, that are coined *perfect* in this paper.

Constructions of classes of matroids with a large span of the parameters (n, k, d, r, δ) and the corresponding local repair sets are given. Using these matroid constructions, new LRCs are constructed with prescribed parameters. The existence results on linear LRCs and the nonexistence results on almost affine LRCs given in this paper strengthen the nonexistence and existence results on perfect linear LRCs given in [W. Song *et al.*, "Optimal locally repairable codes," *IEEE J. Sel. Areas Comm.*].

T. Westerbäck, T. Ernvall, and C. Hollanti are with the Department of Mathematics and Systems Analysis, Aalto University, Finland.

R. Freij-Hollanti is with the Department of Communications and Networking, Aalto University, Finland.

E-mails: {thomas.westerback, ragnar.freij, toni.ernvall, camilla.hollanti}@aalto.fi.

The research of R. Freij-Hollanti is partially supported by the Finnish Academy of Science and Letters. The research of C. Hollanti is supported by the Academy of Finland grants #276031, #282938, and #283262, and by Magnus Ehrnrooth Foundation, Finland. The support from the European Science Foundation under the ESF COST Action IC1104 is also gratefully acknowledged.

Preliminary and partial results of this paper were presented at the 2014 IEEE Information Theory Workshop (ITW) in Hobart, Tasmania [1].

I. Introduction

Due to the ever-growing need for more efficient and scalable systems for cloud storage and data storage in general, distributed storage has become an increasingly important ingredient in many data systems. In their seminal paper [2], Dimakis *et al.* introduced network coding techniques for large-scale distributed storage systems such as data centers, cloud storage, peer-to-peer storage systems and storage in wireless networks. These techniques can, for example, considerably improve the storage efficiency compared to traditional storage techniques such as replication and erasure coding.

Failing devices are not uncommon in large-scale distributed storage systems [3]. A central problem for this type of storage is therefore to design codes that have good distributed repair properties. Several cost metrics and related tradeoffs [2], [4], [5], [6], [7], [8] are studied in the literature, for example *repair bandwidth* [2], [4], *disk-I/O* [9], and *repair locality* [10], [11], [12]. In this paper repair locality is the subject of interest.

The notion of a *locally repairable code* (LRC) was introduced in [13], and such repair-efficient codes are already used in existing distributed storage systems, *e.g.*, in the Hadoop Distributed File System *RAID* used by Facebook and Windows Azure Storage [14]. There are two notions of *symbol locality* considered in the literature: information locality only requires information symbols to be locally repairable, while all-symbol locality requires this to be true for all code symbols. The subject of interest in this paper is the all-symbol locality.

It is well-known that nonlinear codes often achieve better performance than linear ones, *e.g.*, in the context of coding rates for error-correcting codes and maximal throughput for network codes. Almost affine codes were introduced in [15] as a generalization of linear codes. This class of codes contains codes over arbitrary alphabet size, not necessarily prime power. In this paper, we are studying LRCs in the generality of almost affine codes.

We will consider five key invariants (n, k, d, r, δ) of locally repairable codes. The technical definitions are given in Section II-A, but in short, a good code should have large rate k/n as well as high global and local failure tolerance d and δ, respectively. In addition, it is desirable to have small r, which will determine the maximum number of nodes that have to be contacted for repair within a "local" repair set.

In this paper, our main tools for analyzing LRCs come from matroid theory. This is a branch of algebraic combinatorics with natural links to a great number of different topics, *e.g.*, to coding theory, graph theory, matching theory and combinatorial optimization. Matroids were introduced in

[16] in order to abstractly capture properties analogous to linear independence in vector spaces and independence in graphs. Since its introduction, matroid theory has been successfully used to solve problems in many areas of mathematics and computer science. Matroid theory and the theory of linear codes are closely related since every matrix over a field defines a matroid. Despite this fact, until rather recently matroid theory has only played a minor part in the development of coding theory. One pioneering work in this area is the paper by Greene from 1976 [17]. In this paper he describes how the weight enumerator of a linear code C is determined by the Tutte polynomial of the associated matroid of C. Using this result, Greene gives an elegant proof of the MacWilliams identity [18]. Generalizations of these results have then been presented in several papers, for example in [19], [20]. Another important instance of matroidal methods in coding theory is the development of a decomposition theory of binary linear codes [21]. Today, matroid theory also plays an important role in information theory and coding theory, for example in the areas of network coding, secret sharing, index coding, and information inequalities [22], [23], [24]. In this paper, while our main goal is investigating almost affine LRCs with the aid of matroid theory, ideas from the theory of LRCs will also be utilized to acquire new results in matroid theory.

A. Related work

One of the most classical theorems in coding theory is the Singleton bound, discussed in Section II-B [25]. Its classical version bounds the minimum distance d of a code from above in terms of the length n and dimension k. Recent work sharpens the bound in terms of the local parameters (r, δ) [10], [26], [27], [28], as well as in terms of other parameters [13], [29], [30].

There are different constructions of LRCs that are optimal in the sense that they achieve a generalized Singleton bound, *e.g.* [14], [26], [31], [32], [33]. Song *et al.* [32] investigate for which parameters (n, k, r, δ) there exists a linear LRC with all-symbol locality and minimum distance d achieving the generalized Singleton bound from [26]. The parameter set (n, k, r, δ) is divided into eight different classes. In four of these classes it is proven that there are linear LRCs achieving the bound, in two of these classes it is proven that there are no linear LRCs achieving the bound, and the existence of linear LRCs achieving the bound in the remaining two cases is an open question. Independently to the research in this paper, Wang and Zhang used linear programming approaches to strengthen these results when $\delta = 2$ [28].

It was shown in [14], that the r-locality of a linear LRC is a matroid invariant. This was used in

[14] to prove that the minimum distance of a class of linear LRCs achieves a generalized Singleton bound. Moreover, there are several instances of results in the theory of linear codes that have been generalized to all matroids. Examples on how these results can be interpreted for other objects that can represent a matroid, such as graphs, transversals and certain designs can be found in [34].

Recently, the present authors have studied locally repairable codes with all-symbol locality [35]. Methods to modify already existing codes were presented and it was shown that with high probability, a certain random matrix will be a generator matrix for a locally repairable code with a good minimum distance. Constructions were given for three infinite classes of optimal vector-linear locally repairable codes over an alphabet of small size. The present paper extends and deviates from this work by studying the combinatorics of LRCs in general and relating LRCs to matroid theory. This allows for the derivation of fundamental bounds for matroids and linear and almost affine LRCs, as well as for the characterization of the matroids achieving this bound.

In this paper, we have chosen to call the codes and matroids achieving the generalized Singleton bound *perfect* instead of optimal, reserving the term optimal for the best existing solution, *i.e.*, for codes achieving a tight bound instead of the (in some cases loose) Singleton bound. See Definition III.2 and the follow-up footnote for more details.

B. Contributions and organization

The first contribution of this paper is to extend the definitions of the parameters (n, k, d, r, δ) in [26] from linear codes to the much larger class of almost affine codes, and to show that these parameters are matroid invariant for all almost affine LRCs. We then proceed to prove the main results of this paper, which can be summarized as follows:

(i) A matroid analogue of the generalized Singleton bound in [26] is given for (n, k, d, r, δ)-matroids, and in particular to all almost affine codes in Theorem III.3.

(ii) In Theorem III.4, some necessary structural properties are given for an (n, k, d, r, δ)-matroid meeting the generalized Singleton bound.

(iii) In Theorem IV.1, a class of matroids is given with different values of the parameters (n, k, d, r, δ). Simple and explicit constructions of matroids in this class are given in Theorem IV.1 , Theorem IV.2, and Corollary IV.2, and in Examples IV.1, IV.2 and IV.3.

(iv) In Section V-B, we prove that the matroids from Theorem IV.1 are representable over finite fields of large enough size. Hence we obtain four explicit constructions of linear LRCs with given

parameters. The representability is derived by constructing a graph supporting a gammoid isomorphic to the matroid in Theorem IV.1, and using results on representability of gammoids [36].

(v) Theorem IV.4 characterizes values of (n, k, r, δ) for which there exist (n, k, d, r, δ)-matroids meeting the bound (i). In particular, the nonexistence results for linear LRCs in [32] are extended to the nonexistence of almost affine codes and matroids. Moreover, in Theorem V.4 and Theorem V.5, we settle the existence in one of the regimes left open in [32], leaving open only a minor subregime of $b > a \geq \lceil \frac{k}{r} \rceil - 1$, where $a = r \left\lceil \frac{k}{r} \right\rceil - k$ and $b = (r + \delta - 1)\lceil \frac{n}{r+\delta-1} \rceil - n$. This complements recent and independent research by Wang and Zhang [28], where they settle the existence in the subregime $\lceil \frac{n}{r+1} \rceil > b$ and $\delta = 2$ using integer programming techniques.

The proofs of some of the longer theorems and the explicit constructions of matroids with prescribed parameters are given in the Appendix.

II. Preliminaries

A. Parameters (n, k, d, r, δ) of locally repairable codes

In this subsection, we introduce the parameters (n, k, d, r, δ) defined in [26] for linear locally repairable codes. We extend this definition to the much wider class of almost affine codes, to be introduced in II-F. Figure 1 serves as a visual aid for the technical definitions. The information symbols (a, b, c, d, e, f) are stored on twelve nodes as in the figure. Equivalently, we think of the content of the twelve nodes as a codeword, and of the content of an individual node as a code symbol. Within each of the local clouds (or locality sets), three symbols are enough to determine the other two. Thus, Figure 1 depicts a $(12, 6, 3, 3, 3)$-LRC, according to the following definitions.

Let $C \subseteq \mathbb{A}^n$ be a code such that $|C| = |\mathbb{A}|^k$, where $\mathbb{A}$ is a finite set, also referred to as the *alphabet*. For any subset $X = \{i_1, \ldots, i_m\} \subseteq [n] = \{1, 2, \ldots, n\}$, let C_X denote the *projection* of the code into $\mathbb{A}^{|X|}$, that is

$$C_X = \{(c_{i_1}, \ldots, c_{i_m}) : \boldsymbol{c} = (c_1, \ldots, c_n) \in C\}. \tag{1}$$

The code C_X is also called a punctured code in the coding theory literature. The minimum (Hamming) distance d of C can be defined in terms of projections as

$$d = \min\{|X| : X \subseteq [n] \text{ and } |C_{[n]\setminus X}| < |C|\}. \tag{2}$$

For $1 \leq r \leq k$ and $\delta \geq 2$, an (r, δ)-*locality set* of C is a subset $S \subseteq [n]$ such that

(i) $|S| \leq r + \delta - 1$

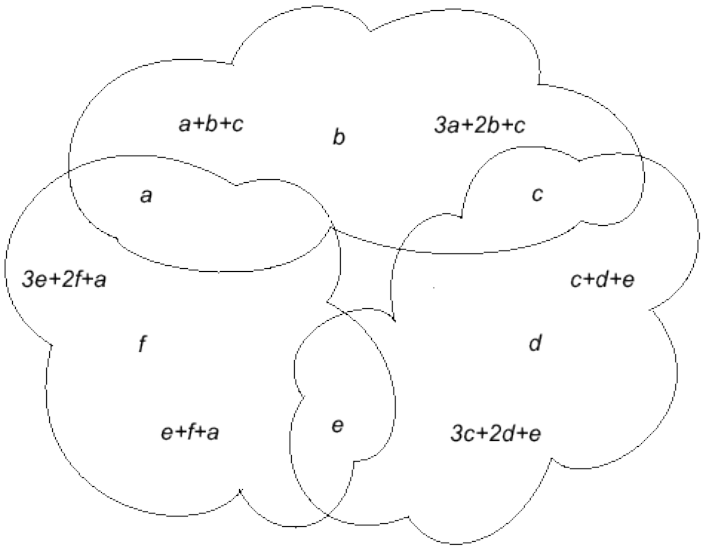

Fig. 1. A storage system from a $(12, 6, 3, 3, 3)$-LRC.

(ii) For every $l \in S$, $L = \{i_1, \ldots, i_{|L|}\} \subseteq S \setminus \{l\}$ and $|L| = |S| - (\delta - 1)$, c_l is a function of $(c_{i_1}, \ldots, c_{i_{|L|}})$, where $\mathbf{c} = (c_1, \ldots, c_n) \in C$.

We say that C is a *locally repairable code (LRC)* with *all-symbol locality* (r, δ) if all the n symbols of the code are contained in an (r, δ)-locality set. The locality sets can be also referred to as the local repair sets.

We remark that the symbols in a locality set S can be used to recover up to $\delta - 1$ lost symbols in the same locality set. Further, we note that each of the following statements are equivalent to statement (ii) above:

(ii′) For any $l \in S$, $L = \{i_1, \ldots, i_{|L|}\} \subseteq S \setminus \{l\}$, and $|L| = |S| - (\delta - 1)$, we have $|C_{L\cup\{l\}}| = |C_L|$,

(ii″) For any $L \subseteq S$ with $|L| \geq |S| - (\delta - 1)$, we have $|C_L| = |C_S|$,

(ii‴) $d(C_S) \geq \delta$, where $d(C_S)$ is the minimum distance of C_S.

An LRC with parameters (n, k), minimum distance d, and all-symbol locality (r, δ) is an (n, k, d, r, δ)-LRC. Since we focus only on all-symbol locality in this paper, we will henceforth use the term LRC to mean a locally repairable code with all-symbol locality.

B. The Singleton bound

For any $[n, k]$-linear code with minimum distance d, the Singleton bound is given by

$$d \leq n - k + 1. \tag{3}$$

This bound was generalized for locally repairable codes in [10] (the case $\delta = 2$) and [26] (general δ) as follows. A linear LRC with parameters (n, k, d, r, δ) satisfies

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1). \tag{4}$$

While the bounds in [10] and [26] are stated assuming only information locality, so are of course in particular still valid under the stronger assumption of all-symbol locality. Other generalizations of the Singleton bound for linear and nonlinear LRCs can be found in [13], [29], [30].

C. Graphs, $G = (V, E)$

Let us fix some standard graph-theoretic notation that will be used at two stages in the constructions. A (finite) directed graph $G = (V, E)$ is a pair of a finite *vertex set* V, whose elements are called nodes or vertices, and an *edge set* $E \subseteq V \times V$ of pairs called arcs or edges. Graphs are often drawn with the vertices as points and arcs (v, u) as arrows $v \to u$. We call v the tail of (v, u), and u the head. A path from $S \subseteq V$ to $T \subseteq V$ is a sequence $v_0, v_1, \ldots, v_n$, where $v_0 \in S$, $v_n \in T$, and $(v_i, v_{i+1}) \in E$ for each $i = 0, \ldots, n-1$. If $v_0 = v_n$, then the path is called a (directed) cycle.

An important case of graphs is when E is *symmetric*, *i.e.*, $(u, v) \in E$ if and only if $(v, u) \in E$. In such case, it is customary to identify the two pairs (u, v) and (v, u) with the set $\{u, v\}$, and erase all the heads of the arrows in the drawing. When talking about a graph without specifying that it is directed, the symmetric situation is assumed. Observe that this definition allows for loops edges (where the tail and the head is the same), but not multiple edges. In this paper, we will assume that all graphs, both symmetric and directed, are without multiple edges and loops.

D. Posets and lattices, $(\mathcal{P}, \subseteq)$

Before studying matroids, we need a minimum of background on poset and lattice theory. We refer the reader to [37] for more information on posets and lattices. The material in this section is used only in the technical work with the lattice of cyclic flats of matroids.

A collection of sets $\mathcal{P} \subseteq 2^E$ ordered by inclusion $\subseteq$ defines a (finite) poset $(\mathcal{P}, \subseteq)$. A *chain* C of $(\mathcal{P}, \subseteq)$ is a set of elements $X_0, \ldots, X_m \in \mathcal{P}$ such that $X_0 \subsetneq X_1 \subsetneq \ldots \subsetneq X_m$. The *length* of a chain C is defined as the integer $l(C) = |C| - 1 = m$. For $X, Y \in \mathcal{P}$, let

$$L_{X,Y} = \{Z \in \mathcal{P} : Z \subseteq X \text{ and } Z \subseteq Y\},$$
$$U_{X,Y} = \{Z \in \mathcal{P} : X \subseteq Z \text{ and } Y \subseteq Z\}.$$

An element $Z \in L_{X,Y}$ is the *meet* of X and Y, denoted by $X \wedge Y$, if it contains every $V \in L_{X,Y}$. Dually, $Z \in U_{X,Y}$ is the *join* of X and Y, denoted by $X \vee Y$, if it is contained in every $V \in U_{X,Y}$. A poset $(\mathcal{P}, \subseteq)$ is a *lattice* if every pair of elements of $\mathcal{P}$ has a meet and a join. If $(\mathcal{P}, \subseteq)$ is a (finite) lattice, then there are two elements $0_{\mathcal{P}}, 1_{\mathcal{P}} \in \mathcal{P}$ such that $0_{\mathcal{P}} \subseteq X$ and $X \subseteq 1_{\mathcal{P}}$ for all $X \in \mathcal{P}$. The *atoms* and *coatoms* of a lattice $(\mathcal{L}, \subseteq)$ are defined as

$$\begin{aligned} A_{\mathcal{L}} &= \{X \in \mathcal{L} \setminus 0_{\mathcal{L}} : \nexists Y \in \mathcal{L} \text{ such that } 0_{\mathcal{L}} \subsetneq Y \subsetneq X\}, \\ coA_{\mathcal{L}} &= \{X \in \mathcal{L} \setminus 1_{\mathcal{L}} : \nexists Y \in \mathcal{L} \text{ such that } X \subsetneq Y \subsetneq 1_{\mathcal{L}}\}, \end{aligned}$$

respectively.

E. Matroids, $M = (\rho, E)$

Matroids can be defined in many equivalent ways, for example by their rank function, nullity function, independent sets, circuits and more [38]. For our purpose, the following definition will be the most useful. Let 2^E denote the set of all subsets of E. A *matroid* M on a finite set E is defined by a *rank function* $\rho : 2^E \to \mathbb{Z}$ satisfying the following axioms:

$$\begin{aligned} &(R1) \quad 0 \leq \rho(X) \leq |X| \text{ for } X \subseteq E, \\ &(R2) \quad X \subseteq Y \subseteq E \Rightarrow \rho(X) \leq \rho(Y), \\ &(R3) \quad X, Y \subseteq E \Rightarrow \rho(X) + \rho(Y) \geq \rho(X \cup Y) + \rho(X \cap Y). \end{aligned} \tag{5}$$

The *nullity function* $\eta : 2^E \to \mathbb{Z}$ of the matroid $M = (E, \rho)$ is defined by

$$\eta(X) = |X| - \rho(X), \text{ for } X \subseteq E.$$

Let X be any subset of E. The subset X is *independent* if $\rho(X) = |X|$, otherwise it is *dependent*. A dependent set X is a *circuit* if all proper subsets of X are independent, *i.e.*, $\rho(X) = |X| - 1$ and $\rho(Y) = |Y|$ for all subsets $Y \subsetneq X$. The *closure* of X is defined as

$$\mathrm{cl}(X) = \{x \in E : \rho(X \cup x) = \rho(X)\}.$$

The subset X is a *flat* if $\mathrm{cl}(X) = X$. It is *cyclic* if it is a (possible empty) union of circuits. The sets of circuits, independent sets, cyclic sets and cyclic flats of a matroid M is denoted by $\mathcal{C}(M)$, $\mathcal{I}(M)$, $\mathcal{U}(M)$ and $\mathcal{Z}(M)$, respectively. We omit the subscript M when the matroid is clear and write $\mathcal{C}$, $\mathcal{I}$, $\mathcal{U}$ and $\mathcal{Z}$, respectively. The set of cyclic flats together with inclusion defines the *lattice of cyclic flats* $(\mathcal{Z}, \subseteq)$ of the matroid. The *restriction* of M to X is the matroid $M|X = (\rho_{|X}, X)$ where

$$\rho_{|X}(Y) = \rho(Y), \text{ for all subsets } Y \subseteq X. \tag{6}$$

F. Almost affine codes and their associated matroids

A code $C \subseteq \mathbb{A}^n$, where $\mathbb{A}$ is a finite set of size $s \geq 2$, is *almost affine* if

$$\log_s(|C_X|) \in \mathbb{Z}$$

for each $X \subseteq [n]$. Note that if C is an almost affine code, then all projections C_X of C are also almost affine.

In [15] it is proven that every almost affine code

$C \subseteq \mathbb{A}^n$

induces a matroid $M_C = (\rho_C, [n])$, where

$$\rho_C(X) = \log_s(|C_X|). \tag{7}$$

Examples of matroids which cannot be represented by any almost affine code are given in [39]. Moreover, an example of a matroid which can be represented by an almost affine code over a three letter alphabet, but not by any linear code is given in [15]. This example is the so-called non-Pappus matroid.

Example II.1. *An example of a matroid $M_G = (\rho, E)$ is defined by the matrix*

$$G = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \end{array} \begin{array}{c} \begin{array}{cccccccccccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} & \mathbf{6} & \mathbf{7} & \mathbf{8} & \mathbf{9} & \mathbf{10} & \mathbf{11} & \mathbf{12} \end{array} \\ \begin{pmatrix} 1 & & & & & & 1 & & 1 & 3 & & 1 \\ & 1 & & & & & 1 & & & 2 & & \\ & & 1 & & & & 1 & 1 & & 1 & 3 & \\ & & & 1 & & & & 1 & & & 2 & \\ & & & & 1 & & & 1 & 1 & & 1 & 3 \\ & & & & & 1 & & & 1 & & & 2 \end{pmatrix} \end{array}, \tag{8}$$

which we think of as a generator matrix of a linear code C over the field $\mathbb{F}_5$. The code C is the row span of G, $E = \{\mathbf{1}, \ldots, \mathbf{12}\}$ is the set of columns, and the rank of a subset of E is the rank of the corresponding submatrix, i.e.*,*

$$\rho(I) = rank(G_I) \text{ for } I \subseteq E,$$

where G_I is the submatrix of G whose columns are the columns indexed by I. Below are some

independent sets, circuits, cyclic flats and rank functions of some subsets of E for the matroid M.

$$\mathcal{I} = \{\emptyset, \{2,3,7\}, \{3,4,5\}, \{7,8,9\}, [6], \ldots\},$$
$$\mathcal{C} = \{\{1,2,3,7\}, \{4,5,8,11\}, \ldots\},$$
$$\mathcal{Z} = \{\{1,2,3,7,10\}, \{3,4,5,8,11\}, \{1,2,3,4,5,7,8,10,11\}, [12], ...\},$$
$$\rho(\emptyset) = 0,\ \rho(\{3,4,5\}) = \rho(\{4,5,8,11\}) = \rho(\{3,4,5,8,11\}) = 3,\ \rho([6]) = \rho([12]) = 6.$$

The reader can verify that the code generated by this matrix corresponds to the storage system in Figure 1, when the rows are the information symbols.

G. Basic properties of matroids and the lattice of cyclic flats

For the applications in this paper, the most important matroid attribute is its lattice of cyclic flats. This is because the minimal cyclic flats of matroids will correspond to local repair sets of the LRC. In this subsection, we present basic properties of the lattice of cyclic flats, that will be needed in later parts of the paper.

Proposition II.1 (see [40])**.** *Let $M = (\rho, E)$ be a matroid. Then*

(i) $\rho(X) = \min\{\rho(F) + |X \setminus F| : F \in \mathcal{Z}\}$*, for* $X \subseteq E$*,*

(ii) Define $\mathcal{D} = \{X :$ *there is* $F \in \mathcal{Z}$ *with* $X \subseteq F$ *and* $|X| = \rho(F) + 1\}$*.*
Then $\mathcal{C}$ is the set of minimal elements in $\mathcal{D}$, ordered by inclusion.

(iii) $(\mathcal{Z}, \subseteq)$ *is a lattice with the following meet and join for* $X, Y \in \mathcal{Z}$*,*
$X \wedge Y = \bigcup_{\{C \in \mathcal{C} : C \subseteq X \cap Y\}} C$ *and* $X \vee Y = \mathrm{cl}(X \cup Y)$*.*

The assertion (i) in Proposition II.1 shows that a matroid is determined by its cyclic flats and their ranks. Conversely, the following theorem gives an axiomatic scheme for a collection of subsets on E and a function on these sets to define the cyclic flats of a matroid and their ranks. This will allow us to construct matroids with prescribed parameters in Section III.

Theorem II.1 (see [40] Th. 3.2)**.** *Let $\mathcal{Z} \subseteq 2^E$ and let ρ be a function $\rho : \mathcal{Z} \to \mathbb{Z}$. There is a matroid M on E for which $\mathcal{Z}$ is the set of cyclic flats and ρ is the rank function restricted to the sets in $\mathcal{Z}$ if*

and only if

$$
\begin{aligned}
&(Z0) \quad \mathcal{Z} \text{ is a lattice under inclusion},\\
&(Z1) \quad \rho(0_{\mathcal{Z}}) = 0,\\
&(Z2) \quad X, Y \in \mathcal{Z} \text{ and } X \subsetneq Y \Rightarrow\\
&\qquad 0 < \rho(Y) - \rho(X) < |Y| - |X|,\\
&(Z3) \quad X, Y \in \mathcal{Z} \Rightarrow \rho(X) + \rho(Y) \geq\\
&\qquad \rho(X \vee Y) + \rho(X \wedge Y) + |(X \cap Y) \setminus (X \wedge Y)|.
\end{aligned}
$$

The results in the proposition below are basic matroid results that will be needed several times in the proofs of other results given later in this paper. We give a proof for the results in Proposition II.2 that we have not been able to find in the literature. For the other results we only give a reference.

Proposition II.2. *Let $M = (\rho, E)$ be a matroid and let X, Y be subsets of E, then*

(i) *If $X \subseteq Y$, then $\eta(X) \leq \eta(Y)$,*

(ii) $\eta(X \cup Y) \geq \eta(X) + \eta(Y) - \eta(X \cap Y)$,

(iii) *If $\rho(X) < \rho(E)$ and $1_{\mathcal{Z}} = E$, then $\eta(X) \leq \max\{\eta(Z) : Z \in coA_{\mathcal{Z}}\}$,*

(iv) $\mathrm{cl}(U) \in \mathcal{Z}(M)$ *for* $U \in \mathcal{U}(M)$,

(v) $\mathcal{U}(M|X) = \{U \subseteq X : U \in \mathcal{U}(M)\}$,

(vi) $\mathcal{C}(M|X) = \{C \subseteq X : C \in \mathcal{C}(M)\}$,

(vii) $\mathcal{Z}(M|X) = \{Z \in \mathcal{Z}(M) : Z \subseteq X\}$ *if* $X \in \mathcal{F}(M)$,

(viii) $X \notin \mathcal{U}(M)$ *if and only if* $\exists x \in X$ *such that* $\rho(X - x) < \rho(X)$,

(ix) $\rho(\mathrm{cl}(X)) = \rho(X)$,

(x) *If $X \subseteq Y$, then* $\mathrm{cl}(X) \subseteq \mathrm{cl}(Y)$.

Proof: Properties (i), (ii), (v), (vii) and (viii) can be found in [41, Lemma 2.2.4, Lemma 2.3.1, the paragraph under Lemma 2.4.5]. Property (iv) is a consequence of [38, Proposition 1.4.10 (ii)]. For (iii), assume that $\rho(X) < \rho(E)$ and $1_{\mathcal{Z}} = E$. Thus, $\mathrm{cl}(X) \neq E$ and $\eta(X) \leq \eta(\mathrm{cl}(X))$. Let U be the largest cyclic set such that $U \subseteq \mathrm{cl}(X)$. From [41, Lemma 2.4.8, Lemma 2.5.2], we have that $\eta(\mathrm{cl}(X)) = \eta(U)$ and that U is a cyclic flat. Property (iv) now follows from the fact that

$$\rho(U) \leq \rho(\mathrm{cl}(X)) < \rho(E) = \rho(1_{\mathcal{Z}}).$$

Property (vi) is a direct consequence of (v). Property (ix) is a consequence of property (x) which can be found in [38, Lemma 1.4.2] ■

Example II.2. *Continuing with Example II.1, and remembering that the elements of M_G are the columns of G, we see that the cyclic flats of M_G are the submatrices in Figure 2. The atomic cyclic flats are thus the submatrices corresponding to column sets $\{1,2,3,7,10\}$, $\{3,4,5,8,11\}$ and $\{1,5,6,9,12\}$. Remembering from* (8) *that the rows are indexed by the information symbols (a,b,c,d,e,f), these atomic cyclic flats agree exactly with the local clouds in Figure 1.*

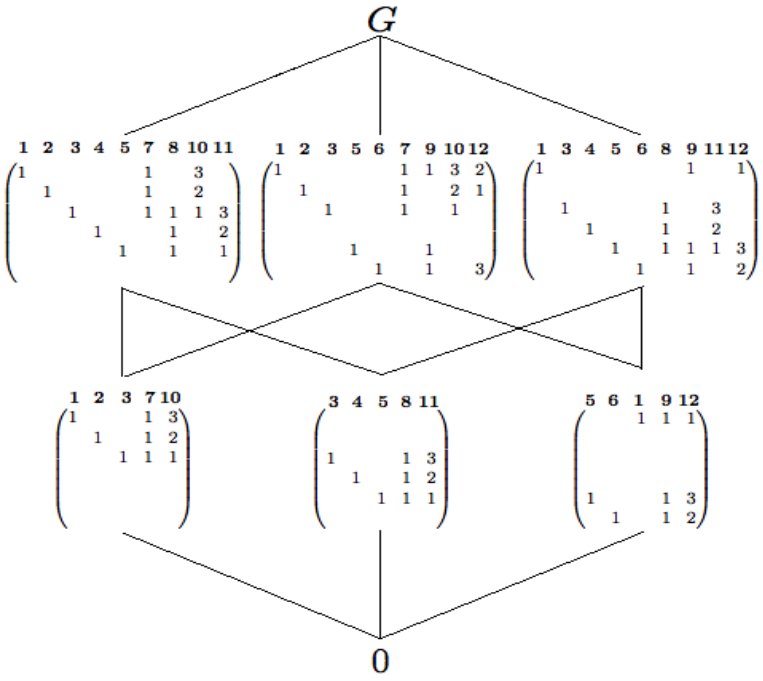

Fig. 2. The lattice $\mathcal{Z}(M_G)$ of cyclic flats of the matroid $M(G)$ in Example II.2.

III. Locally Repairable Matroids

A. The parameters (n,k,d,r,δ) for matroids

In this subsection we show that the parameters (n,k,d,r,δ) are matroid invariants for an almost affine LRC. This will allow us to extend the definition of these parameters to matroids in general.

Let C be an almost affine (n,k,d,r,δ)-LRC over some finite alphabet $\mathbb{A}$. By the definition given in Eq. (7), we know that $|C_X| = |\mathbb{A}|^{\rho_C(X)}$, which specializes to $k = \rho_C([n])$ when $X = [n]$. In [15] it is proven that $M_{C_X} = M_C|X$ for $X \subseteq [n]$. Consequently, since the projection C_X is also almost affine, (2) implies that

$$d(C_X) = \min\{|Y| : Y \subseteq X \text{ and } \rho_C(X \setminus Y) < \rho_C(X)\},$$

where $d(C_X)$ denotes the minimum distance of C_X.

Using the observations above and the definition of an (n, k, d, r, δ)-LRC given in Section II-A, we conclude the following theorem.

Theorem III.1. *Let C be an almost affine LRC with the associated matroid $M_C = (\rho_C, [n])$. Then, the parameters (n, k, d, r, δ) of C are matroid invariants, where*

(i) $k = \rho_C([n])$,

(ii) $d = \min\{|X| : X \subseteq [n]$ *and* $\rho_C([n] \setminus X) < k\}$,

(iii) C *has all-symbol locality* (r, δ) *if and only if, for every* $j \in [n]$ *there exists a subset* $S_j \subseteq [n]$ *such that*

a) $j \in S_j$,

b) $|S_j| \leq r + \delta - 1$,

c) $d(C_{S_j}) = \min\{|X| : X \subseteq S_j$ *and* $\rho_C(X) < \rho_C(S_j)\} \geq \delta$.

These results can now be taken as the definition of the parameters (n, k, d, r, δ) for an arbitrary matroid.

Definition III.1. *Let $M = (\rho, E)$ be a matroid. Then we call M an (n, k, d, r, δ)-matroid, where*

(i) $n = |E|$,

(ii) $k = \rho(E)$,

(iii) $d = \min\{|X| : X \subseteq E$ *and* $\rho(E \setminus X) < k\}$,

(iv) The parameters $0 < r \leq \rho(E)$ *and* $\delta \geq 2$ *are such that for all* $x \in E$, *there exists a subset* $S_x \subseteq E$ *with*

a) $x \in S_x$,

b) $|S_x| \leq r + \delta - 1$,

c) $d(M|S_x) = \min\{|X| : X \subseteq S_x$ *and* $\rho(S_x \setminus X) < \rho(S_x)\} \geq \delta$.

A subset $S \subseteq E$ is called a (r, δ)-*locality set* of the elements $x \in S$ if the statements b)–c) above are satisfied by S. The parameters n and k are obviously defined for all matroids. We note that the parameter d is finite if and only $k > 0$. Furthermore, we notice that every element $x \in E$ is contained in some cyclic set S_x if and only if $1_{\mathcal{Z}} = E$. If this is the case, and $r = \max\{|S_x| - 1 : x \in X\}$, then M has $(r, 2)$-locality. As a consequence of the observations above, we get the following proposition.

Proposition III.1. *A matroid $M = (\rho, E)$ is an (n, k, d, r, δ)-matroid with finite values of (n, k, d, r, δ) if and only if $0 < \rho(E)$ and $1_{\mathcal{Z}} = E$.*

Observe that if M has (r,δ)-locality, then by Definition III.1 (iv), M has (r',δ')-locality for $r \leq r' \leq k$ and $2 \leq \delta' \leq \delta$ with $r'+\delta' \geq r+\delta$. So neither the values of (r,δ) nor the locality sets S_x are in general uniquely determined for a

matroid

M.

B. A generalized Singleton bound for (n,k,d,r,δ)-matroids

The main result of this subsection is Theorem III.3 which gives a Singleton-type bound on the parameters (n,k,d,r,δ) for matroids. In the case of linear LRCs with information locality and trivial failure tolerance $\delta = 2$, *i.e.*, only tolerating one failure, the bound was given in [10].

The core ingredients of the proof of Theorem III.3 are the same as in [10], interpreted for matroids. First, we relate the parameters (n,k,d,r,δ) of a matroid to its lattice of cyclic flats in Theorem III.2. Then in Lemma III.1, we obtain a large cyclic flat Y_{m-1} of rank less than k. In Theorem III.3 we relate Y_{m-1} to d, thereby proving the theorem.

Theorem III.2. *Let $M=(\rho,E)$ be an (n,k,d,r,δ)-matroid with $0<\rho(E)$ and $1_{\mathcal{Z}}=E$. Then*

(i) $d = n-k+1-\max\{\eta(Z) : Z \in coA_{\mathcal{Z}}\}$,

(ii) For each $x \in E$, there is a cyclic set $S_x \in \mathcal{U}(M)$ such that

a) $x \in S_x$,

b) $|S_x| \leq r+\delta-1$,

c) $d(M|S_x) = \eta(S_x)+1-\max\{\eta(Z) : Z \in coA_{\mathcal{Z}(M|S_x)}\} \geq \delta$.

Proof: The proof is given in the Appendix. ∎

As $\eta(Z)$ is non-negative for every Z, Theorem III.2 (ii) c) gives $\delta+\rho(S_x)-1 \leq |S_x|$, which together with Theorem III.2 (ii) b) shows that

$$\rho(S_x) \leq r \tag{9}$$

for any (r,δ)-locality S_x. Moreover, we observe that for any atom S in a lattice of cyclic flats with $0_{\mathcal{Z}}=\emptyset$, we can use any subset $S' \subseteq S$ as a locality set when $|S'| > \rho(S)$. However, different choices of locality sets may give different values on the parameters (r,δ).

Example III.1. *Representing the cyclic flats associated to the matroid M_G from Example II.2 just by their corresponding sets and ranks in Figure 3, we use Theorem III.2 to get the parameters (n,k,d,r,δ) of the linear LRC that is generated by the matrix G given in Example II.1.*

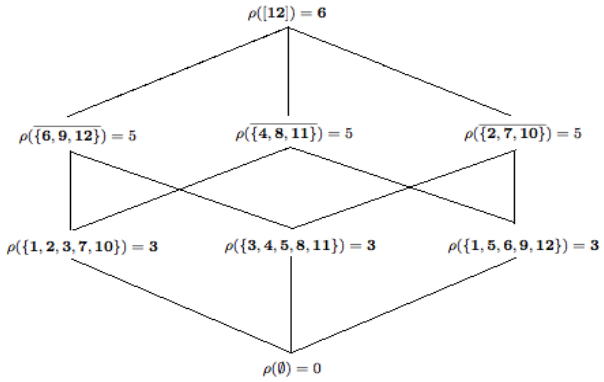

Fig. 3. The lattice $\mathcal{Z}(M_G)$ of cyclic flats of the matroid $M(G)$ in Example II.2, without reference to the matrix G.

The values for (n, k, d) *are*

$$n = 12,$$
$$k = 6,$$
$$d = 12 - 6 + 1 - 4 = 3.$$

Using $S_1 = \{1, 2, 3, 7, 10\}$, $S_2 = \{3, 4, 5, 8, 11\}$ *and* $S_3 = \{1, 5, 6, 9, 12\}$ *as the locality sets, we get the parameters* $(r, \delta) = (3, 3)$.

From Theorem III.2, we derive a chain of cyclic flats, from which we will extract a large cyclic flat, to be used in the proof of Theorem III.3.

Lemma III.1. *Let* $M = (\rho, E)$ *be an* (n, k, d, r, δ)*-matroid. Then there is a chain*

$$0_{\mathcal{Z}} = Y_0 \subsetneq Y_1 \subsetneq \ldots \subsetneq Y_m = E$$

in $(\mathcal{Z}(M), \subseteq)$ *such that for* $j = 1, \ldots, m$ *we have*

(i) $\rho(Y_j) \leq \rho(Y_{j-1}) + r$,

(ii) $\eta(Y_j) \geq \eta(Y_{j-1}) + (\delta - 1)$.

Proof: The proof is given in the Appendix. ∎

We are now ready to prove the generalized Singleton bound for matroids.

Theorem III.3. *Let* $M = (\rho, E)$ *be an* (n, k, d, r, δ)*-matroid. Then*

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Proof: Let

$$C : 0_{\mathcal{Z}} = Y_0 \subsetneq Y_1 \subsetneq \ldots \subsetneq Y_m = E$$

be a chain of $(\mathcal{Z}, \subseteq)$ given in Lemma III.1. Then $\eta(Y_{m-1}) \geq (m-1)(\delta-1)$, by Lemma III.1 (ii). On the other hand, by Lemma III.1 (i) we have that $k = \rho(Y_m) \leq mr$, so $m \geq \lceil \frac{k}{r} \rceil$.

Combining these results we get

$$\eta(Y_{m-1}) \geq (m-1)(\delta-1) \geq (\left\lceil \frac{k}{r} \right\rceil - 1)(\delta-1).$$

Since $Y_{m-1} \in \mathcal{Z} \setminus \{1_{\mathcal{Z}}\}$, we have

$$\max\{\eta(Z) : \eta(Z) \in coA_{\mathcal{Z}}\} \geq \eta(Y_{m-1}),$$

so Theorem III.2 (i) yields

$$d \leq n - k + 1 - \eta(Y_{m-1}) \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right)(\delta - 1).$$

■

We also give three additional bounds on the parameters of a matroid.

Proposition III.2. *Let $M = (\rho, E)$ be an (n, k, d, r, δ)-matroid. Then*

(i) $\delta \leq d$,

(ii) $k \leq n - \lceil \frac{k}{r} \rceil (\delta - 1)$,

(iii) $\frac{k}{n} \leq \frac{r}{r+\delta-1}$.

Proof: The proof is given in the appendix. ■

In the case of codes, Proposition III.2 (i) and (iii) have natural interpretations. Indeed, (i) says that the local minimum distance is bounded from above by the global minimum distance, and (iii) says that the global code rate is bounded from above by the local code rate.

C. A structure theorem for matroids achieving the generalized Singleton bound

Definition III.2. *We will call an (n, k, d, r, δ)-matroid* perfect *if it meets the generalized Singleton bound of Theorem III.3 with equality,* i.e. *if*

$$d = n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right)(\delta - 1). \tag{10}$$

In analogy, we will call a LRC satisfying (10) *a* perfect *LRC*[1].

[1]We point out that, typically, codes achieving these kind of bounds have been called optimal in the literature. However, we feel that the notion *optimal* should be saved for the code that is the best we can do. Thus, saying that an optimal code does not exist when the bound cannot be reached with equality feels wrong, since we can still find a code with minimum distance only slightly smaller than the bound, and this code is the best possible solution in this case and thus deserves to be called optimal. Therefore, we have opted to call the codes achieving the bound *perfect*. This is say that, even though perfect codes do not exist for all parameters, optimal solutions can still be found.

These notions should not be confused with those of a perfect matroid design or a perfect code in classical coding theory literature. Theorem III.4 gives some necessary structural properties for perfect (n, k, r, δ)-matroids with $r < k$. We will use this structure theorem to prove that for certain values of (n, k, r, δ), there are no perfect (n, k, r, δ)-matroids, and consequently no perfect LRCs. The degenerate case when $r = k$ is easier, and is considered in Section IV-B1.

A collection of sets $X_1 \ldots, X_j$ is said to have a *non trivial union* if

$$X_l \nsubseteq \bigcup_{i \in [j] \setminus \{l\}} X_i, \text{ for } l = 1, \ldots, j.$$

Theorem III.4. *Let $M = (\rho, E)$ be an (n, k, d, r, δ)-matroid with $r < k$ and*

$$d = n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Let then $\{S_x : x \in E\} \subseteq \mathcal{U}(M)$ be a collection of cyclic sets for which the statements a) – c) in Theorem III.2 (iv) are satisfied. Then

(i) $0_{\mathcal{Z}} = \emptyset$,

(ii) for each $x \in E$,

a) $\eta(S_x) = (\delta - 1)$,

b) S_x *is an atom in* $\mathcal{Z}(M)$, *and in particular a cyclic flat.*

(iii) For each collection $F_1, \ldots F_j$ of cyclic flats in $\{S_x : x \in E\}$ that has a non trivial union,

c) $\eta(\bigvee_{i=1}^{j} F_i) = \begin{cases} j(\delta - 1) & \text{if} \quad j < \lceil \frac{k}{r} \rceil, \\ n - k \geq \lceil \frac{k}{r} \rceil (\delta - 1) & \text{if} \quad j \geq \lceil \frac{k}{r} \rceil, \end{cases}$

d) $\bigvee_{i=1}^{j} F_i = \begin{cases} \bigcup_{i=1}^{j} F_i & \text{if} \quad j < \lceil \frac{k}{r} \rceil, \\ E & \text{if} \quad j \geq \lceil \frac{k}{r} \rceil, \end{cases}$

e) $\rho(\bigvee_{i=1}^{j} F_i) = \begin{cases} |\bigcup_{i=1}^{j} F_i| - j(\delta - 1) & \text{if} \quad j < \lceil \frac{k}{r} \rceil, \\ k & \text{if} \quad j \geq \lceil \frac{k}{r} \rceil. \end{cases}$

f) $|F_j \cap (\bigcup_{i=1}^{j-1} F_i)| \leq |F_j| - \delta$ *if* $j \leq \lceil \frac{k}{r} \rceil$.

Proof: The proof is given in the Appendix. ∎

By the structure theorem III.4 above we get the following corollary.

Corollary III.1. *Let $M = (\rho, E)$ be an (n, k, d, r, δ)-matroid with $r < k$ and*

$$d = n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Then M has a collection of cyclic flats $F_1, \ldots, F_m$ such that

(i) $\{F_i\}_{i\in[m]}$ *has a non trivial union,*

(ii) $|F_i| \leq r+\delta-1$ *for* $i=1,\ldots,m$*,*

(iii) $\eta(F_i)=\delta-1$*,*

(iv) $\{X\in\mathcal{Z}(M) : X\subseteq F_i\}=\{\emptyset,F_i\}$ *for* $i=1,\ldots,m$*,*

(v) $\bigcup_{i\in[m]} F_i = E$*,*

(vi) statements c)–f) in Theorem III.4 holds for every collection of flats $\{F_i\}_{i\in I}$ *with* $I\subseteq[m]$ *and* $|I|\leq\lceil\frac{k}{r}\rceil$*,*

(vii) $k\leq|\bigcup_{i\in I\setminus\{j\}}F_i|+|F_j|-\lceil\frac{k}{r}\rceil(\delta-1)-|(\bigcup_{i\in I\setminus\{j\}}F_i)\cap F_j|$ *for* $I\subseteq[m]$*,* $|I|=\lceil\frac{k}{r}\rceil$ *and* $j\in I$*.*

Proof: The statements (i)–(v) follows directly from Theorem III.4 (i)–(ii) and Theorem III.2 (iv). Statement (vi) is a consequence of (i) and Theorem III.4 (iii), since (i) implies that $\{F_i\}_{i\in I}$ has a non trivial union. For statement (vii) we first observe by (iv), (vi) and Proposition II.1 (iii) that

$$(\bigvee_{i\in I\setminus\{j\}} F_i)\wedge F_j = (\bigcup_{i\in I\setminus\{j\}} F_i)\wedge F_j = \emptyset.$$

Hence, by (vi) and axiom (Z3) in Theorem II.1,

$$\begin{aligned} k &= \rho(\textstyle\bigvee_{i\in I}F_i) \\ &\leq \rho(\textstyle\bigvee_{i\in I\setminus\{j\}}F_i)+\rho(F_j)-\rho(\emptyset)-|(\bigvee_{i\in I\setminus\{j\}}F_i)\cap F_j| \\ &= |\textstyle\bigcup_{i\in I\setminus\{j\}}F_i|-(\lceil\frac{k}{r}\rceil-1)(\delta-1)+|F_j|-(\delta-1)-|(\bigcup_{i\in I\setminus\{j\}}F_i)\cap F_j|. \end{aligned}$$

■

We remark that structure theorems similar in spirit to the above have been given for linear (n,k,d,r,δ)-LRCs in [10] and [42]. Namely, Theorem 2.2 in [42] covers the case when $r|k$, showing that local repair sets correspond to linear $[r+\delta-1,r,\delta]$-MDS codes and are mutually disjoint. Theorem 7 in [10] proves the same in the special case $\delta=2$.

Corollary III.1 (iv) means that the local matroid $M|F_i$ is uniform of rank $|F_i|-(\delta-1)$, for $i=1,\ldots,m$. When the matroid comes from a linear code, the code in question is thus an $[|F_i|,|F_i|-(\delta-1),\delta]$-MDS code. By (vi) and (vii) in Corollary III.1, we obtain conditions on how large the intersections of union of subsets of the cyclic flats $\{F_i\}_{i\in[m]}$ can be. These results imply the corresponding results on linear LRCs.

IV. Constructions and classes of (n,k,d,r,δ)-Matroids

The generalized Singleton bound theorem for matroids gives an upper bound for the value of d in terms of the parameters (n,k,r,δ) for a matroid. In subsection IV-A we will give some constructions

on (n, k, d, r, δ)-matroids. These constructions will then be used in Subsection IV-B, where we will investigate, given different classes of the parameters (n, k, r, δ), whether or not perfect (n, k, r, δ)-matroids exist.

A. Combinatorial constructions of (n, k, d, r, δ)-matroids

In this section we will give four increasingly specialized constructions of (n, k, d, r, δ)-matroids. The constructions are purely combinatorial, and proceed by assigning the atomic cyclic flats, together with the rank function on the lattice of cyclic flats. In Section V, we prove that the matroids we have constructed can be represented by linear codes.

1) General construction of (n, k, d, r, δ)-matroids: Let $F_1, \ldots, F_m$ be a collection of subsets of a finite set E and define $F_I = \bigcup_{i \in I} F_i$ for $I \subseteq [m]$. Further, let k be a nonnegative integer and ρ a function $\rho : \{F_i\}_{i \in [m]} \to \mathbb{Z}$ satisfying

(i) $0 < \rho(F_i) < |F_i|$ for $i \in [m]$,

(ii) $F_{[m]} = E$,

(iii) $k \leq |F_{[m]}| + \sum_{i \in [m]} (\rho(F_i) - |F_i|)$,

(iv) $I \subseteq [m], j \in [m] \setminus I \Rightarrow |F_I \cap F_j| < \rho(F_j)$.

Define

$$\mathcal{Z}_{<k} = \{F_J : |F_J| + \sum_{i \in J} (\rho(F_i) - |F_i|) < k\}$$

and $\mathcal{Z} = \mathcal{Z}_{<k} \cup \{E\}$.

Now, we extend the function ρ to a function $\mathcal{Z} \to \mathbb{Z}$, by

$$\begin{cases} \rho(F_J) &= |F_J| + \sum_{i \in J} (\rho(F_i) - |F_i|) \text{ for } F_J \in \mathcal{Z}_{<k}, \\ \rho(E) &= k. \end{cases} \tag{11}$$

Note that the extension of ρ given in (11) is well defined, as by (iii), E is not in $\mathcal{Z}_{<k}$. Also note that $F_\emptyset = \emptyset$ and $\rho(F_\emptyset) = 0$. Finally, we define $\mathcal{I} = \{X \subseteq E : |F_I \cap X| \leq \rho(F_I) \text{ for all } I \subseteq [m]\}$.

Theorem IV.1. *Let $F_1, \ldots, F_m$ be a collection of subsets of a finite set E, k a nonnegative integer and $\rho : \{F_i\}_{i \in [m]} \to \mathbb{Z}$ a function satisfying (i)–(iv). Then $\mathcal{Z}$ and $\rho : \mathcal{Z} \to \mathbb{Z}$, defined in* (11)*, define an (n, k, d, r, δ)-matroid $M(F_1, \ldots, F_m; k; \rho)$ on E for which $\mathcal{Z}$ is the collection of cyclic flats, ρ is the rank function restricted to the cyclic flats, $\mathcal{I}$ is the set of independent sets, and*

(i) $n = |E|$,

(ii) $k = \rho(E)$,

(iii) $d = n - k + 1 - \max\{\sum_{i\in I} \eta(F_i) : F_I \in \mathcal{Z}_{<k}\}$,

(iv) $\delta = 1 + \min_{i\in[m]}\{|F_i| - \rho(F_i)\}$,

(v) $r = \max_{i\in[m]}\{\rho(F_i)\}$.

Proof: The proof is given in the Appendix. ∎

Example IV.1. *Let F_1, F_2, F_3 be disjoint sets of cardinality 4, with $\rho(F_1) = \rho(F_2) = 3$ and $\rho(F_3) = 2$. Moreover, let $(k, r, \delta) = (7, 3, 3)$. By Theorem IV.1, this corresponds to a matroid of size 14 and minimum distance 4.*

2) Specialized construction of (n, k, d, r, δ)-matroids: To construct (n, k, d, r, δ)-matroids with large d in Section IV-B, we will use a special case of the construction in IV.1. We represent the atomic cyclic flats F_i by nodes in a graph, with labelled edges representing the intersections between the flats. The construction of a lattice of cyclic flats from a weighted graph can be made much more general by assigning weights to the nodes, representing the size and rank of the corresponding flats. However, in this section we specialize all parameters to obtain matroids that achieve the Singleton bound.

Let G be a graph with vertices $[m]$ and edges W, and let $\gamma : W \to \mathbb{Z}_{\geq 1}$ be a positive integer-valued function on the edge set. Moreover, let (k, r, δ) be three integers with $0 < r < k$ and $\delta \geq 2$, such that

(i) G has no triangles,

(ii) $k \leq rm - \sum_{w\in W} \gamma(w)$,

(iii) $r > \sum_j \gamma(\{i, j\})$ for every $i \in [m]$.

From the graph G we construct the sets $F_1, \ldots, F_m$ and the rank function ρ by first assigning the following:

(iv) $\rho(F_i) = r$ for $i \in [m]$,

(v) $|F_i| = r + \delta - 1$ for $i \in [m]$,

(vi) $|F_i \cap F_j| = \gamma(\{i, j\})$ for $\{i, j\} \in W$.

Note that (v)–(vii) uniquely defines the sets $F_1, \ldots, F_m$ and their ranks, up to isomorphism. This can be seen by induction over m, observing that (iv) guarantees that the intersections $F_i \cap F_j$ can be chosen to be disjoint for different j. This is required, as there is no 3-cycle in the graph G, so

$$|F_h \cap F_i \cap F_j| = 0 \text{ for all three distinct elements } h, i, j \in [m].$$

Also note that, while n is not a parameter of the graph construction, it is a function of the parameters,

as we have

$$n = |\cup_i F_i| = m(r+\delta-1) - \sum_{w\in W}\gamma(w).$$

Theorem IV.2. *Let $F_1,\ldots,F_m$ and $\rho : \{F_i\} \to \mathbb{Z}$ be constructed from a weighted graph (G,γ) with parameters (k,r,δ) according to (i)–(vi). Then $(\{F_i\},\rho)$ satisfies (i)–(iv) in IV-A1. In particular, $\{F_i\}$ are the atomic cyclic flats of an (n,k,d,r,δ)-matroid with*

(i) $n = (r+\delta-1)m - \sum_{w\in W}\gamma(w)$,

(ii) $d = n-k+1-(\delta-1)\max\{|I| : r|I| - \sum_{w\in W\cap I\times I}\gamma(w) < k\}$.

Proof: The proof is given in the Appendix. ∎

Now, in addition, we assume that G has girth at least $\max\{4, \lceil\frac{k}{r}\rceil + 1\}$, and that the weight function γ does not take too large values. Then we get the following theorem, on the existence of perfect (n,k,r,δ)-matroids.

Corollary IV.1. *Let (G,γ) be a weighted graph, and let (k,r,δ) be integers such that (i)–(iii) is satisfied. Let $b = \sum_{w\in W}\gamma(w)$, and $a = \lceil\frac{k}{r}\rceil r - k$. Assume moreover that G has no l-cycles, for $l \le \lceil\frac{k}{r}\rceil$, and that $\sum_{w\in W\cap I\times I}\gamma(w) \le a$ for every $I\subseteq[m]$ with $|I| = \lceil\frac{k}{r}\rceil$.*

Then there exists a (n,k,d,r,δ)-matroid with

(i) $n = (r+\delta-1)m - b$,

(ii) $d = n-k+1-(\lceil\frac{k}{r}\rceil - 1)(\delta-1)$.

Proof: We need to prove that

$$\left\lceil\frac{k}{r}\right\rceil - 1 \le \max\left\{|I| : r|I| - \sum_{w\in W\cap I\times I}\gamma(w) < k = \left\lceil\frac{k}{r}\right\rceil r - a\right\}.$$

If $|I| = \lceil\frac{k}{r}\rceil - 1$, then

$$r|I| - \sum_{w\in W\cap I\times I}\gamma(w) \le r(\left\lceil\frac{k}{r}\right\rceil - 1) < k.$$

If, on the other hand, $|I| = \lceil\frac{k}{r}\rceil$, then

$$r|I| - \sum_{w\in W\cap I\times I}\gamma(w) = r\left\lceil\frac{k}{r}\right\rceil - \sum_{w\in W\cap I\times I}\gamma(w) \ge r\left\lceil\frac{k}{r}\right\rceil - a = k,$$

by assumption. Thus, the corollary follows from Theorem IV.2. ∎

Corollary IV.2. *Let (G,γ) be a weighted graph, and let (k,r,δ) be integers such that (i)–(iii) is satisfied. Let $b = \sum_{w\in W}\gamma(w)$, and $a = \lceil\frac{k}{r}\rceil r - k$. Assume moreover that G has no l-cycles, for*

$l \leq \lceil \frac{k}{r} \rceil$, *and that* $1 \leq \gamma(w) \leq \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor$ *for every* $w \in W$. *Then there is an* (n, k, d, r, δ)*-matroid with*

(i) $n = (r + \delta - 1)m - b$,

(ii) $d = n - k + 1 - (\lceil \frac{k}{r} \rceil - 1)(\delta - 1)$.

Proof: Since G has no l-cycles for $l \leq \lceil \frac{k}{r} \rceil$, we have for every $I \subseteq [m]$ with $|I| = \lceil \frac{k}{r} \rceil$ that $|W \cap I \times I| \leq \lceil \frac{k}{r} \rceil - 1$. Since $\gamma(w) \leq \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor$, we then get

$$\sum_{w \in W \cap I \times I} \gamma(w) \leq \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) \leq a,$$

so Theorem IV.1 applies. ■

We remark that in order to find as small n as possible for a chosen (k, r, δ, a, b) in Corollary IV.2, we want to find a good graph with as few nodes as possible. To find such a graph, preferable properties for the graph are: many small cycles of length $\max\{4, \lceil \frac{k}{r} \rceil + 1\}$, large values of γ on every edge, *i.e.* $\gamma(w) = \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor$ for $w \in W$, and that the sum of γ-values incident to each node is large, *i.e.* $\sum_j \gamma(\{i, j\}) = r - 1$ for all nodes $i \in [m]$).

Example IV.2. *Let* G *denote the graph below on the vertex set* $[6]$*, where the values of* γ *are written above the edges in the graph, and* $(k, r, \delta) = (14, 4, 2)$. *We get* $b = \sum \gamma(w) = 3$ *and* $a = r \lceil \frac{k}{r} \rceil - k = 2$

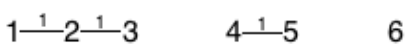

Fig. 4. The graph $G(\gamma; 14, 4, 2, 2, 3)$

By Corollary IV.2, this graph corresponds to a $(27, 14, 11, 4, 2)$*-matroid on the ground set* $[27]$*, with six atomic cyclic flats* $F_1, \ldots F_6$*, where*

$$F_1 = \{1, \ldots, 5\},\ F_2 = \{1, 6, \ldots, 9\},\ F_3 = \{6, 10, \ldots, 13\},$$
$$F_4 = \{14, \ldots, 18\},\ F_4 = \{14, 19, \ldots, 22\} \text{ and } F_5 = \{23, \ldots, 27\}.$$

Example IV.3. *Let* $G = G(\gamma; k, r, \delta, a, b)$ *denote the graph below on the vertex set* $[11]$. *The* γ*-values for the edges are written in the graph and* $(k, r, \delta, a, b) = (19, 9, 5, 8, 21)$.

By Corollary IV.2, this graph corresponds to a $(122, 19, 96, 9, 5)$*-matroid, whose lattice of cyclic flats has* 11 *atoms.*

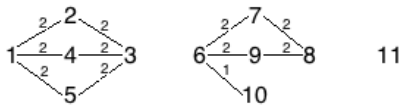

Fig. 5. The graph $G(\gamma; 19, 9, 5, 8, 21)$

B. The maximal d for (n, k, r, δ)-matroids

We know by Theorem III.3, that the inequality

$$d \leq n - k - (\left\lceil \frac{k}{r} \right\rceil - 1)(\delta - 1)$$

holds for any (n, k, d, r, δ)-matroid. It is then very natural to ask what is the maximal value of d, for which there exists an (n, k, r, δ)-matroid, for given (n, k, r, δ) with $0 < r \leq k \leq n - (\delta - 1)\lceil \frac{k}{r} \rceil$ and $\delta \geq 2$. We will denote this maximal value $d_{\max} = d_{\max}(n, k, r, \delta)$. The case $r = k$ is degenerate, and we will consider this first. The case when $r < k$ will be further divided into four subcases in Theorem IV.4. Theorem IV.4 will later translate into results for linear LRCs in Theorem V.4 and Theorem V.5.

1) The maximal value of d when $r = k$: A well known class of matroids is the class of *uniform matroids* [38], defined as $U_n^k = (\rho, E)$, where

$$|E| = n \text{ and } \rho(X) = \min\{|X|, k\}. \tag{12}$$

This implies that the cyclic sets of U_n^k is

$$\mathcal{U}(U_n^k) = \{\emptyset\} \cup \{X \subseteq E : |X| \geq k + 1\},$$

and that the cyclic flats are

$$\mathcal{Z} = \{0_{\mathcal{Z}}, 1_{\mathcal{Z}}\}, \text{ with } 0_{\mathcal{Z}} = \emptyset,\ 1_{\mathcal{Z}} = E,\ \rho(0_{\mathcal{Z}}) = 0 \text{ and } \rho(1_{\mathcal{Z}}) = k.$$

If $k = r$, the generalized Singleton bound given in Theorem III.3 reduces to the classical Singleton bound, $d = n - k + 1$. Then using Theorem III.2 (iii), we get that $\mathcal{Z} = \{\emptyset, E\}$, so M is the uniform matroid U_n^k. For (r, δ)-locality, let $S_x = U_n^k$ for each $x \in E$ and $\delta = d = n - k + 1$. Then $|S_x| = r + (\delta - 1)$ and $d(S_x) = \delta$. Consequently, U_n^k is a matroid with parameters $(n, k, d, r, \delta) = (n, k, n - k + 1, r, n - k + 1)$.

2) The maximal value of d when $r < k$: As the first result of this section, we prove that

$$n - k - \left\lceil \frac{k}{r} \right\rceil (\delta - 1) \leq d_{\max}(n, k, r, \delta) \leq n - k - (\left\lceil \frac{k}{r} \right\rceil - 1)(\delta - 1),$$

where the second inequality is Theorem III.3 revisited. We will then use the graph constructions given in Theorem IV.2 and Theorem IV.1, in order to construct matroids with larger d. In the cases when $d_{\max} < n - k - (\lceil \frac{k}{r} \rceil - 1)(\delta - 1)$, we will use Theorem III.4 to prove this.

Theorem IV.3. *For any (n, k, r, δ) satisfying $1 \leq r < k \leq n - \lceil \frac{k}{r} \rceil (\delta - 1)$ and $2 \leq \delta < n$, there exists a (n, k, d, r, δ)-matroid, where*

$$d = n - k - \left\lceil \frac{k}{r} \right\rceil (\delta - 1).$$

Proof: Let $m = \left\lceil \frac{n}{r+\delta-1} \right\rceil$, and let $F_1, \ldots F_{m-1}$ be disjoint sets with rank r and size $r + \delta - 1$. Let F_m be disjoint from all of $F_1, \ldots F_{m-1}$, with size $|F_m| = n - (m-1)(r + \delta - 1)$ and rank $\rho(F_m) = |F_m| - \delta + 1$. Finally, let M be defined by $\mathcal{Z}(M) = \{F_I\}$, where $F_I = \cup_{i \in I} F_i$, and

$$\rho(F_I) = \min\{\sum_{i \in I} \rho(F_i), k\}.$$

It is readily seen that M has minimum distance

$$d = n - k + 1 - (\delta - 1) \max\{|I| : \rho(F_I) < k\} \geq n - k + 1 - \left\lceil \frac{k}{r} \right\rceil (\delta - 1).$$

■

In particular, when $\delta = 2$, this means that the optimal minimum distance is one of $n - k + 1 - \lceil \frac{k}{r} \rceil$ and $n - k - \lceil \frac{k}{r} \rceil$. The remainder of this section aims at deciding which of these two possibilities is the case for fixed

Before stating the technical theorem on $d_{\max}$, we need the following qualitative result.

Proposition IV.1. *Let M be an (n, k, d, r, δ)-matroid and let $a = \lceil \frac{k}{r} \rceil r - k$ and $b = \left\lceil \frac{n}{r+\delta-1} \right\rceil (r + \delta - 1) - n$. Then the following hold,*

$$\left\lceil \frac{n}{r + \delta - 1} \right\rceil \geq \begin{cases} \lceil \frac{k}{r} \rceil & \text{if} \quad b \leq a, \\ \lceil \frac{k}{r} \rceil + 1 & \text{if} \quad b > a, \end{cases}$$

Proof: Let $\left\lceil \frac{n}{r+\delta-1} \right\rceil = \lceil \frac{k}{r} \rceil + t$. Note that $n - k \geq \lceil \frac{k}{r} \rceil (\delta - 1)$ by Proposition III.2. Hence,

$$\begin{aligned} \lceil \tfrac{k}{r} \rceil (\delta - 1) &\leq n - k \\ &= (\lceil \tfrac{k}{r} \rceil + t)(r + \delta - 1) - b - (\lceil \tfrac{k}{r} \rceil r - a) \\ &= \lceil \tfrac{k}{r} \rceil (\delta - 1) + t(r + \delta - 1) - (b - a). \end{aligned}$$

This implies that $t \geq 0$ if $b \leq a$ and $t \geq 1$ if $b > a$. ■

Theorem IV.4. *Let (n,k,r,δ) be integers such that $0<r<k\leq n-\left\lceil\frac{k}{r}\right\rceil(\delta-1)$, $k=\left\lceil\frac{k}{r}\right\rceil r-a$ and $n=\left\lceil\frac{n}{r+\delta-1}\right\rceil(r+\delta-1)-b$. Let $d_{\max}=d_{\max}(n,k,r,\delta)$ be the largest d such that there exists an (n,k,d,r,δ)-matroid. Then the following hold.*

(i) If $a\geq b$, then $d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)$;

(ii) If $b>a$ and $b\geq r$, then $d_{\max}\geq n-k+1-\left\lceil\frac{k}{r}\right\rceil(\delta-1)+(b-r)$.

(iii) If $b>a$ and $a<\left\lceil\frac{k}{r}\right\rceil-1$, then

$d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)$ *if and only if* $\left\lfloor\left\lceil\frac{k}{r}\right\rceil/2\right\rfloor\leq a$ *and*

$$\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\left\lceil\frac{k}{r}\right\rceil-1+(b-a)\left(1+\frac{1}{t}\right),$$

where $t=\left\lfloor a/\left(\left\lceil\frac{k}{r}\right\rceil-1-a\right)\right\rfloor$;

(iv) If $b>a\geq\left\lceil\frac{k}{r}\right\rceil-1$, $\left\lceil\frac{k}{r}\right\rceil\geq 3$ and

$$\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\left\lfloor\frac{b}{stu}\right\rfloor(t(u-1)+2)+y,$$

where $s=\left\lfloor\frac{a}{\left\lceil\frac{k}{r}\right\rceil-1}\right\rfloor$, $t=\left\lfloor\frac{r-1}{s}\right\rfloor$, $u=\left\lceil\frac{\left\lceil\frac{k}{r}\right\rceil+1}{2}\right\rceil$, $x=\left\lceil\frac{b-\left\lfloor\frac{b}{stu}\right\rfloor stu}{s}\right\rceil$, and

$$y=\begin{cases}0 & \text{if } stu\mid b,\\ x-\left\lfloor\frac{x}{u}\right\rfloor+1+\min\{\left\lfloor\frac{x}{u}\right\rfloor,1\} & \text{if } stu\nmid b,\end{cases}$$

then $d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)$;

(v) If $b>a\geq\left\lceil\frac{k}{r}\right\rceil-1$, $\left\lceil\frac{k}{r}\right\rceil=2$, and

$$\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\begin{cases}\left\lceil\frac{b}{a}\right\rceil+1 & \text{if } 2a\leq r-1,\\ \left\lceil\frac{b}{\left\lfloor\frac{r-1}{2}\right\rfloor}\right\rceil+1 & \text{if } 2a>r-1,\end{cases}$$

then $d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)$.

Proof: The proof is given in the Appendix. ■

In the proof of Theorem IV.4(iv), we will notice that a simpler bound, but in general not as good, is $\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\left\lceil\frac{b}{stu}\right\rceil(t(u-1)+2)$.

Example IV.4. *Examples of constructions of matroids in Theorem IV.4(i), (iii) and (iv) given by the proofs of the theorem are given in Example IV.1, IV.2 and IV.3 respectively.*

V. Applications of (n, k, d, r, δ)-Matroids to (n, k, d, r, δ)-LRCs

In this section we will use the previous results on (n, k, d, r, δ)-matroids to get new results on linear and almost affine (n, k, d, r, δ)-LRCs. All the proofs of the non-existence of matroids immediately give corresponding bounds for codes. To verify the other direction, obtaining codes with prescribed parameter values from matroids with the same parameters, we will show that the class of matroids given in Theorem IV.1 is a subclass of a class of matroids called gammoids. Gammoids have the property that they are representable over any finite field of sufficiently large size.

The main result in this section is Theorem V.1.

Theorem V.1. *Let $M(F_1, \ldots, F_m; k; \rho)$ be an (n, k, d, r, δ)-matroid that we get in Theorem IV.1. Then for every large enough finite field there is a linear LRC over the field with parameters (n, k, d, r, δ).*

A. Transversal matroids and gammoids

We start by giving a short introduction to gammoids. For more information on this fascinating class of matroids we refer the reader to [38], [43].

A gammoid is associated to a directed graph G as follows.

Definition V.1. *Let $G = (V, D)$ be a directed graph, with $S \subseteq V$ and $T \subseteq V$.The* gammoid $M(G)$ *is a matroid $M(G)$ on S where the independent sets of $M(G)$ equals*

$$\mathcal{I}(M(G)) = \{X \subseteq S : \exists \text{ a set of } |X| \text{ vertex-disjoint paths from } X \text{ to } T\}.$$

Our interest in gammoids in this paper stems from the following result.

Theorem V.2 ([36])**.** *Every gammoid over a finite set E is representable over every finite field of size greater than or equal to $2^{|E|}$.*

Many natural classes of gammoids, can be represented over fields of much smaller size than 2^n. For example, a uniform matroid U_n^k (12) is a gammoid associated to a complete bipartite graph with $V = S \cup T$, $|S = n|$, $|T| = k$ and $D = S \times T$. However, uniform matroids are represented by linear $[n, k, d = n - k + 1]$-MDS codes, which exist over $\mathbb{F}_q$ when $q \geq n$.

B. Constructions of linear (n, k, d, r, δ)-LRCs $C(F_1, \ldots, F_m; k; \rho)$

Theorem V.1 follows immediately from Lemma V.1 and Theorem V.2. The key element is the construction of a directed graph whose associated gammoid is the matroid from Theorem IV.1. This construction is detailed in Algorithm 1.

Algorithm 1 Input: $(F_1, \ldots, F_m; k; \rho)$. Output: $G = (V, D, S, T)$

$S = E$, $H = \emptyset$, $D = \emptyset$, $T = [k]$

Label $e \in S$ with $s(e) = \{i : e \in F_i\}$.

h *is a function* $H \to 2^{[m]}$

for all $e \in E$ **do**

  if $|s(e)| \geq 2$ **then**

    $H \leftarrow H \cup \{u_e\}$

    $h(u_e) = s(e)$

for all $i \in [m]$ **do**

  $l_i = \rho(F_i) - |\{u \in H : i \in h(u)\}|$

  $H \leftarrow H \cup \{v_1^i, \ldots, v_{l_i}^i\}$

  $h(v_1^i) = \ldots = h(v_{l_i}^i) = \{i\}$

for all $(e, u) \in S \times H$ **do**

  if $s(e) \subseteq h(u)$ **then**

    $D \leftarrow D \cup (\overrightarrow{e, u})$

$D \leftarrow D \cup H \times T$

$V = S \cup (H \cup T)$

Output (V, D, S, T)

Lemma V.1. *Let $F_1, \ldots, F_m$ be a collection of subsets of a finite set E whose union is all of E, and write $F_I = \cup_{i \in I} F_i$. Let $\rho : \{F_i\}_{i \in [m]} \to \mathbb{Z}$ satisfy*

(i) $0 \leq \rho(F_i) \leq |F_i|$,

(ii) $k \leq |F_{[m]}| + \sum_{i \in [m]} (\rho(F_i) - |F_i|)$,

(iii) $|F_I \cap F_j| < \rho(F_j)$ *whenever* $j \notin I$.

Then the gammoid $M(G)$, that we get from Algorithm 1 is equal to the matroid $M(F_1, \ldots, F_m; k; \rho)$ that we get in TheoremIV.1.

Proof: The proof is given in the Appendix. ■

Example V.1. *Consider the matroid M_G, associated to the storage system in Figure 1 and the code in Example II.1, and whose lattice of cyclic flats is written out in Example II.2. By Lemma V.1, this is the gammoid associated to the following graph, with $|T| = k = 6$ and $|S| = n = 12$. Note that in*

this particular setting, Line 15 in Algorithm 1 is superfluous, could be replaced by assigning $H = T$, since H already has only 6 nodes. Indeed, the inclusion of the bipartite graph (H, T) corresponds to truncating the gammoid at rank k.

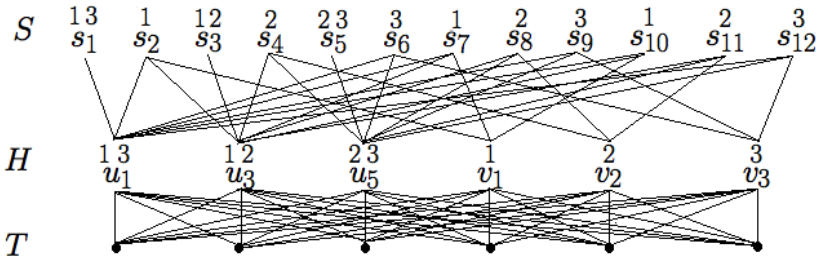

Fig. 6. A downward directed graph supporting the matroid associated to a $(12, 6, 3, 3, 3)$-LRC.

C. Bounds on the parameters (n, k, d, r, δ) for LRCs

In this section we will give results on the parameters (n, k, d, r, δ) for linear, and more generally almost affine LRCs. The results are direct consequences of the corresponding results for matroids, thanks to the representability results in Theorem V.1 and the matroid invariance of the parameters (n, k, d, r, δ), from Theorem III.1. We will therefore not give any further proofs in this section. Observe that this means that the same bounds are valid for matroids, almost affine codes, and linear codes.

Theorem V.3. *If C is an almost affine LRCs with the parameters (n, k, d, r, δ), then*

$$d \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Proposition V.1. *Let C be an almost affine LRC with parameters (n, k, d, r, δ). Then*

(i) $\delta \leq d$,

(ii) $k \leq n - \left\lceil \frac{k}{r} \right\rceil (\delta - 1)$,

(iii) $\frac{k}{n} \leq \frac{r}{r+\delta-1}$.

Theorem V.4. *Let C be an almost affine LRC with parameters (n, k, r, d, δ), and let $a = \left\lceil \frac{k}{r} \right\rceil r - k$ and $b = \left\lceil \frac{n}{r+\delta-1} \right\rceil (r + \delta - 1) - n$. Then the following hold.*

(i) If $b > a$ and $a < \left\lfloor \left\lceil \frac{k}{r} \right\rceil / 2 \right\rfloor$, then $d < n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1)$;

(ii) If $b > a$ and $\left\lfloor \left\lceil \frac{k}{r} \right\rceil / 2 \right\rfloor \leq a \leq \left\lceil \frac{k}{r} \right\rceil - 1$, then $d < n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1)$.

Theorem V.5. *Let (n,k,r,δ) be integers such that $0<r<k\leq n-\left\lceil\frac{k}{r}\right\rceil(\delta-1)$, $a=\left\lceil\frac{k}{r}\right\rceil r-k$ and $b=\left\lceil\frac{n}{r+\delta-1}\right\rceil(r+\delta-1)-n$. Let $d_{\max}=d_{\max}(n,k,r,\delta)$ be the largest d such that there exists a linear LRC with parameters (n,k,d,r,δ). Then the following hold.*

(i) If $a\geq b$, then $d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)$;

(ii) If $b>a$, then

$$d_{\max}\geq\begin{cases}n-k+1-\left\lceil\frac{k}{r}\right\rceil(\delta-1) & \text{if}\quad b\leq r-1,\\ n-k+1-\left\lceil\frac{k}{r}\right\rceil(\delta-1)+(b-r) & \text{if}\quad b\geq r;\end{cases}$$

(iii) If $b>a$, $\left\lfloor\left\lceil\frac{k}{r}\right\rceil/2\right\rfloor\leq a<\left\lceil\frac{k}{r}\right\rceil-1$ and $\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\left\lceil\frac{k}{r}\right\rceil-1+(b-a)\left(1+\frac{1}{t}\right)$, where $t=\left\lfloor a/\left(\left\lceil\frac{k}{r}\right\rceil-1-a\right)\right\rfloor$, then

$$d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1);$$

(iv) If $b>a\geq\left\lceil\frac{k}{r}\right\rceil-1$, $\left\lceil\frac{k}{r}\right\rceil\geq 3$ and $\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\left\lfloor\frac{b}{stu}\right\rfloor(t(u-1)+2)+y$, where $s=\left\lfloor\frac{a}{\left\lceil\frac{k}{r}\right\rceil-1}\right\rfloor$, $t=\left\lfloor\frac{r-1}{s}\right\rfloor$, $u=\left\lceil\frac{\left\lceil\frac{k}{r}\right\rceil+1}{2}\right\rceil$, $x=\left\lceil\frac{b-\left\lfloor\frac{b}{stu}\right\rfloor stu}{s}\right\rceil$ and

$$y=\begin{cases}0 & \text{if}\quad stu\mid b,\\ x-\left\lfloor\frac{x}{u}\right\rfloor+1+\min\{\left\lfloor\frac{x}{u}\right\rfloor,1\} & \text{if}\quad stu\nmid b,\end{cases}$$

then

$$d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1);$$

(v) If $b>a\geq\left\lceil\frac{k}{r}\right\rceil-1$, $\left\lceil\frac{k}{r}\right\rceil=2$ and

$$\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\begin{cases}\left\lceil\frac{b}{a}\right\rceil+1 & \text{if}\quad 2a\leq r-1,\\ \left\lceil\frac{b}{\left\lfloor\frac{r-1}{2}\right\rfloor}\right\rceil+1 & \text{if}\quad 2a>r-1,\end{cases}$$

then

$$d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1).$$

Just like in the remark below Theorem IV.4, a simpler bound, but in general not as good, in Theorem V.5(iv) is $\left\lceil\frac{n}{r+\delta-1}\right\rceil\geq\left\lceil\frac{b}{stu}\right\rceil(t(u-1)+2)$.

It was proven in [42] Corollary 2.3 that linear LRCs with all-symbol locality in the case when $r|k$ and $r+\delta-1\nmid n$ cannot achieve the Singleton-type bound given in Theorem V.3. This corresponds to the case $a=0$ and $b>0$ in Theorem V.4. Hence, by Theorem V.5 (ii), we obtain that

$$d_{\max}=n-k+1-\left(\left\lceil\frac{k}{r}\right\rceil-1\right)(\delta-1)-1,$$

for linear (n,k,d,r,δ)-LRCs when $r|k$ and $b=r+\delta-2$.

VI. Conclusions

Recent progress in coding theory has proven matroid theory to be a valuable tool in many different contexts. This trend carries over to locally repairable codes. Especially the lattice of cyclic flats is a useful object to study, as its elements correspond to the local repair sets.

We have thoroughly studied linear and more generally almost affine LRCs with all-symbol locality, as well as the connections of these codes to matroid theory. We derived a generalized Singleton bound for matroids and nonexistence results for certain classes of (n, k, d, r, δ)-matroids. These results can then be directly translated to nonexistence results for almost affine LRCs.

Further, we have given several constructions of matroids with prescribed values of the parameters (n, k, d, r, δ). Using these matroid constructions, novel constructions of linear LRCs are given, using the representability of gammoids. Several classes of optimal linear LRCs then arise from these constructions.

As future work, (non)existence results for matroids and linear and almost affine LRCs achieving the generalized Singleton bound remain open for certain classes of parameters (n, k, r, δ), when $\lceil \frac{k}{r} \rceil - 1 \leq a < b$. In addition, the size of the underlying finite field that our linear (n, k, d, r, δ)-LRCs can be constructed over is left for future research. We expect that the upper bound 2^n arising from the related bound for all gammoids is loose for our class of matroids. We conjecture that all our matroids from Section IV-A are representable over fields of size polynomial in n.

References

[1] T. Westerbäck, T. Ernvall, and C. Hollanti, "Almost affine locally repairable codes and matroid theory," *2014 IEEE Inf. Theory Workshop (ITW)*, 2014.

[2] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," *IEEE Trans. Inf. Theory*, 56(9), pp. 4539–4551, September 2010.

[3] S. Ghemawat, H. Gobioff, and S. T. Leung, "The Google file system," In *SOSP03, Proceedings of the nineteenth ACM symposium on Operating systems principles*, 2003.

[4] T. Ernvall, S. El Rouayheb, C. Hollanti, and H. V. Poor, "Capacity and security of heterogeneous distributed storage systems," *IEEE J. Sel. Areas Comm.*, 31(12), pp. 2701–2709, Dec. 2013.

[5] B. Sasidharan, and P. V. Kumar, "High-rate regenerating codes through layering," *2013 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1611–1615, 2013.

[6] C. Tian, V. Aggarwal, and V. A. Vaishampayan, "Exact-repair regenerating codes via layered erasure correction and block designs," *2013 IEEE Int. Symp. Inf. Theory (ISIT)* , pp. 1431–1435, 2013.

[7] T. Ernvall, "Codes between MBR and MSR points with exact repair property," *IEEE Trans. Inf. Theory*, 60(11), pp. 6993–7005, Nov. 2014.

[8] S. Goparaju, S. El Rouayheb, and R. Calderbank, "New codes and inner bounds for exact repair in distributed storage systems," *2014 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1036–1040, 2014.

[9] I. Tamo, Z. Wang, and J. Bruck, "MDS array codes with optimal rebuilding," *2011 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1240–1244, 2011.

[10] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the locality of codeword symbols," *IEEE Trans. Inf. Theory*, 58(11), pp. 6925–6934, 2012.

[11] F. Oggier and A. Datta, "Self-repairing homomorphic codes for distributed storage systems," *2011 IEEE INFOCOM*, pp. 1215–1223.

[12] D. S. Papailiopoulos, J. Luo, A. G. Dimakis, and C. Huang, J. Li "Simple regenerating codes: Network coding for cloud storage," *2012 IEEE INFOCOM*, pp. 2801–2805.

[13] D. S. Papailiopoulos, and A. G. Dimakis, "Locally repairable codes," *2012 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2771–2775.

[14] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis, "Optimal locally repairable codes and connections to matroid theory," *2013 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1814–1818.

[15] J. Simonis and A. Ashikhmin, "Almost affine codes", *Design, codes and cryptography*, 14, pp. 179–197, 1998.

[16] H. Whitney, 'On the abstract properties of linear dependence", *Amer. J. Math*, 57, pp. 509–533, 1935.

[17] C. Greene, 'Weight enumeration and the geometry of linear codes", *Stud. Appl. Math*, 55, pp. 119–128, 1976.

[18] F. J. MacWilliams 'A theorem on the distribution of weights in a systematic code', *Bell Syst. Tech J.*, 42, pp. 79–94, 1963.

[19] A. Barg, "The matroid supports of a linear code", *Appl. Algebra Engrg. Comm. Comput.*, 8, pp. 165–172, 1997.

[20] T. Britz, "Code enumerators and Tutte polynomials", *IEEE Trans. Inf. Theory*, 56, pp. 4350–4358, 2010.

[21] N. Kashyap, "A decomposition theory for binary linear codes", *IEEE Trans. Inf. Theory*, 54, pp. 3035–3058, 2008.

[22] R. Dougherty, C. Freiling, and K. Zeger, "Networks, matroids, and non-Shannon information inequalities", *IEEE Trans. Inf. Theory*, 53(6), pp. 1949–1969, 2007.

[23] J. Martí-Farré and C. Padró, "On secret sharing schemes, matroids and polymatroids", In S. Vadhan edt., *4th Theory of Crypt. Conf. TCC 2007, Lecture Notes in Computer Science*, vol. 4392, pp. 253–272, 2007.

[24] S. El Rouayheb, A.Sprintson, and C. Georghiades, 'On the index coding problem and its relation to network coding and matroid theory", *IEEE Trans. Inf. Theory*, 56(7), pp. 3187–3195, 2010.

[25] R. C. Singleton, "Maximum distance q-nary codes", *IEEE Trans. Inf. Theory*, 10, pp. 116–118, 1964.

[26] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar, "Optimal linear codes with a local-error-correction property", *2012 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2776 – 2780.

[27] N. Prakash, V. Lalitha, and P. V. Kumar, "Codes with locality for two erasures", *2014 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1962 – 1966.

[28] A. Wang, and Z. Zhang, "An integer programming-based bound for locally repairable codes" *IEEE Trans. Inf. Theory*, pp. 5280 – 5294, 2015.

[29] V. Cadambe and A. Mazumdar, "An upper bound on the size of locally recoverable codes", In *Proc. IEEE Symp. Netw. Coding*, pp. 1–5, Jun. 2013.

[30] A. S. Rawat, O. O. Koyluoglu, N. Silberstein, and S. Vishwanath, "Optimal locally repairable and secure codes for distributed storage systems", *IEEE Trans. Inf. Theory*, 60(1), pp. 212–236, 2014.

[31] N. Silberstein, A. S. Rawat, O. O. Koyluoglu, and S. Vishwanath, "Optimal locally repairable codes via rank-metric codes," *2013 IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1819–1823.

[32] W. Song, S. H. Dau, C. Yuen, and T. J. Li, "Optimal locally repairable linear codes",*IEEE J. Sel. Areas Comm.*, 32(5), pp. 1019–1036, 2014.

[33] I. Tamo and A. Barg, "A family of optimal locally recoverable codes" *IEEE Trans. Inf. Theory*, 60(8), pp. 4661–4676, 2014

[34] T. Britz, T. Johnsen, D. Mayhew, and K. Shiromoto, "Wei-type duality theorems for matroids", *Designs, Codes and Cryptography*, 62, pp. 331–341, 2012.

[35] T. Ernvall, T. Westerbäck, C. Hollanti, and R. Freij, "Constructions and properties of linear locally repairable codes," submitted, arXiv:1410.6339.

[36] B. Lindström, "On the vector representations of induced matroids" *Bull. London Math. Soc.*, 5, pp. 85–90, 1973.

[37] R. Stanley "Enumerative combinatorics, vol 1" *2:ed Cambridge University Press*, 2011.

[38] J. Oxley, "Matroid Theory" *Oxford Graduate Texts in Mathematics*, 3rd ed., Oxford University Press, 1992.

[39] F. Matúŝ, "Matroid representation by partitions", *Discrete Math.*, 203, pp. 169–194, 1999.

[40] J. E. Bonin and A. de Mier, "The lattice of cyclic flats of a matroid", *Annals of Combinatorics*, 12, pp. 155–170, 2008.

[41] K. Shoda, "Large families of matroids with the same Tutte polynomial", *PhD thesis, The Geroge Washington University*, August 2012.

[42] Govinda M. Kamath, N. Prakash, V. Lalitha, and P. Vijay Kumar, "Codes With Local Regeneration and Erasure Correction", *IEEE Trans. Inf. Theory*, 60(8), pp. 4637–4660, 2014.

[43] A. Schrijver, "Combinatorial Optimization: Polyhedra and Efficiency. Vol B: Matroids, Trees, Stable Sets" *Algorithms and Combinatorics 24*, Springer-Verlag, 2003.

[44] P. Hall, "On Representatives of Subsets", *J. London Math. Soc.*,10, pp. 26–30, 1935.

APPENDIX

Proof of Theorem III.2: For statement (i), we first claim that

$$\begin{aligned} d &= \min\{|X| : X \subseteq E \text{ and } \rho(E \setminus X) < k\} \\ &= n - \max\{|Y| : Y \subseteq E \text{ and } \rho(Y) < k\} \\ &= n - \max\{|Y| : Y \in \mathcal{F} \setminus \{E\}\} \\ &= n - k + 1 - \max\{\eta(Y) : Y \in \mathcal{F} \setminus \{E\}\} \\ &= n - k + 1 - \max\{\eta(Z) : Z \in \mathrm{coA}_{\mathcal{Z}}\} \end{aligned}$$

The first equality is the definition of d from Definition III.1. The third equality claims that the maximum is obtained when Y is a flat, which follows from Proposition II.2 (ix) and the fact that $Y \subseteq \mathrm{cl}(Y)$. By maximality, Y must have rank $\rho(Y) = k - 1$, which gives the forth equality. The fifth equality now follows directly from Proposition II.2 (iii).

For statement (ii), we first observe that S_x in Definition III.1 can be chosen to be cyclic, as we could otherwise find a smaller set with the same nullity and smaller size, by Proposition II.2 (viii). Statements a) and b) in this Theorem follows directly from Definition III.1. Finally, statement c) in this theorem follows by applying (i) to $M|S_x$, observing that $1_{M|S_x} = S_x$. ■

Proof of Lemma III.1:

First, let $\{S_x\}_{x \in E}$ be a collection of cyclic sets of M for which the statements a) - c) in Theorem III.2 (ii) are satisfied. We construct the chain $\{Y_j\}_{j=0}^{m}$ inductively by first letting $Y_0 = \emptyset$. Given $Y_{j-1} \subsetneq E$, we choose $x_j \in E \setminus Y_{j-1}$ arbitrarily, and assign $Y_j = \mathrm{cl}(Y_{j-1} \cup S_x)$. If $Y_j = E$, we set $m = j$.

Let j be any integer in $[m]$. We first observe that $\mathrm{cl}(S_j)$ is a cyclic flat, by Proposition II.2 (iv). Hence, by Proposition II.1 (iii), we see that inductively Y_j is a cyclic flat with

$$Y_j = \mathrm{cl}(Y_{j-1} \cup S_j) = \mathrm{cl}(Y_{j-1} \cup \mathrm{cl}(S_j)) = Y_{j-1} \vee \mathrm{cl}(S_j). \tag{13}$$

As $x_j \in Y_j \setminus Y_{j-1}$, we indeed have an increasing chain

$$C : 0_{\mathcal{Z}} = Y_0 \subsetneq Y_1 \subsetneq \ldots \subsetneq Y_m = E.$$

We remark as in (9) that $\rho(S_j) \leq r$ for any $j \in [m]$. Hence, by axiom (R3) in (5), we have

$$\rho(Y_j) = \rho(Y_{j-1} \cup S_j) \leq \rho(Y_{j-1}) + \rho(S_j) \leq \rho(Y_{j-1}) + r.$$

Moreover, by the statements (i) - (iii) in Proposition II.2 and Theorem III.2 (ii) c), we have

$$\begin{aligned}
\eta(Y_j) &= \eta(\mathrm{cl}(Y_{j-1} \cup S_j)) \\
&\geq \eta(Y_{j-1} \cup S_j) \\
&\geq \eta(Y_{j-1}) + \eta(S_j) - \eta(Y_{j-1} \cap S_j) \\
&\geq \eta(Y_{j-1}) + \eta(S_j) - \max\{\eta(X) : X \in coA_{\mathcal{Z}(M|S_j)}\} \\
&= \eta(Y_{j-1}) + d(M|S_j) - 1 \\
&\geq \eta(Y_{j-1}) + \delta - 1.
\end{aligned}$$

This concludes the proof. ■

Proof of Proposition III.2: For (i), let Y be any subset of E with $|Y| < \delta$. From Definition III.1, we conclude for every $x \in E$ that there is a subset $S_x \subseteq E$ with $x \in S_x$ and $\rho(S_x \setminus Y) = \rho(S_x)$.

Hence $\rho(E \setminus Y) = \rho(E)$, since every flat containing $S_x \setminus Y$ must contain S_x, and $E = \bigcup_{x \in E} S_x$. Consequently, from the definition $d = \min\{|X| : X \subseteq E,\ \rho(E \setminus X) < \rho(E)\}$, it follows that $\delta \leq d$.

For (ii), by (i) and Theorem III.3,

$$\delta \leq n - k + 1 - \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) (\delta - 1).$$

Therefore

$$k \leq n - \left\lceil \frac{k}{r} \right\rceil (\delta - 1).$$

For (iii), by (ii), we have

$$\frac{n}{k} \geq 1 + \left\lceil \frac{k}{r} \right\rceil \frac{(\delta - 1)}{k} \geq 1 + \frac{\delta - 1}{r} = \frac{r + \delta - 1}{r}.$$

■

Proof of Theorem III.4: Let

$$C : 0_{\mathcal{Z}} = Y_0 \subsetneq Y_1 \subsetneq \ldots \subsetneq Y_m = E, \tag{14}$$

be a chain of $(\mathcal{Z}(M), \subseteq)$ as given in Lemma III.1 (i), from a subset $\{S_j\}_{j \in [m]}$ of $\{S_x\}_{x \in E}$. Since d achieves the generalized Singleton bound in Theorem III.3, we get that $m = \lceil \frac{k}{r} \rceil$ and $\eta(Y_{m-1}) \leq (\lceil \frac{k}{r} \rceil - 1)(\delta - 1)$. Hence,

$$\eta(Y_j) = j(\delta - 1) \text{ for } j = 0, 1, \ldots, m-1, \quad \text{where} \quad m = \left\lceil \frac{k}{r} \right\rceil \geq 2, \tag{15}$$

by Lemma III.1 (iii) and the proof given for Theorem III.3.

For (i), observe that $\eta(0_{\mathcal{Z}}) = |0_{\mathcal{Z}}|$. Hence, if $0_{\mathcal{Z}} \neq \emptyset$, then $\eta(Y_0) = \eta(0_{\mathcal{Z}}) > 0$. This is a contradiction by (15).

To prove (ii), first observe that for any S_x we can select the chain in (14), such that $Y_1 = \mathrm{cl}(S_x)$. By (15), and since $\eta(X) \leq \eta(\mathrm{cl}(X))$ for any $X \subseteq E$, we get that

$$\delta - 1 = \eta(Y_1) \geq \eta(S_x).$$

Moreover, as we know from Theorem III.2 (iv) c),

$$d(M|S_x) \leq \delta \iff \min\{|X| : X \subseteq S_x,\, \rho(S_x \setminus X) < \rho(S_x)\} \geq \delta,$$

which implies that $\eta(S_x) \geq \delta - 1$, proving (ii) a).

To prove (ii) b), assume that S_x is not a cyclic flat. Then

$$\eta(Y_1) = \eta(\mathrm{cl}(S_x)) > \eta(S_x) = \delta - 1,$$

which contradicts (15). Thus,

$$0_{\mathcal{Z}(M|S_x)} = \emptyset,\, 1_{\mathcal{Z}(M|S_x)} = S_x \text{ and } \mathcal{Z}(M|S_x) = \{X \in \mathcal{Z}(M) : X \subseteq S_x\}$$

by Proposition II.2 (vii). Now suppose there were a cyclic flat $Z \in \mathcal{Z}(M)$ such that $\emptyset \subsetneq Z \subsetneq S_x$. Then $\rho(Z) < \rho(S_x)$ and $\eta(Z) > \eta(\emptyset) = 0$ by axiom (Z2) in Theorem II.1. Consequently, by Proposition II.2 (iii) and Theorem III.2 (iv) (c),

$$d(M|S_x) \leq \eta(S_x) + 1 - \eta(Z) \leq \delta - 1,$$

contradicting the (r, δ)-locality.

For (iii), we will first prove that any collection $F_1, \ldots, F_m$ of cyclic sets from $\{S_x : x \in E\}$ with a non-trivial union, and $m = \lceil \frac{k}{r} \rceil$, constitutes a chain as given in (14), with $Y_j = Y_{j-1} \vee F_j$ for $j = 1, \ldots, m$. Indeed, the chain in the proof of (14) is obtained by sequentially choosing an arbitrary S_x with $S_x \not\subseteq Y_j$, which can be chosen from $\{F_i\}$ as this is a collection with non-trivial union.

If $|Y_j \cap F_{j+1}| \geq \rho(F_{j+1})$, then we obtain that $\mathrm{cl}(Y_j \cap F_{j+1}) = F_{j+1} \subseteq \mathrm{cl}(Y_j) = Y_j$ by Proposition II.2 (x). This is a contradiction, and consequently

$$|Y_j \cap F_{j+1}| < \rho(F_{j+1}). \tag{16}$$

Now, by (ii) b) and Proposition II.1 (ii), any subset $X \subseteq S_x$ contains a circuit if and only if $|X| > \rho(S_x)$, *i.e.*, $\rho(X) = |X|$ if and only if $|X| \leq \rho(S_x)$. Consequently, $\eta(Y_j \cap F_{j+1}) = 0$. This implies, using Proposition II.2 (ii), (15) and statement (ii), that

$$\eta(Y_j \cup F_{j+1}) \geq \eta(Y_j) + \eta(F_{j+1}) - \eta(Y_j \cap F_{j+1}) = (j+1)(\delta - 1). \tag{17}$$

Furthermore by (15), if $j+1 \leq m-1$, then

$$\eta(Y_j \cup F_{j+1}) \leq \eta(\mathrm{cl}(Y_j \cup F_{j+1})) = \eta(Y_{j+1}) = (j+1)(\delta-1).$$

Hence, $Y_{j+1} = F_1 \cup \ldots \cup F_{j+1}$ if $j+1 \leq m-1$. If $Y_m \neq E$, then $\rho(Y_m) < \rho(E)$ and $\eta(Y_m) > (\lceil \frac{k}{r} \rceil - 1)(\delta-1)$. Then it follows, by Proposition II.2 (iii) and Theorem III.2 (iii), that

$$d < n-k+1-\left(\left\lceil \frac{k}{r} \right\rceil - 1\right)(\delta-1).$$

This is a contradiction. Consequently, $F_1, \ldots, F_m$ constitutes a chain as given in (14), with

$$Y_j = \bigvee_{i=1}^{j} F_i = \begin{cases} \bigcup_{i=1}^{j} F_i & \text{if} \quad j < \lceil \frac{k}{r} \rceil, \\ E & \text{if} \quad j = m = \lceil \frac{k}{r} \rceil. \end{cases} \tag{18}$$

For statement (iii) c), we first notice that the statement follows directly from (15) when $j < \lceil \frac{k}{r} \rceil$. When $j \geq m = \lceil \frac{k}{r} \rceil$ we conclude, using (18), that $\bigvee_{i=1}^{j} F_i = E$. Also, by (15),

$$n-k \geq \eta(F_1 \cup \ldots \cup F_m) \geq \left\lceil \frac{k}{r} \right\rceil (\delta-1).$$

Statement (iii) d) follows directly from (18), and statement (iii) e) is a immediately consequence of (iii) c)–d). Statement (iii) f) follows from (16) and (ii) a). ■

Proof of Theorem IV.1: We will show that $\mathcal{Z}$ and ρ define a matroid, by proving that the axioms (Z0)-(Z3) in Theorem II.1 are satisfied by $\mathcal{Z}$ and ρ. We let I, J be two subsets of $[m]$.

(Z0) Since the collection of sets $F_1, \ldots, F_m$ has a non trivial union by assumption (iv), it follows that $F_I \subsetneq F_J$ if and only if $I \subsetneq J$. Hence, we immediately get that $\mathcal{Z}$ is a lattice under inclusion with

$$F_I \wedge F_J = F_{I \cap J} \quad \text{and} \quad F_I \vee F_J = \begin{cases} F_{I \cup J} & \text{if} \quad F_{I \cup J} \in \mathcal{Z}_{<k}, \\ E & \text{if} \quad F_{I \cup J} \notin \mathcal{Z}_{<k}, \end{cases}$$

for $F_I, F_J \in \mathcal{Z}_{<k}$. Also, the bottom element in the lattice $0_{\mathcal{Z}}$ equals $\emptyset$ and by assumption (ii) in the top element $1_{\mathcal{Z}}$ equals E.

(Z1) Since $0_{\mathcal{Z}} = F_\emptyset$, we obtain that

$$\rho(0_{\mathcal{Z}}) = \rho(F_\emptyset) = 0.$$

(Z2) Since $F_I \subsetneq F_J$ if and only if $I \subsetneq J$, it is enough to prove that the axiom (Z2) holds for $F_I \subsetneq F_J$ in the following two cases:

$$(i) \quad F_J \in \mathcal{Z}_{<k} \text{ and } J = I \cup \{j\} \text{ with } j \in [m] \setminus I,$$
$$(ii) \quad F_I \in \mathcal{Z}_{<k} \text{ and } J = [m], \text{ i.e. } F_J = E.$$

In the first case, by the construction of ρ,

$$\begin{aligned}\rho(F_{I\cup\{j\}}) - \rho(F_I) &= |F_{I\cup\{j\}}| - \textstyle\sum_{l\in I\cup\{j\}} \eta(F_l) - (|F_I| - \sum_{i\in I}\eta(F_i)) \\ &= |F_j| - |F_j \cap F_I| - \eta(F_j) \\ &= \rho(F_j) - |F_j \cap F_I| \\ &> 0.\end{aligned}$$

Moreover, we have

$$\begin{aligned}&(|F_{I\cup\{j\}}| - \rho(F_{I\cup\{j\}})) - (|F_I| - \rho(F_I)) = \\ &\textstyle\sum_{l\in I\cup\{j\}} \eta(F_l) - \sum_{i\in I}\eta(F_i) = \\ &\eta(F_j) > 0.\end{aligned}$$

For case (ii), we immediately get that $\rho(E) - \rho(F_I) = k - \rho(F_I) > 0$. Now, we claim that for any $j \in [m] \setminus \{I\}$ with $F_{I\cup\{j\}} \notin \mathcal{Z}_{<k}$, we have

$$(|F_{I\cup\{j\}}| - k) - (|F_I| - \rho(F_I)) > 0. \tag{19}$$

By construction of $\mathcal{Z}_{<k}$,

$$(|F_{I\cup\{j\}}| - k) - (|F_I| - \rho(F_I)) \geq \sum_{l\in I\cup\{j\}} \eta(F_l) - \sum_{i\in I}\eta(F_i) = \eta(F_j) > 0.$$

Hence, by case (i) and (19), it follows

$$(|E| - \rho(E)) - (|F_I| - \rho(I)) > 0.$$

(Z3) Suppose that $F_I, F_J \in \mathcal{Z}_{<k}$. Then

$$\begin{aligned}&\rho(F_I) + \rho(F_J) - (\rho(F_I \vee F_J) + \rho(F_I \wedge F_J) + |(F_I \cap F_J) \setminus (F_I \wedge F_J)|) = \\ &\rho(F_I) + \rho(F_J) - \rho(F_I \vee F_J) - \rho(F_{I\cap J}) - |F_I \cap F_J| + |F_{I\cap J}| = \\ &|F_I| - \textstyle\sum_{i\in I}\eta(F_i) + |F_J| - \sum_{j\in J}\eta(F_j) - \rho(F_I \vee F_J) + \sum_{j\in I\cap J}\eta(F_j) - |F_I \cap F_J| = \\ &|F_I \cup F_J| - \textstyle\sum_{j\in I\cup J}\eta(F_j) - \rho(F_I \vee F_j) = \\ &|F_{I\cup J}| - \textstyle\sum_{j\in I\cup J}\eta(F_j) - \rho(F_I \vee F_j).\end{aligned}$$

If $F_{I\cup J} \in \mathcal{Z}_{<k}$, then

$$|F_{I\cup J}| - \sum_{j\in I\cup J}\eta(F_j) - \rho(F_I \vee F_j) = |F_{I\cup J}| - \sum_{j\in I\cup J}\eta(F_j) - (|F_{I\cup J}| - \sum_{j\in I\cup J}\eta(F_j)) = 0.$$

If $F_{I\cup J} \notin \mathcal{Z}_{<k}$, then

$$\begin{aligned}|F_{I\cup J}| - \textstyle\sum_{j\in I\cup J}\eta(F_j) - \rho(F_I \vee F_j) &= |F_{I\cup J}| - \textstyle\sum_{j\in I\cup J}\eta(F_j) - k \\ &\geq |F_{I\cup J}| - \textstyle\sum_{j\in I\cup J}\eta(F_j) - (|F_{I\cup J}| - \sum_{j\in I\cup J}\eta(F_j)) \\ &= 0.\end{aligned}$$

Moreover, for E and F_I we have that

$$\rho(F_I) + \rho(E) - (\rho(F_I \vee E) + \rho(F_I \wedge E) + |(F_I \cap E) \setminus (F_I \wedge E)|) =$$
$$\rho(F_I) + \rho(E) - \rho(E) - \rho(F_I) + |F_I| - |F_I| = 0.$$

We have now proven that the axioms (Z0)–(Z3) in Theorem II.1 are satisfied by $\mathcal{Z}$ and ρ. Hence, $\mathcal{Z}$ and ρ define a matroid $M = M(F_1, \ldots, F_m; k; \rho)$ over E.

The parameters (n, k, d, r, δ) will now be investigated using Theorem III.2. Firstly, the parameters (n, k, d, r, δ) are defined for M with $n = |1_{\mathcal{Z}}| = |E|$ and $k = \rho(1_{\mathcal{Z}}) = \rho(E)$, since $E \in \mathcal{Z}$ and $\rho(E) = k > 0$. By Axiom $(Z2)$ in Theorem II.1, $\eta(Y) > \eta(X)$ for all $X, Y \in \mathcal{Z}$ when $X \subsetneq Y$. Hence, by Theorem III.2 (iii),

$$\begin{aligned} d &= n - k + 1 - \max\{\eta(F_I) : F_I \in \mathcal{Z}_{<k}\} \\ &= n - k + 1 - \max\{|F_I| - (|F_I| - \textstyle\sum_{i \in I} \eta(F_i)) : F_I \in \mathcal{Z}_{<k}\} \\ &= n - k + 1 - \max\{\textstyle\sum_{i \in I} \eta(F_i) : F_I \in \mathcal{Z}_{<k}\}. \end{aligned}$$

Let $\delta - 1 = \min_{i \in [m]}\{\eta(F_i)\}$ and S be a subset of F_i such that $|S| = \rho(F_i) + \delta - 1$. By construction and Proposition II.2 (vii),

$$\mathcal{Z}(M|F_i) = \{Z \in \mathcal{Z}(M) : Z \subseteq F_i\} = \{\emptyset, F_i\}.$$

Hence, from Proposition II.1 (i) and (ii),

$$\rho(X) = \{|X|, \rho(F_i)\} \text{ for } X \subseteq F_i$$

and

$$\mathcal{C}(M) \cap F_i = \{X \subseteq F_i : |X| = \rho(F_i) + 1\}.$$

This implies that S is a cyclic set and that

$$d(M|S) = |S| - \rho(S) + 1 = \rho(F_i) + \delta - 1 - \rho(F_i) + 1 = \delta.$$

by Definition III.1 (iv) c). Therefore, with $r = \max_{i \in [m]}\{\rho(F_i)\}$ and as $F_{[m]} = E$, statements (iv) a)–c) in Theorem III.2 are satisfied. Consequently, M has (r, δ)-locality and S is a locality set.

It remains to show that the independent sets $\mathcal{I}(M)$ equals $\mathcal{I}$. We first point out that

$$|F_J| - \sum_{j \in J} \eta(F_j) > |F_I| - \sum_{i \in I} \eta(F_i),$$

for $I \subsetneq J \subseteq [m]$. Noting that $X \subseteq E$ is independent if and only if X does not contain any circuits, and applying Proposition II.1 (ii), we get

$$\begin{aligned}\mathcal{I}(M) &= \{X \subseteq E : |X| \leq \rho(Y) \text{ for all } Y \in \mathcal{Z}\} \\ &= \{X \subseteq E : |F_I \cap X| \leq \min\{|F_I| - \textstyle\sum_{i\in I} \eta(F_i), k\} \text{ for all } I \subseteq [m]\} \\ &= \mathcal{I}.\end{aligned}$$

■

Proof of Theorem IV.2: To prove the theorem, we will first show that the assumptions (i)–(iv) in Section IV-A1 are satisfied by $(F_1, \ldots, F_m; k; \rho)$, obtained from the graph (G, γ) in Section IV-A2. We will then show that the values of the parameters (n, k, d, r, δ) of $M(F_1, \ldots, F_m; k; \rho)$ are the ones requested in Theorem IV.2.

Statement IV-A1 (i) follows directly from IV-A2 (ii) and (iii). IV-A1 (ii) is obvious. For IV-A1 (iii), we first notice that by IV-A2 (iv) and (vi), we can define γ as the size of a nonempty intersection of two sets F_i and F_j Hence, as $F_h \cap F_i \cap F_j = \emptyset$ for all $h, i, j \in [m]$, we know that

$$|F_{[m]}| = \sum_{i\in[m]} |F_i| - \sum_{w\in W} \gamma(w). \tag{20}$$

Moreover, for $i \in [m]$, we have

$$\eta(F_i) = |F_i| - \rho(F_i) = \delta - 1 + \beta(i).$$

Consequently,

$$\begin{aligned}|F_{[m]}| - \textstyle\sum_{i\in[m]} \eta(F_i) &= \textstyle\sum_{i\in[m]} |F_i| - m(\delta - 1) - \sum_{i\in[m]} \beta(i) - \sum_{w\in W} \gamma(w) \\ &= mr - \textstyle\sum_{i\in[m]} \alpha(i) - \sum_{w\in W} \gamma(w).\end{aligned}$$

Therefore, by IV-A2 (v), IV-A1 (iii) holds. For IV-A1 (iv), we first remark that

$$F_{[m]\setminus i} \cap F_i = \sum_{w=\{i,j\}\in W} \gamma(w)$$

and $\rho(F_i) = r - \alpha(i)$ for $i \in [m]$. Hence IV-A1 (iv) holds, by IV-A2 (vi).

We will now determine the parameters (n, k, d, r, δ), proving that they agree for the graph and the matroid. The given parameters (r, δ) for the graphs also give (r, δ)-locality of the matroid as $\rho(F_i) \leq r$ and $\eta(F_i) \geq \delta - 1$ by (IV-A2) (ii) and (iii), and IV-A2 (i) and (ii). We have already proven that the parameter k of the graph is the rank of the matroid. Moreover, by (20),

$$n = |F_{[m]}| = |\sum_{i\in[m]} |F_i| - \sum_{w\in W} \gamma(w) = (r + \delta - 1)m - \sum_{i\in[m]} \alpha(i) + \sum_{i\in[m]} \beta(i) - \sum_{w\in W} \gamma(w).$$

The statement about d in Theorem IV.2 (ii) holds as a consequence of Theorem IV.1 (iii) and the properties that

$$\sum_{i\in I}\eta(F_i) = |I|(\delta-1) + \sum_{i\in I}\beta(i)$$

and

$$|F_I| - \sum_{i\in I}\eta(F_i) = r|I| - \sum_{i\in I}\alpha(i) - \sum_{w\subseteq I, w\in W\in I}\gamma(w).$$

This concludes the proof. ■

Proof of Theorem IV.4: We will divide the proof of Theorem IV.4 into the the parts (i)–(v). First, we recall that a and b are the integers where

$$k = \left\lceil \frac{k}{r} \right\rceil r - a \text{ and } n = \left\lceil \frac{n}{r+\delta-1} \right\rceil (r+\delta-1) - b.$$

Proof of Theorem IV.4 (i): We will mimic the proof of Theorem IV.3, using the assumption that $a \geq b$ to tighten the bounds. Hence, let $m = \left\lceil \frac{n}{r+\delta-1} \right\rceil$, and let $F_1, \ldots F_{m-1}$ be disjoint sets with rank r and size $r+\delta-1$. Let F_m be disjoint from all of $F_1, \ldots F_{m-1}$, with size

$$|F_m| = n - (m-1)(r+\delta-1) = r+\delta-1-b$$

and rank $\rho(F_m) = |F_m| - \delta + 1 = r - b$. Finally, let M be defined by $\mathcal{Z}(M) = \{F_I\}$, where $F_I = \cup_{i\in I} F_i$, and

$$\rho(F_I) = \min\{\sum_{i\in I}\rho(F_i), k\}.$$

Now, the union of any $\lceil \frac{k}{r} \rceil - 1$ sets among $F_1, \ldots, F_{m-1}$, together with F_m, has rank

$$r\left\lceil \frac{k}{r} \right\rceil - b = k + a - b \leq k.$$

Thus we have $\max\{|I| : \rho(F_I) < k\} = \lceil \frac{k}{r} \rceil$, so M has minimum distance

$$d = n - k + 1 - (\delta-1)\max\{|I| : \rho(F_I) < k\} = n - k - (\left\lceil \frac{k}{r} \right\rceil - 1)(\delta-1).$$

■

Proof of Theorem IV.4 (ii): We will use Theorem IV.1 to construct a (n, k, d, r, δ)-matroid with

$$d = n - k + 1 - \left\lceil \frac{k}{r} \right\rceil (\delta-1) + (b-r),$$

where

$$b = (r+\delta-1)\left\lceil \frac{n}{r+\delta-1} \right\rceil - n > a = r\left\lceil \frac{k}{r} \right\rceil - k.$$

For this purpose, let $m = \left\lceil \frac{n}{r+\delta-1} \right\rceil - 1 = \lceil \frac{k}{r} \rceil + t$, where $t \geq 0$ by Proposition IV.1. Let $F_1, \dots F_{m-1}$ be disjoint sets with rank r and size $r+\delta-1$. Let F_m be disjoint from all of $F_1, \dots F_{m-1}$, with size

$$|F_m| = n - (m-1)(r+\delta-1) = 2(r+\delta-1) - b$$

and rank $\rho(F_m) = r$. Finally, let M be defined by $\mathcal{Z}(M) = \{F_I\}$, where $F_I = \cup_{i \in I} F_i$, and

$$\rho(F_I) = \min\{\sum_{i \in I} \rho(F_i), k\}.$$

Clearly, this has the desired values of r and n. To guarantee that M has rank $\rho(M) = \rho(F_{[m]}) = k$ is the rank function of a matroid, we verify that

$$k = \left\lceil \frac{k}{r} \right\rceil r - a = (m-t)r \leq mr = \sum_{i \in [m]} \rho(F_i).$$

Statements IV-A2 (vi) follows as $r - \alpha(i) \geq 1$ for $i \in [m]$ and as G has no edges. Now, by Theorem IV.2, there is an (n, k, d, r, δ)-matroid with

$$n = (r+\delta-1)m + (r+\delta-1-b) = (r+\delta-1)\left\lceil \frac{n}{r+\delta-1} \right\rceil - b.$$

Moreover, by Theorem IV.1,

$$d = n - k - 1 - \left(\left\lceil \frac{k}{r} \right\rceil (\delta-1) - (b-r) \right),$$

as

$$\max_{I \in V_{<k}} \{(\delta-1)|I|\} = \left(\left\lceil \frac{k}{r} \right\rceil - 1 \right) + (r+\delta-1-b)$$

for

$$V_{<k} = \{I \subseteq [m] : r|I| < k\}.$$

This concludes the proof. ∎

Proof of Theorem IV.4 (iii), right implication: By the structure theorem III.4, we see that the existence of an (n, k, d, r, δ)-matroid with $d = n - k + 1 - \left(\lceil \frac{k}{r} \rceil - 1\right)(\delta - 1)$ implies the existence of subsets $F_1, \dots, F_m$ of $[n]$ such that:

(i) $F_j \not\subseteq \bigcup_{i \in [m] \setminus \{j\}} F_i$ for $j = 1, \dots, m$,

(ii) $|F_i| \leq r + \delta - 1$ for $i = 1, \dots, m$,

(iii) $|\bigcup_{i \in [m]} F_i| = n$,

(iv) $|F \cap (\bigcup_{i \in I} F_i)| \leq |F| - \delta$ for every $F \in \{F_i\}_{i \in [m] \setminus I}$ with $I \subseteq [m]$ and $|I| < \lceil \frac{k}{r} \rceil$,

(v) $|\bigcup_{i \in I} F_i| - |I|(\delta - 1) \geq k = \lceil \frac{k}{r} \rceil r - a$ for every $I \subseteq [m]$ with $|I| \geq \lceil \frac{k}{r} \rceil$.

For simplicity, denote $\lceil \frac{k}{r} \rceil = h$. For any set system $F_1, \ldots, F_m$ where $|F_i| \leq r + \delta - 1$ for every$i$, construct a graph $\mathcal{G}$ on vertex set $[m]$, with an edge between i and j if and only if $F_i \cap F_j \neq \emptyset$. Note that, when $I \subseteq [m]$ is such that the induced subgraph $\mathcal{G}[I]$ on I is connected, then

$$|F_I| \leq (r + \delta - 2)|I| + 1.$$

If $\mathcal{G}[I]$ is connected and equality holds in the above inequality, then I is said to be *full*. Note that for every full component I in $\mathcal{G}$ and integer $1 \leq u \leq |I|$, there is a subset $I' \subseteq I$ such that $|I'| = t$ and I' is full. Denoting by $c(\mathcal{G}[I])$ the number of full components of $\mathcal{G}[I]$, we get

$$|F_I| \leq (r + \delta - 2)|I| + c(\mathcal{G}[I]).$$

Let J be the union of the $h-a-1$ largest full components of $\mathcal{G}$ together with all non-full components of $\mathcal{G}$. If $|J| \geq h$, then we have a subset of nodes $J' \subseteq J$ with $|J'| = h$, such that $c(\mathcal{G}[J']) \leq h-a-1$. However, assuming

$$|F_I| \geq h(r + \delta - 1) - a = h(r + \delta - 2) + h - a$$

for every subset $I \subseteq [m]$ with $|I| = h$, then $c(\mathcal{G}[J']) \geq h - a$. Hence, $|J| \leq h - 1$ and $\mathcal{G}[[m] \setminus J]$ is a union of full components $I_1, \ldots, I_s$ of $\mathcal{G}$, and these full components contain at most $\left\lfloor \frac{h-1}{h-1-a} \right\rfloor$ nodes each.

When bounding $\left\lceil \frac{n}{r+\delta-1} \right\rceil$, we first notice that

(i) $|I|(r + \delta - 1) - |F_I| = |I| - 1$ if I is connected and full,

(ii) $|I|(r + \delta - 1) - |F_I| \geq |I|$ if I is connected and not full,

(iii) $h(r + \delta - 1) - |F_I| \leq a$ if $|I| \leq h$,

(iv) $|I_i|(r + \delta - 1) - |F_{I_i}| \leq \left\lfloor \frac{h-1}{h-1-a} \right\rfloor - 1$ for $1 \leq i \leq s$.

Hence,

$$\begin{aligned} b &= m(r + \delta - 1) - |F_{[m]}| \\ &= |J|(r + \delta - 1) - |F_J| + \textstyle\sum_{i=1}^{s} |I_i|(r + \delta - 1) - |F_{I_i}|. \end{aligned}$$

Also, as $|J| < h$, we get

$$|J|(r + \delta - 1) - |F_J| + \sum_{i=1}^{s} |I_i|(r + \delta - 1) - |F_{I_i}| \leq a + s\left(\left\lfloor \frac{h-1}{h-1-a} \right\rfloor - 1 \right).$$

Hence, as $b > a$, we have $\left\lfloor \frac{h-1}{h-1-a} \right\rfloor \geq 2$, or equivalently $a \geq \lceil \frac{h}{2} \rceil$. Now, assume that $a \geq \lceil \frac{h}{2} \rceil$. For the cardinality of $F_{[m]}$ we have that

$$|F_{[m]}| = (|J| + \sum_{i=1}^{s} |I_i|)(r + \delta - 1) - b.$$

Using (iii), (iv) and the property that $|J| < h$, we now obtain that

$$\left\lceil \frac{|F_{[m]}|}{r+\delta-1} \right\rceil \geq h-1+\left\lfloor \frac{b-a}{\left\lfloor \frac{h-1}{h-1-a} \right\rfloor - 1} \right\rfloor \left\lfloor \frac{h-1}{h-1-a} \right\rfloor + t, \tag{21}$$

where

$$t = \begin{cases} 0 \text{ if } (\left\lfloor \frac{h-1}{h-1-a} \right\rfloor - 1) | (b-a) \\ (b-a) - \left\lfloor \frac{b-a}{\left\lfloor \frac{h-1}{h-1-a} \right\rfloor - 1} \right\rfloor \left(\left\lfloor \frac{h-1}{h-1-a} \right\rfloor - 1 \right) + 1 \text{ otherwise.} \end{cases}$$

Rearranging equation (21), we find the bound

$$\left\lceil \frac{n}{r+\delta-1} \right\rceil \geq \left\lceil \frac{k}{r} \right\rceil - 1 + (b-a)\left(1+\frac{1}{t}\right), \tag{22}$$

where

$$t = \lfloor a/\left(h-1-a\right) \rfloor = \left\lfloor a/\left(\left\lceil \frac{k}{r} \right\rceil - 1 - a \right) \right\rfloor.$$

■

Construction 3: To prove Theorem IV.4 (iii), we will construct graphs (G, γ) that satisfy the assumptions in IV-A2 with (k, r, δ), and then use Theorem IV.1. For simplicity, denote

$$s = \left\lfloor \frac{\lceil \frac{k}{r} \rceil - 1}{\lceil \frac{k}{r} \rceil - 1 - a} \right\rfloor, \; u = \left\lceil \frac{k}{r} \right\rceil - 1 - a + \left\lceil \frac{b-a}{s-1} \right\rceil \text{ and } x = \left\lceil \frac{k}{r} \right\rceil - 1 - s\left(\left\lceil \frac{k}{r} \right\rceil - 1 - a \right)$$

Let

(i) $m \geq \lceil \frac{k}{r} \rceil - 1 + (b-a)\left(1+\frac{1}{t}\right)$, where $t = \lfloor a/\left(\lceil \frac{k}{r} \rceil - 1 - a \right) \rfloor$,

(ii) G be the graph consisting of vertex-disjoint paths $P_1, \ldots, P_u$ with

$$|P_i| = \begin{cases} s+1 & \text{if} \quad 1 \leq i \leq x, \\ s & \text{if} \quad x+1 \leq i \leq u-1, \\ s & \text{if} \quad i = u \text{ and } s-1 \mid b-a, \\ b-a-\lfloor \frac{b-a}{s-1} \rfloor (s-1)+1 & \text{if} \quad i = u \text{ and } s-1 \nmid b-a, \end{cases} \tag{23}$$

(iii) $\gamma(w) = 1$ for each $w \in W$.

Proof of Theorem IV.4 (iii), left implication: : We first note that statement IV-A2 (i) follows directly as G has no cycles. Statement IV-A2 (ii) is a consequence of (23) (iii). For statement IV-A2 (iii), we first remark that by (21) and (22), we get

$$\sum_{w \in W} \gamma(w) = |W| = (\sum_{i \in u} |P_i|) - u = (s+1)x + (u-1-x)s + |P_u| - u = b.$$

Statement IV-A2 (iv) follows directly from (23) (i). Statement IV-A2 (v) follows from the fact that

$$\sum_{i=1}^{\lceil \frac{k}{r} \rceil - 1 - a} |P_i| = \left\lceil \frac{k}{r} \right\rceil - 1$$

and

$$\sum_{i,j\in P, w=\{i,j\}\in W} \gamma(w) = \left\lceil \frac{k}{r} \right\rceil - 1 - (\left\lceil \frac{k}{r} \right\rceil - 1 - a) = a,$$

where $P = \bigcup_{1\leq i\leq \lceil \frac{k}{r} \rceil - 1 - a} P_i$. Finally, statement IV-A2 (vi) follows from the property that $\gamma(w) = 1$ for all $w \in W$. The result now follows using (21) and (22), which imply that

$$|\bigcup_{i=1}^{u} P_i| = \left\lceil \frac{k}{r} \right\rceil - 1 + (b-a)\left(1 + \frac{1}{t}\right),$$

where $t = \left\lfloor a/\left(\left\lceil \frac{k}{r} \right\rceil - 1 - a\right) \right\rfloor$. ■

Construction 4: To prove Theorem IV.4 (iv), we will construct graphs $G = G(\gamma; k, r, \delta, a, b)$ that satisfy the statements in Corollary IV.2, and then use Theorem IV.1. For simplicity, denote

$$s = \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor,\ t = \left\lfloor \frac{r-1}{s} \right\rfloor,\ u = \left\lceil \frac{\lceil \frac{k}{r} \rceil + 1}{2} \right\rceil \text{ and } x = \left\lceil \frac{b - \lfloor \frac{b}{stu} \rfloor stu}{s} \right\rceil.$$

Before we are ready to construct G, we need some subgraphs that will be the building blocks of G. For $1 \leq i \leq t$, let P_i denote a path containing $u+1$ vertices, with p_i as start vertex and q_i as end vertex. Now, let B denote the graph obtained from $\sqcup_i P_i$ by identifying all p_i to the same vertex $p \in B$, all all the end vertices q_i the same vertex $q \in B$

We will now define a subgraph of $B'(h)$ of B, where h denotes the number of edges that the subgraph should have. First we remark that the number of edges in B equals tu. Now, order the edges in B from 1 to tu by starting from the start vertex p and ending in the end vertex q for each path, ordering the edges path by path from P_1 to P_t. This is

- the path P_i is the sequence of vertices $p = v_1^{(i)}, v_2^{(i)}, \ldots v_u^{(i)}, v_{u+1}^{(i)} = q$, then edge $\{v_j^{(i)}, v_{j+1}^{(i)}\}$ is ordered as edge number $(i-1)u + j$.

The subgraph $B'(h)$ is now defined as the subgraph of B that consists of the edges numbered from 1 to x and the vertices associated to these edges. By $B'(0)$ we mean the graph with no vertices.

The number of vertices of B equals the number of internal nodes in paths $P_1, \ldots, P_t$ plus 2, *i.e.*,

$$t(u-1) + 2.$$

Moreover, the number of vertices in $B'(h)$, when $h \neq 0$, equals

$$\left\lfloor \frac{h}{u} \right\rfloor (u-1) + (h - \left\lfloor \frac{h}{u} \right\rfloor u) + (1 + \min\{\left\lfloor \frac{h}{u} \right\rfloor, 1\}) = h - \left\lfloor \frac{h}{u} \right\rfloor + 1 + \min\{\left\lfloor \frac{h}{u} \right\rfloor, 1\}.$$

Now, for the construction of G, let

(i) $m \geq \left\lfloor \frac{b}{stu} \right\rfloor (t(u-1)+2) + y$, where

$$y = \begin{cases} 0 & \text{if} \quad stu \mid b, \\ x - \left\lfloor \frac{x}{u} \right\rfloor + 1 + \min\{\left\lfloor \frac{x}{u} \right\rfloor, 1\} & \text{if} \quad stu \nmid b; \end{cases}$$

(ii) G be the graph with vertices $[m]$ and edges W, where G consists of $\left\lfloor \frac{b}{stu} \right\rfloor$ copies of B, one copy of $B'(x)$ and possibly some additional isolated vertices; (24)

(iii)
- If $s \mid b$ then $\gamma(w) = s$ for all $w \in W$,
- If $s \nmid b$ then

$$\gamma(w) = \begin{cases} s & \text{if} \quad w \text{ is not the vertex number x in } B'(x), \\ b - \lfloor \frac{b}{s} \rfloor s & \text{if} \quad w \text{ is the vertex number x in } B'(x); \end{cases}$$

Proof of Theorem IV.4 (iv): As $\lceil \frac{k}{r} \rceil \geq 3$, and the smallest size of a cycle in the graph is $2u \geq \lceil \frac{k}{r} \rceil + 1$, it follows that G has no l-cycles for $l \leq \max\{3, \lceil \frac{k}{r} \rceil\}$. Also, the property that $1 \leq \gamma(w) \leq \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor$ for all edges w in G follows from (24) (iii) as $s = \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor$. For statement IV-A2 (iii), we remark that for a copy of B in G, the total sum of $\gamma(w)$ for all edges w in B equals stu. Moreover, the total sum of $\gamma(w)$ for all edges w in $B'(x)$ equals sx if $s|b$, and $s(x-1)+b-\lfloor \frac{b}{s} \rfloor s$ if $s \nmid b$. Hence

$$\sum_{w \in W} \gamma(w) = \left\lfloor \frac{b}{stu} \right\rfloor \sum_{edges \ \ w \in B} \gamma(w) + \sum_{edges \ \ w \in B'(x)} \gamma(w) = b.$$

Statement IV-A2(vi) follows as $ts \leq r-1$ and $2s \leq 2\left\lfloor \frac{a}{2} \right\rfloor \leq a \leq r-1$. Hence, by Corollary IV.2 and Theorem IV.1, the theorem is now proven. ■

Construction 5 when $2a \leq r-1$*:* We will construct graphs (G, γ) that satisfy the statements in IV-A2 and then use Theorem IV.1. To construct G, let

(a) $m \geq \lceil \frac{b}{a} \rceil + 1$;

(b) G be the graph with vertices $[m]$ and edges $W = \{\{i, i+1\} : 1 \leq i \leq \lceil \frac{b}{a} \rceil\}$;

(c) For $\{i, i+1\} \in W$ let, (25)

$$\gamma(\{i, i+1\}) = \begin{cases} a & \text{if} \quad i < \lceil \frac{b}{a} \rceil, \\ a & \text{if} \quad i = \lceil \frac{b}{a} \rceil \text{ and } a|b, \\ b - \lfloor \frac{b}{a} \rfloor a & \text{if} \quad i = \lceil \frac{b}{a} \rceil \text{ and } a \nmid b. \end{cases}$$

Proof of Theorem IV.4(v) when $2a \leq r-1$: That G has no l-cycles for $l \leq \max\{3, \lceil \frac{k}{r} \rceil\}$ follows directly as G has no cycles. Also, that $1 \leq \gamma(w) \leq \left\lfloor \frac{a}{\lceil \frac{k}{r} \rceil - 1} \right\rfloor$ for all edges w in G follows

directly from (25) (c). For statement IV-A2 (iii), we obtain from (25) (c) that

$$\sum_{w\in W}\gamma(w)=\left(\frac{b}{a}-1\right)a+a=b \text{ if } a|b,$$

and

$$\sum_{w\in W}\gamma(w)=\left\lfloor\frac{b}{a}\right\rfloor a+b-\left\lfloor\frac{b}{a}\right\rfloor a=b \text{ if } a\nmid b.$$

As the maximal number of neighbors of a vertex in G is 2, we get that for any $i\in[m]$,

$$\sum_{w=\{i,j\}\in W}\gamma(w)\leq 2a\leq r-1.$$

Hence, by Corollary IV.2 and Theorem IV.1, the theorem is now proven. ■

Construction 5 when $2a>r-1$*:* In order to prove Theorem IV.4(v), we will construct graphs (G,γ) that satisfy the statements in Corollary IV.2, and then use Theorem IV.1. For simplicity, denote $h=\lfloor\frac{r-1}{2}\rfloor$. Now, to construct G, let

(a) $m\geq\lceil\frac{b}{h}\rceil+1$;

(b) G be the graph with vertices $[m]$ and edges $W=\{\{i,i+1\}:1\leq i\leq\lceil\frac{b}{h}\rceil\}$;

(c) For $\{i,i+1\}\in W$, let

$$\gamma(\{i,i+1\})=\begin{cases}h & \text{if } \ i<\lceil\frac{b}{h}\rceil,\\ h & \text{if } \ i=\lceil\frac{b}{h}\rceil \text{ and } h|b,\\ b-\lfloor\frac{b}{h}\rfloor h & \text{if } \ i=\lceil\frac{b}{h}\rceil \text{ and } h\nmid b.\end{cases}$$

Proof of Theorem IV.4(v) when $2a>r-1$: The proof is completely analogous to the proof Theorem IV.4(v) when $2a\leq r-1$, replacing a by h. ■

Proof of Lemma V.1: We want to prove that the matroid $M(\mathbf{F})$ obtained from the set system $\mathbf{F}$ in Theorem IV.1 is isomorphic to the gammoid $M(G)$ associated to the graph G in Algorithm 1. We will proceed by proving that the independent sets $\mathcal{I}(M(\mathbf{F}))$ and $\mathcal{I}(M(G))$ are equal.

The independent sets of $M(G)$ are

$$\mathcal{I}(M(G))=\{X\subseteq E:\exists \text{ a set of } \ |X| \text{ vertex-disjoint paths from } X \text{ to } T\}.$$

As each path from E to T goes through the complete bipartite graph between H and T, with $|T|=k$, we can equivalently write

$$\mathcal{I}(M(G))=\{X\subseteq E:\exists \text{ a matching of size } |X| \text{ between } X \text{ and } H, \text{ and } |X|\leq k\}.$$

By Theorem IV.1, the independent sets of the matroid $M(\mathbf{F})$ are

$$\begin{aligned}\mathcal{I}(M(\mathbf{F})) &= \{X \subseteq E : |X \cap F_I| \leq \min\{|F_I| - \textstyle\sum_{i\in I} \eta(F_i), k\} \text{ for each } I \subseteq [m]\} \\ &= \{X \subseteq E : |X \cap F_I| \leq |F_I| - \textstyle\sum_{i\in I} \eta(F_i) \text{ for each } I \subseteq [m], \text{ and } |X| \leq k\}\end{aligned}$$

Hence, it remains to show that there is a matching of size $|X|$ in G between X and H, if and only if

$$|X \cap F_I| \leq |F_I| - \sum_{i\in I} \eta(F_i) \text{ for each } I \subseteq [m]. \tag{26}$$

By Halls theorem [44], a bipartite graph (U, V, E) has a matching of size U if and only if $|N(A)| \geq |A|$ for every $A \subseteq U$. Here, U and V are the two parts of the bipartition, and

$$N(A) = \{x \in V : \exists u \in A \text{ with } ux \in E\}$$

is the neighborhood of A.

Assume that there is a matching of size $|X|$ in G between X and H. Then, in particular, $X \cap F_I$ has at least $|X \cap F_I|$ neighbors in H, for every $I \subseteq [m]$. But all these neighbors $v \in H$ must have $h(v) \cap I \neq \emptyset$, by construction of G. Now as

$$|\{v \in H : i \in h(v)\}| = \rho(F_i) = |F_i| - \eta(F_i)$$

and

$$|\{v \in H : \{i, j\} \subseteq h(v)\}| = |F_i \cap F_j|,$$

it follows by induction on $|I|$ that

$$|\{u \in H : h(u) \cap I \neq \emptyset\}| = |F_I| - \sum_{i\in I} \eta(F_i).$$

This number of neighbors must be at least $|X \cap F_I|$, wherefore (26) holds.

Assume, on the other hand, that (26) holds, and let $A \subseteq X$ be an arbitrary subset of X. To apply Hall's Theorem, we need to prove that $|N(A)| \geq |A|$.

Write $A = A' \cup A''$ where $A' = \{u \in A : |s(u)| = 1\}$ and $A'' = \{u \in A : |s(u)| \geq 2\}$ respectively. For $x \in A''$, by (7–9) in Algorithm1, there is a node $u_x \in H$ for which $(\overrightarrow{x, u_x}) \in D$. Consequently, for $H'' = \{u_x \in H : x \in X''\}$, we have

$$|H''| = |\{u_x \in H : x \in X''\}| = |X''|. \tag{27}$$

Moreover, for $x \in A$, let $H_x = \{u \in H : (\overrightarrow{x, u}) \in D\}$. By construction,

$$|H_x| = |H_{s(x)}| = |F_{s(x)}| - \eta(F_{s(x)}).$$

Hence, for $H' = \{u : \exists x \in A' \text{ such that } (\overrightarrow{x,u}) \in D\}$ and $I' = \{s(x) : x \in A'\}$, we get

$$|H'| = |H_{I'}| = |F_{I'}| - \sum_{i \in I'} \eta(F_i). \tag{28}$$

Since $A \subset X$, and X satisfies (26), we know by (27) and (28), we now obtain that $|A| \leq |H' \cup H''| \leq |N(A)|$. As $A \subset X$ was chosen arbitrarily, we can apply Hall's theorem to the effect that there is a matching between X and H of size $|X|$. This concludes the proof. ■