Minimizing the Alphabet Size in Codes with Restricted Error Sets

Mira Gonen, Michael Langberg, Alex Sprintson

***Abstract*—This paper focuses on error-correcting codes that can handle a predefined set of specific error patterns. The need for such codes arises in many settings of practical interest, including wireless communication and flash memory systems. In many such settings, a smaller field size is achievable than that offered by MDS and other standard codes. We establish a connection between the minimum alphabet size for this generalized setting and the combinatorial properties of a hypergraph that represents the prespecified collection of error patterns. We also show a connection between error and erasure correcting codes in this specialized setting. This allows us to establish bounds on the minimum alphabet size and show an advantage of non-linear codes over linear codes in a generalized setting. We also consider a variation of the problem which allows a small probability of decoding error and relate it to an approximate version of the hypergraph coloring problem.**

I. Introduction

In many practical settings, there is a need to design error-correcting codes that can handle specific error patterns. For example, in wireless communications, magnetic recording, flash memory systems, and Dynamic Random-Access Memories (DRAMs) the errors can appear in correlated locations such as bursts, single-row errors, or crisscrosss errors, e.g., [1]–[6]. These settings benefit from customized error correcting codes, that may improve on the best known parameters of standard error correcting codes. For example, the optimal error-correcting capabilities of the classical (n,k) Maximum Distance Separable (MDS) code, such as the Reed-Solomon code, come at the price of a significant alphabet size of $q \geq n-k+1$, [7].[1] As we show in this paper, in many settings with specific error patters, a much smaller alphabet size is needed.

In this work, we present a general framework for code design that can handle any possible collection of predefined error patterns. Our framework applies to both linear and non-linear codes. For an error-correcting code of length n, we use an n-vertex hypergraph G to represent the given collection of error sets. Specifically, nodes of G represent the coordinates (symbols) of the codewords, while the hyperedges of G represent possible locations for errors, i.e., each hyperedge e represents the set of coordinates that can be corrupted in the specific scenario represented by e. For each collection of error sets represented by G, we are interested in finding the minimum alphabet size over which there exists a code that can correct all error sets specified by edges in G. In our setting, (n,k)-MDS codes can correct error patterns corresponding to the complete $(n-k)/2$-uniform n-vertex hypergraph.

In this work, we relate the minimum alphabet size of error-correcting codes with predefined error patterns to certain variants of hypergraph coloring. Through reductive arguments to erasure codes, and in particular to our prior work [8] in the context of erasure codes with generalized decoding sets, we propose code design for the error setting at hand, and show that non-linear error-correcting codes outperform linear ones. We then turn to study a variation of the problem which allows a small probability of decoding error and relate it to an approximate version of hypergraph coloring.

Our work is structured as follows. In Section II, we give some preliminaries and, in particular, we introduce our model for generalized erasure and error patterns. We also review our previous study on erasure codes in the generalized setting of a predefined collection of decoding sets [8]. In Section III, we present bounds on the minimum alphabet size of the corresponding codes through hypergraph coloring. In Section IV, we reduce the error-correcting setting to the erasure setting. In Section V, we extend our studies to the problem of error detection. Finally, in Section VI, we relax the zero-error requirement for decoding a correct message and analyze settings which allow small $\varepsilon > 0$ probability of decoding error.

II. Model and Preliminaries

Since our paper makes a connection between erasure and error correction in a generalized setting, we present definitions for both scenarios. We begin by presenting a definition and our prior results for erasure correction scenarios.

A. *Erasure Correction with predefined decoding sets.*

We start by studying the design of erasure-codes in a generalized setting in which decoding is required from a collection of predefined decoding sets. In this setting, the decoding sets include the set of coordinates that can be used to decode the message. The setting is represented by a hypergraph $G = ([n], E)$, with the set $[n] = \{1, \ldots, n\}$ of nodes representing coordinates and set of hyperedges E representing decoding sets.

We define the q_k parameter of a given hypergraph $G = ([n], E)$ as the minimum alphabet size of a (n,k) erasure code that enables the receiver to decode the original message from every subset $e \in E$.

Mira Gonen is with the Department of Computer Science, Ariel University, Ariel 40700, Israel (e-mail: mirag@ariel.ac.il).

Michael Langberg is with the Department of Electrical Engineering, University at Buffalo (State University of New-York), Buffalo, NY 14260, USA (e-mail: mikel@buffalo.edu). Work supported in part by NSF grant 1909451.

Alex Sprintson is with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128, USA (e-mail: spalex@tamu.edu).

[1]The minimum alphabet size of an (n,k) MDS code is unknown, see Conjectures 1 and 2 in the paper.

Definition 1 (The q_k parameter [8]) *Let $G = ([n], E)$ be a hypergraph on the vertex set $[n] = \{1, \ldots, n\}$. Let k be integer. Let $q_k(G)$ denote the smallest size q of an alphabet F for which there exist an encoding function*

$$C : F^k \to F^n$$

and a decoding function

$$D : (F \cup \{\perp\})^n \to F^k$$

such that for every edge $e \in E$ and every message $m \in F^k$ it holds that

$$D(C_e(m)) = m.$$

Here, $C_e(m)$ stands for the word obtained from the codeword $C(m)$ by replacing the symbols in the locations of $[n] \setminus e$ by the erasure symbol $\perp$.

Similarly, let $q_{k,lin}(G)$ denote the smallest prime power q for which there exist linear *encoding and decoding functions defined above when F is a field of size q.*

In Definition 1, notice that for G that includes edges of size less than k no such (C, D) pair exists (no matter what the size of F is). In this case we define $q_k(G)$ and $q_{k,lin}(G)$ to be ∞. Moreover, for every G with edges of size at least k, MDS codes satisfy the requirements on (C, D) and thus $q_k(G) < \infty$. Specifically, observe that for the complete n-vertex k-uniform hypergraph, denoted by $\kappa_{n,k}$, the values of $q_k(\kappa_{n,k})$ and $q_{k,lin}(\kappa_{n,k})$ are equal to the minimum alphabet sizes of general and linear (n, k) MDS codes, respectively. We state below the MDS conjectures for general and for linear codes (see, e.g., [7], [9]–[11]).

Conjecture 1 (MDS Conjecture for general codes) *For given integers $k < q \neq 6$, let $n(q, k)$ be the largest integer n such that $q_k(\kappa_{n,k}) \leq q$. Then,*

$$n(q,k) \leq \begin{cases} q+2 & \text{if } 4|q \text{ and } k \in \{3, q-1\} \\ q+1 & \text{otherwise.} \end{cases} \tag{1}$$

Conjecture 2 (MDS Conjecture for linear codes) *For given integers $k < q$ where q is a prime power, let $n(q, k)$ be the largest integer n such that $q_{k,lin}(\kappa_{n,k}) \leq q$. Then,*

$$n(q,k) \leq \begin{cases} q+2 & \text{if } q \text{ is even and } k \in \{3, q-1\} \\ q+1 & \text{otherwise.} \end{cases} \tag{2}$$

There are strong relations between the q parameter of hypergraphs and certain colorings.

Definition 2 (Hypergraph strong-coloring) *A* valid *strong-coloring of a hypergraph G is an assignment of colors to its vertices so that the vertices of each edge are assigned to distinct colors. The chromatic number $\chi(G)$ of G is the minimum number of colors that allows a valid strong-coloring of G. At times, we refer to χ simply as the chromatic number of G.*

Definition 3 (Hypergraph k-coloring) *A* valid *k-coloring of a hypergraph G is an assignment of colors to its vertices so that the vertices of each edge are assigned to at least k distinct colors. The k-chromatic number $\chi_k(G)$ of G is the minimum number of colors that allows a valid k-coloring of G. If G has edges of size less than k, we define $\chi_k(G) = \infty$.*

Note that a k-coloring of a k-uniform hypergraph is exactly a strong-coloring. Also, note that every hypergraph G for which $q_k(G) < \infty$ (i.e., all edges are of size at least k) satisfies $\chi_k(G) \leq \chi(G)$. In particular, for k-uniform hypergraphs G, $\chi_k(G) = \chi(G)$.

Theorem 1 (Connecting $q_k(G)$ with $\chi_k(G)$, [8]) *For every hypergraph G for which $q_k(G) < \infty$,*

$$q_k(G) \leq q_k(\kappa_{\chi_k(G),k}) \quad \text{and} \quad q_{k,lin}(G) \leq q_{k,lin}(\kappa_{\chi_k(G),k}).$$

In particular,

$$q_k(G) \leq q_{k,lin}(G) \leq [\chi_k(G) - 1]_{pp}.$$

Here, for an integer x, $[x]_{pp}$ represents the smallest prime power that is greater or equal to x.

Theorem 1 formalizes the natural intuition that for *simple* collections of erasure patterns G, i.e., the setting in which $\chi_k(G)$ is small, a *small* alphabet size q suffices for a suitable erasure code. In particular, the theorem states that $q_k(G)$ is upper bounded by $q_k(\kappa_{\chi_k(G),k})$, which is the minimum alphabet size of a $(\chi_k(G), k)$ MDS code.

The graph family $G_{q,k}$, defined next, is helpful in analyzing the tightness of the upper bound provided by Theorem 1.

Definition 4 (The graph family $G_{q,k}$) *For integers q and k, let $G_{q,k}$ be the k-uniform hypergraph whose vertex set consists of all the balanced vectors of length q^k over $F = \{0, 1, \ldots, q-1\}$, that is, the vectors $u \in F^{q^k}$ such that $|\{i \in [q^k] \mid u_i = j\}| = q^{k-1}$ for every $j \in F$, where k vertices $u^1 = (u^1_1, \ldots, u^1_{q^k}), \ldots, u^k = (u^k_1, \ldots, u^k_{q^k})$ form an edge if the collection of k-tuples $\{(u^1_i, u^2_i, \ldots, u^k_i)\}_{i \in [q^k]}$ is equal to $[q]^k$.*

The following lemma identified hypergraphs G for which the gap between $q_k(G)$ and $\chi_k(G)$ is maximal.

Lemma 1 (The extremal nature of $G_{q,k}$, [8]) *For integers q and k,*

1) *$q_k(G_{q,k}) \leq q$, and*
2) *$\chi_k(G) \leq \chi_k(G_{q,k})$ for every graph G with $q_k(G) = q$.*

Extending results in [8], below we present (rater loose) bounds on $\chi(G_{q,k})$. Proof appears in the full version of this work [12].

Proposition 1 (Bounds of $\chi_k(G_{k,q})$) *For every prime power q and $k \geq 2$,*

$$\frac{q^k - 1}{q - 1} \leq \chi_k(G_{q,k}) \leq \binom{q^{k-1} + 1}{q^{k-2} + 1}.$$

Lemma 1 and Proposition 1 imply a gap between $q_k(G_{q,k})$ and $\chi_k(G_{q,k})$ which can be extended to one between $q_{k,lin}$ and the k-chromatic number of the subgraph of $G_{q,k}$ induced by vertices that correspond to normalized linear functions.

Proposition 2 (Gap between $q_{k,lin}(G)$ and $\chi_k(G)$, [8]) *For every $k \geq 3$ and every prime power q, there exists a k-uniform hypergraph G with $q_{k,lin}(G) \leq q$ and yet $\chi_k(G) \geq \frac{q^k - 1}{q - 1}$.*

We finally state a modest known gap between $q_{k,lin}$ and q_k. Identifying graphs that exhibit a larger gap than that presented below is a problem left open in this work.

Proposition 3 (Gap between $q_{k,lin}$ and q_k, [13]) *For $q=3$ and $k=2$ it holds that*

$$q_{k,lin}(G_{q,k}) = [\chi_k(G_{q,k})-1]_{pp} = 5 > 3 \geq q_k(G_{q,k}).$$

B. Error Correction with predefined error sets.

In what follows, we extend our discussion beyond erasures to the context of *errors*. As we will see, several of our results on the q-parameter corresponding to erasures extend naturally to the p-parameter (defined below) corresponding to codes with restricted error sets. Similarly, to the erasure setting, we represent the collection of error sets by using a hypergraph $G=([n],E)$, in which the set of vertices $[n]$ represents coordinates of a codeword. Each edge $e \in E$ of G represents an error set, i.e., the set of the coordinates that can be altered. Note that this is different from the notation used in Definition 1 for the erasure case in which edges e represented decoding sets (i.e., sets of uncorrupted symbols).

Definition 5 (The p_k parameter) *Let $G=([n],E)$ be a hypergraph on the vertex set $[n]=\{1,\dots,n\}$. Let k be an integer. Let $p_k(G)$ denote the smallest size p of an alphabet F for which there exist an encoding function*

$$C: F^k \to F^n$$

and a decoding function

$$D: F^n \to F^k$$

such that for every edge $e \in E$, every message $m \in F^k$, and every error vector $v=(v_1,\dots,v_n) \in F^n$,

$$D(C(m) \diamond_e v) = m.$$

Here, for $C(m)=c_1,\dots,c_n$, the term $C(m)\diamond_e v$ refers to the vector $y=y_1,\dots,y_n$ for which for $i\in[n]$, $y_i=v_i$ if $i\in e$, and otherwise $y_i=c_i$ (i.e., we overwrite $C(m)$ with values of v in the coordinates $i\in e$). If such codes (C,D) do not exist we define $p_k(G)=\infty$.

Similarly, let $p_{k,lin}(G)$ denote the smallest prime power p for which there exist linear *encoding and decoding functions as above when F is a field of size p. If linear codes (C,D) do not exist we define $p_{k,lin}(G)=\infty$.*

In Definition 5, the pair (C,D) corresponds to a code that is resilient to errors on locations corresponding to an edge $e\in E$. That is, the edge set E represents the possible error patterns (i.e., sets of potentially corrupted symbols).

Similar to the case of erasures, when $n-k$ is even, for the complete hypergraph $\kappa_{n,\frac{n-k}{2}}$, the values of $p_k(\kappa_{n,\frac{n-k}{2}})$ and $p_{k,lin}(\kappa_{n,\frac{n-k}{2}})$ are equal to the minimum alphabet sizes of general and linear (n,k) MDS codes, respectively. That is, $p_k(\kappa_{n,\frac{n-k}{2}})=q_k(\kappa_{n,k})$ and $p_{k,lin}(\kappa_{n,\frac{n-k}{2}})=q_{k,lin}(\kappa_{n,k})$.

Note that Definitions 1 and 5 assume zero-error decoding. We relax this requirement in Section VI.

III. Bounds on the Alphabet Size

Proposition 4 (Analog of Theorem 1) *Let k be an integer and let $G=([n],E)$ be a hypergraph for which it holds that $|e|\leq\left\lfloor\frac{n-k}{2}\right\rfloor$ for all $e\in E$. Then, it holds that*

$$p_k(G)\leq q_k(\kappa_{\chi,k}),$$

where $\chi=\chi(\bar{G})$ and $\bar{G}=(V,\bar{E})$ is the hypergraph with vertex set $V=[n]$ and edges $\bar{E}=\{V\setminus e | e\in E\}$.

Proof: To ease our notation, we assume that $n-k$ is even (minor modifications in the proof are needed otherwise). Let G be as above and let $\chi=\chi(\bar{G})$. Denoting $q=q_k(\kappa_{\chi,k})$, it follows that there exists a (χ,k) MDS code C over an alphabet F of size q. To prove that $p_k(G)\leq q$, we define a coding scheme for G over the alphabet F that includes the following two steps. First, fix a valid strong coloring $g:[n]\to[\chi]$ of $\bar{G}$. Second, consider the encoding function $\widetilde{C}:F^k\to F^n$ that given a message $m\in F^k$ outputs the vector in F^n whose i'th entry $\widetilde{C}_i(m)$ is $C_{g(i)}(m)$, i.e., $\widetilde{C}_i(m)$ is the coordinate in the codeword $C(m)$ which corresponds to the color of the i'th vertex. Here, and throughout, we use the notation $C_i(m)$ to denote the i'th entry in the codeword $C(m)$.

The decoder $\widetilde{D}:F^n\to F^k$ for G is now defined using the following procedure. Consider an error vector $v\in F^n$, edge $e_0\in E$, and the corresponding received word $y=\widetilde{C}(m)\diamond_{e_0} v$. For each edge $\bar{e}$ in $\bar{E}$, the decoder $\widetilde{D}$ considers $y_{\bar{e}}$ consisting of the entries of y restricted to the indices in $\bar{e}$, and detects whether $y_{\bar{e}}$ has been corrupted, i.e., whether $\widetilde{C}_{\bar{e}}(m)=y_{\bar{e}}$. As for at least one such edge $\bar{e}_0$ it holds that $\widetilde{C}_{\bar{e}_0}(m)=y_{\bar{e}_0}$ (e.g. for $\bar{e}_0=[n]\setminus e_0$), the decoder $\widetilde{D}$ can use $y_{\bar{e}_0}$ to decode m. We are left to show, given $\bar{e}\in\bar{E}$, how $\widetilde{D}$ can detect whether $\widetilde{C}_{\bar{e}}(m)=y_{\bar{e}}$, and if so decode m.

To detect whether a given $\bar{e}$ in $\bar{E}$ satisfies $\widetilde{C}_{\bar{e}}(m)=y_{\bar{e}}$ we note, by the definition of $\widetilde{C}$ and the fact that all vertices in $\bar{e}$ have distinct colors under the coloring g, that the entries in $\widetilde{C}_{\bar{e}}(m)$ correspond to at least $(n+k)/2$ distinct entries in $C(m)$. The latter, in turn, implies that $\widetilde{C}_{\bar{e}}$ is itself a $(|\bar{e}|,k)$ MDS code. As such, $\widetilde{C}_{\bar{e}}(m)$ can detect up to $|\bar{e}|-k\geq\frac{n-k}{2}$ errors and correct up to $(|\bar{e}|-k)/2\geq\frac{n-k}{4}$ errors. We conclude, as all error sets e are of size at most $(n-k)/2$, that given $\bar{e}$ in $\bar{E}$, the decoder $\widetilde{D}$ can detect whether or not $y_{\bar{e}}$ has been corrupted, and if not, recover m as required. ■

Proposition 4 is not tight, meaning that $p_k(G)$ might be smaller than $q_k(\kappa_{\chi,k})$. For $k=2$, take for example $G=([6],E)$ to be the 6-cycle, i.e., the graph on 6 vertices in which its edges $E=\{(i,i+1)|i=0,1,\dots,5\}$ (with addition mod 6). Then $p_2(G)=2$, since the binary encoding $C:F^2\to F^6$ in which for a message $m=(x,y)\in F^2$ equals $C(x,y)=(x,y,x,y,x,y)$ allows majority decoding for any 2 errors along an edge in G. However, $\chi=\chi(\bar{G})=6$, since every pair of vertices in $\bar{G}$ is included in some edge in $\bar{E}$, and by [14] it holds that $q_2(\kappa_{\chi,k})=q_2(\kappa_{6,2})=5$. In the next section, we improve on Proposition 4 by connecting the p_k and q_k parameters.

IV. Connecting Error and Erasure Correcting Codes

For parameters n and k, we say that encoder $C : F^k \to F^n$ is *good* for a given hypergraph $G = ([n], E)$ with respect to erasures (res., errors) if there exists a decoder D satisfying Definition 1 (res., Definition 5). The following proposition is proven from basic principles.

Proposition 5 (From errors to erasures) *Let n and k be parameters. Consider a hypergraph $G^{\mathsf{err}} = ([n], E^{\mathsf{err}})$ corresponding to errors. Let $G^{\mathsf{era}} = ([n], E^{\mathsf{era}})$ be the hypergraph (corresponding to erasures) for which*

$$E^{\mathsf{era}} = \{[n] \setminus (e_1^{\mathsf{err}} \cup e_2^{\mathsf{err}}) \mid e_1^{\mathsf{err}}, e_2^{\mathsf{err}} \in E^{\mathsf{err}}\}.$$

Let $C : F^k \to F^n$ be any encoder. Then, C is good for G^{err} if and only if C is good for G^{era}.

Proof: First assume that C is good for G^{err}. We show that for every edge $e = e^{\mathsf{era}} \in E^{\mathsf{era}}$, one can decode m from $C_e(m)$. Assume in contradiction that there are two messages $m_1 \neq m_2$ such that $C_e(m_1) = C_e(m_2)$. Recall that $e = [n] \setminus (e_1 \cup e_2)$ for $e_1 = e_1^{\mathsf{err}} \in E^{\mathsf{err}}$ and $e_2 = e_2^{\mathsf{err}} \in E^{\mathsf{err}}$. Consider the word $y = (y_1, \dots, y_n) \in F^n$ such that for $i \in e = [n] \setminus (e_1 \cup e_2)$: $y_i = C_i(m_1) = C_i(m_2)$, for $i \in e_1 \setminus e_2$: $y_i = C_i(m_2)$, and for $i \in e_2$: $y_i = C_i(m_1)$. It is not hard to verify that there exist vectors v_1 and v_2 such that $y = C(m_1) \diamond_{e_1} v_1 = C(m_2) \diamond_{e_2} v_2$. Namely, y could be obtained from the codeword $C(m_1)$ with error vector v_1 corresponding to e_1 or from the codeword $C(m_2)$ with error vector v_2 corresponding to e_2, contradicting the existence of a decoder D according to Definition 5.

For the other direction, if code C is not good for G^{err} then there exist two messages, m_1 and m_2, two error vectors v_1 and v_2, and two edges e_1 and e_2 in E^{err} such that $C(m_1) \diamond_{e_1} v_1 = C(m_2) \diamond_{e_2} v_2$. Otherwise, it is not hard to verify the existence of a natural decoder D according to Definition 5. Let $e = [n] \setminus (e_1 \cup e_2) \in E^{\mathsf{era}}$. The equality $C(m_1) \diamond_{e_1} v_1 = C(m_2) \diamond_{e_2} v_2$ now implies that $C_e(m_1) = C_e(m_2)$, which in turn implies that C is not good for G^{era}. ■

The proposition above has an operational perspective. Namely, one can design an error-correcting code C and decoder D for a given graph G^{err}, by designing an erasure-code for the graph G^{era}. The latter can be done, e.g., using Theorem 1 to obtain the following corollary (which improves on Proposition 4).

Corollary 2 *Let k be an integer. For every hypergraph $G^{\mathsf{err}} = ([n], E)$ for which $p_k(G^{\mathsf{err}}) < \infty$ it holds that*

$$p_k(G^{\mathsf{err}}) \le q_k(G^{\mathsf{era}}) \le q_k(\kappa_{\chi_k(G^{\mathsf{era}}),k}) \le [\chi_k(G^{\mathsf{era}}) - 1]_{pp},$$

which, in turn, implies that

$$p_k(G^{\mathsf{err}}) \le q_k(\kappa_{\chi,k}),$$

where $\chi = \chi(\overline{G^{\mathsf{err}}})$ as in Proposition 4.

We now extend the connections implied by Proposition 5 to capture the p_k and q_k parameters.

Theorem 3 (Connecting p_k with q_k) *Let n, k be parameters such that $n - k \ge k$. Let $G_0^{\mathsf{era}} = ([n], E_0^{\mathsf{era}})$ be a hypergraph corresponding to erasures such that $q_k(G_0^{\mathsf{era}}) < \infty$. Then, for $N = 2n - k$ there exists a hypergraph G^{err} on N vertices such that $p_k(G^{\mathsf{err}}) = q_k(G_0^{\mathsf{era}})$ and $p_{k,lin}(G^{\mathsf{err}}) = q_{k,lin}(G_0^{\mathsf{era}})$.*

Proof: Let $G_0^{\mathsf{era}} = ([n], E_0^{\mathsf{era}})$ be as above. We define two graphs according to G_0^{era}. First consider the graph $G^{\mathsf{err}} = ([n] \cup U, E^{\mathsf{err}})$ corresponding to errors for which U is a vertex set of size $n - k$ and

$$E^{\mathsf{err}} = \{U\} \cup \{[n] \setminus e^{\mathsf{era}} | e^{\mathsf{era}} \in E_0^{\mathsf{era}}\}.$$

Here, we use the fact that edges in E_0^{era} are subsets of $[n]$. Namely, the vertex set $[n] \cup U$ of G^{err} is of size $N = 2n - k$ and each edge in E^{err} is of size at most $\frac{N-k}{2} = n - k$. We refer to the edges in $\{[n] \setminus e^{\mathsf{era}} | e^{\mathsf{era}} \in E_0^{\mathsf{era}}\} \subset E^{\mathsf{err}}$ as *ordinary* edges, and to the edge $U \in E^{\mathsf{err}}$ as the *special* edge.

Let G^{era} be the graph corresponding to erasures defined by G^{err} as in Proposition 5. Namely, $G^{\mathsf{era}} = ([n] \cup U, E^{\mathsf{era}})$ where

$$E^{\mathsf{era}} = \{([n] \cup U) \setminus (e_1^{\mathsf{err}} \cup e_2^{\mathsf{err}}) \mid e_1^{\mathsf{err}}, e_2^{\mathsf{err}} \in E^{\mathsf{err}}\}.$$

Taking a closer look at the edge set E^{era}, if an edge e^{era} in E^{era} is defined by two ordinary edges of E^{err}, then it is not hard to verify that $U \subseteq e^{\mathsf{era}}$. If an edge e^{era} in E^{era} is defined by the special edge U and an ordinary edge $e \in E^{\mathsf{err}}$, then $e^{\mathsf{era}} = [n] \setminus e$. As the ordinary edge $e \in E^{\mathsf{err}}$, by definition, equals $[n] \setminus e_0^{\mathsf{era}}$ for an edge $e_0^{\mathsf{era}} \in E_0^{\mathsf{era}}$ we conclude that $e^{\mathsf{era}} = e_0^{\mathsf{era}}$. Finally, if an edge e in E^{era} is defined solely by U (i.e., we set $e_1 = e_2 = U$), then $e = [n]$. All in all, we conclude that the edge set E^{era} equals the edges $E_0^{\mathsf{era}} \cup \{[n]\}$ and an additional set of edges e^{era} for which $U \subseteq e^{\mathsf{era}}$.

We now show that $q_k(G^{\mathsf{era}}) = q_k(G_0^{\mathsf{era}})$. We start by studying codes for G_0^{era} and G^{era}. For any code $C_0 : F^k \to F^n$ for G_0^{era}, define the code $C : F^k \to F^{n+(n-k)}$ for G^{era} in which for every message m it holds that $C(m) = C_0(m)$ on the first $[n]$ entries, that $C(m) = m$ on entries $n+1, \dots, n+k$ and that $C(m)$ equals the symbol $a \in F$ for the remaining entries $n+k+1, \dots, 2n-k$. Here, we use the fact that $n - k \ge k$. Similarly, for any code $C : F^k \to F^{n+n-k}$ for G^{era}, let the code $C_0 : F^k \to F^n$ for G_0^{era} be the restriction of C to the first n entries. It is now not hard to verify that C_0 is good for G_0^{era} if and only if C is good for G^{era}. More specifically, let C_0 be a code that is good for G_0^{era}, and let D_0 be the corresponding decoder. For any message m and edge $e = e_0^{\mathsf{era}}$ it holds that $D_0((C_0)_e(m)) = m$. To show that C is good for G^{era} we define the decoder D, that for $e^{\mathsf{era}} \in E^{\mathsf{era}}$ either runs D_0 on the first n entries of C if $e^{\mathsf{era}} \subseteq [n]$, or decodes using the identity mapping from U if $U \subseteq e^{\mathsf{era}}$. For the opposite direction, let C be good for G^{era}, and let D be the corresponding decoder. To show that C_0 is good for G_0^{era} we define the decoder D_0 as the restriction of D that takes into account only the first n entries of C. Correctness follows as $E_0^{\mathsf{era}} \subseteq E^{\mathsf{era}}$ and as C_0 is a restriction of C to the first n entries.

To show that $p_k(G^{\mathsf{err}}) = q_k(G_0^{\mathsf{era}})$, let $N = 2n - k$ and let $C : F^k \to F^N$ be any encoder. By Proposition 5, C is good for G^{err} if and only if C is good for G^{era}. By the discussion above, C is good for G^{era} if and only if the corresponding C_0

is good for $G_0^{\texttt{era}}$. Thus, C_0 is good for $G_0^{\texttt{era}}$ if and only if C is good for $G^{\texttt{err}}$. Optimizing over $|F|$, we conclude $p_k(G^{\texttt{err}}) = q_k(G_0^{\texttt{era}})$. As the reductions described above between C and C_0 preserves linearity, we also conclude that $p_{k,lin}(G^{\texttt{err}}) = q_{k,lin}(G_0^{\texttt{era}})$. ■

By Theorem 3 the gap between the q_k parameter and the $q_{k,lin}$ parameter for erasure codes stated in Proposition 3 implies a gap between the p_k parameter and the $p_{k,lin}$ parameter for error correcting codes. We summarize this results in the following corollary.

Corollary 4 (Non-linear codes outperform linear codes) *For $k = 2$, there exists a hypergraph G with $p_{k,lin}(G) = 5$ and yet $p_k(G) = 3$.*

V. Error detection

Similar to the case of errors and erasures, one can define analogs of Definitions 1 and 5 for the case of error detection. Namely, for a given hypergraph $G = ([n], E)$ the r_k parameter defined below equals the minimum size alphabet of an (n, k) error detection code that can detect error patters represented by E.

Definition 6 (The r_k parameter) *Let $G = ([n], E)$ be a hypergraph on the vertex set $[n] = \{1, \ldots, n\}$, and let k be an integer. Let $r_k(G)$ denote the smallest size r of an alphabet F for which there exist an encoding function*

$$C : F^k \to F^n$$

and a decoding function

$$D : F^n \to \{\text{error, no-error}\}$$

such that for every edge $e \in E$, every message $m \in F^k$, and every error vector $v = (v_1, \ldots, v_n) \in F^n$,

$$D(C(m) \diamond_e v) = \text{"error"} \text{ if and only if } C(m) \neq C(m) \diamond_e v,$$

($\diamond_e$ is defined in Definition 5).

Similar to Definition 1, in Definition 6, $r_k(G)$ is defined if and only if all edges of G are of size at most $n - k$, otherwise we define $r_k(G) = \infty$. Also, similar to Proposition 5, the following proposition is proven from basic principles (its proof is sketched here for completeness).

Proposition 6 (Detecting errors vs. correcting erasures) *Let n and k be parameters such that $n - k \geq k$. For a hypergraph $G = ([n], E)$, let $\bar{G} = ([n], \bar{E})$ be the hypergraph for which $\bar{E} = \{[n] \setminus e | e \in E\}$. Then, $r_k(G) = q_k(\bar{G})$.*

Proof: Assume that $\bar{C}$ is a good erasure code for $\bar{G}$. The same code can be used for detection on G. Namely, given a received word y, to check if y is corrupted in locations corresponding to $e \in E$, decode to m using $y_{\bar{e}}$ (via the erasure decoding) and compare $\bar{C}(m)$ to y. For the other direction, assume that C is a good detection code for G. Use the same code C for erasures. To decode from $C_{\bar{e}}(m)$, construct the collection Y of size $|F|^{n-|\bar{e}|}$ of words $y \in F^n$ that equal $C_{\bar{e}}(m)$ on the locations of $\bar{e}$ and otherwise equal a (distinct) word in $F^{n-|\bar{e}|}$. As C is a detection code for errors with support $e = [n] \setminus \bar{e}$, we can detect the unique $y \in Y$ that is a codeword, and accordingly decode m. ■

VI. Average error ε

In what follows, we generalize the q_k, p_k, and r_k parameters to include a decoding error. In our prior work [8], for $k = 2$ in the context of erasures, we considered decoding error when averaged over the message set F^k. We here consider a looser notion of error that is also averaged over edges in the edge set E of the hypergraph at hand. As shown below, allowing a slight error in decoding will in turn allow the construction of codes with small alphabet sizes (independent of the blocklength n).

Definition 7 (The $q_{\varepsilon,k}$, $p_{\varepsilon,k}$, and $r_{\varepsilon,k}$ parameters) *Let k be an integer. Let $G = ([n], E)$ be a hypergraph on the vertex set $[n]$ and let $\varepsilon > 0$. Let $q_{\varepsilon,k}(G)$ denote the smallest size q of an alphabet F for which there exist an encoding function $C : F^k \to F^n$ and a decoding function $D : (F \cup \{\perp\})^n \to F^k$ such that*

$$\Pr_{e,m}[D(C_e(m)) = m] \geq 1 - \varepsilon,$$

where m is uniformly chosen from F^k, and e is uniformly chosen from E. One may define $p_{\varepsilon,k}(G)$ and $r_{\varepsilon,k}(G)$ in an analogous manner.

We will need the following *approximate* version of coloring.

Definition 8 (Hypergraph $(1-\varepsilon)$-k-coloring) *A* valid *$(1-\varepsilon)$-k-coloring of a hypergraph $G = (V, E)$ is an assignment of colors to its vertices V so that for at least $(1-\varepsilon)|E|$ edges $e \in E$, the vertices of e are assigned to at least k colors. The $(1-\varepsilon)$-k-chromatic number $\chi_{\varepsilon,k}(G)$ of G is the minimum number of colors that allows a valid $(1-\varepsilon)$-k-coloring of G.*

Theorem 5 *Let $G = (V, E)$ be a hypergraph, and $\varepsilon > 0$ a parameter. Then $q_{\varepsilon,k}(G) \leq [\chi_{\varepsilon,k}(G) - 1]_{pp}$. In particular, for any two integers, n and k, it holds that $q_{\varepsilon,k}(\kappa_{n,k}) \leq O(k^2/\varepsilon)$.*

Proof: The proof that $q_{\varepsilon,k}(G) \leq [\chi_{\varepsilon,k}(G) - 1]_{pp}$ is almost identical to the proof of Theorem 1 (presented in [8]) and is obtained by replacing q_k and χ_k by $q_{\varepsilon,k}$ and $\chi_{\varepsilon,k}$ respectively. The second part of the theorem follows by showing that $\kappa_{n,k}$ can be $(1-\varepsilon)$-k colored using k^2/ε colors. Consider partitioning $[n]$ into k^2/ε subsets, each of size $\varepsilon n/k^2$. Assign the same color to all the vertices in the same subset, and distinct colors to vertices in distinct subsets. We now show that this is a $(1-\varepsilon)$-k coloring. The fraction of edges that are assigned to at least k colors is

$$\frac{\binom{k^2/\varepsilon}{k} \cdot (\varepsilon n/k^2)^k}{\binom{n}{k}}.$$

Now, for integers a and b, $\binom{a}{b} = \frac{\prod_{j=0}^{b-1}(a-j)}{b!}$, and $a^b \geq \prod_{j=0}^{b-1}(a-j) \geq (a-b)^b \geq (1 - b^2/a)a^b$, thus

$$\frac{(1-b^2/a)a^b}{b!} \leq \binom{a}{b} \leq \frac{a^b}{b!}.$$

Therefore $\frac{\binom{k^2/\varepsilon}{k} \cdot (\varepsilon n/k^2)^k}{\binom{n}{k}} \geq 1 - \varepsilon$. ■

Notice that implications corresponding to those in Theorem 5 on parameters $p_{\varepsilon,k}$ and $r_{\varepsilon,k}$ can be derived using Theorem 3 and Proposition 6, respectively.

References

[1] Y. Wu. Novel burst error correction algorithms for reed-solomon codes. *IEEE Transactions on Information Theory*, 58(2):519–529, 2012.

[2] Eitan Yaakobi, Jing Ma, Laura Grupp, Paul H Siegel, Steven Swanson, and Jack K Wolf. Error characterization and coding schemes for flash memories. In *2010 IEEE Globecom Workshops*, pages 1856–1860, 2010.

[3] G. Forney. Burst-correcting codes for the classic bursty channel. *IEEE Transactions on Communication Technology*, 19(5):772–781, 1971.

[4] J. Moon and J. Park. Detection of prescribed error events: application to perpendicular recording. In *IEEE International Conference on Communications, ICC 2005, Seoul*, volume 3, pages 2057–2062. IEEE, 200520.

[5] L. Yohananov and E. Yaakobi. Codes for graph erasures. *IEEE Transactions on Information Theory*, 65(9):5433–453, 2019.

[6] Vilas Sridharan and Dean Liberty. A study of dram failures in the field. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, SC '12, Washington, DC, USA, 2012. IEEE Computer Society Press.

[7] K. A. Bush. Orthogonal arrays of index unity. *Ann. Math. Statistics*, 23:426 – 434, 1952.

[8] M. Gonen, I. Haviv, M. Langberg, and A. Sprintson. Minimizing the alphabet size of erasure codes with restricted decoding sets. In *IEEE International Symposium on Information Theory, ISIT 2020, Los Angeles, CA, USA, June 21-26, 2020*, pages 144–149. IEEE, 2020.

[9] B. Segre. Curve razionali normali e k-archi negli spazi finiti. *Ann. Mat. Pura Appl.*, 39(4):357–379, 1955.

[10] F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-CorrectingCodes*. Amsterdam: North-Holland, 1977.

[11] S. Huntemann. The upper bound of general maximum distance separable codes, 2012. Honours Project.

[12] M. Gonen, M. Langberg, and A. Sprintson. Minimizing the alphabet size in codes with restricted error sets. *Manuscript. Available on arXiv.org*, 2021.

[13] A. R. Lehman and E. Lehman. Network coding: does the model need tuning? In *SODA'05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 499–504, 2005.

[14] S. Huntemann. The Upper Bound of General Maximum Distance Separable Codes, 2012.