On Fixed Length Systematic All Limited Magnitude Zero Deletion/Insertion Error Control Codes*

Luca G. Tallini†, Hoang Vu‡ and Bella Bose‡
†Dipartimento di Scienze della Comunicazione, Università degli Studi di Teramo, Teramo, Italy. E-mail: ltallini@unite.it
‡School of EECS, Oregon State University, Corvallis, OR, USA. E-mails: {vuh,Bella.Bose}@oregonstate.edu

***Abstract*—In a systematic code (systematic in the strict sense) a check symbol is appended to the data word. Here, the theory and design of systematic binary block codes capable of correcting t insertion and/or deletion of the symbol 0 in each and every 0-run is studied. This problem is related to the zero error capacity achieving systematic codes in limited magnitude error channels. Optimal and sub-optimal systematic code designs and the encoding/decoding algorithms are given.**

***Index Terms*—m-ary codes, 0-insertion/deletion errors, max L_1 distance, error control codes, limited magnitude error channels, communication systems, synchronization errors, sticky channels, constant weight codes.**

I. Introduction

In various communication and magnetic recording systems, the channel may cause two types of synchronization errors. The first one is not receiving a transmitted symbol (a deletion error), and the second one is receiving a spurious symbol (an insertion error). Furthermore, the propagation of these errors will significantly reduce the performance of the systems. The Insertion/Deletion Channel, along with channels like replication, substitutions, and combinations thereof, finds applications in diverse fields, from computational biology to document exchange and DNA data storage systems.

The general problem of designing efficient codes for insertion/deletion of any symbol has been an open research problem for over 55 years even though various results have been given in [1], [7], [9]–[12], [15], [18], [19], [23], [25]–[29] (and the references in these papers). Let $\mathbb{Z}_2 \stackrel{\text{def}}{=} \{0,1\}$ be the binary alphabet. Some efficient binary code designs for controlling the insertions/deletions of some fixed symbol, say 0, are given in [5], [13], [14], [16], [18], [22], [30]–[33], [35]. This simpler 0-error model finds application in achieving reliable communication for the repetition (or, sticky) channel model where the insertion and/or deletion of repeated symbols may occur [5], [22], [35]. In [32] the 0-error model with the limited magnitude error [3] constraints is studied from the zero error capacity perspective. In this setting, it comes natural to assume that the number of insertions of the fixed symbol 0 (i. e., 0-insertions) and the number of the symbol 0 deletions (i. e., 0-deletions) in each and every 0-run (also referred to as bucket of zeros in [32]) is at most t_i and t_d, respectively. In [32], some theory and efficient optimal and suboptimal non-systematic binary block code designs are presented and these codes are capable of correcting up to t_i 0-insertions and t_d 0-deletions in each and every 0-run ((t_i,t_d)-LM0EC codes), for fixed $t_i, t_d \in \mathbb{N}$. For example, let $t_i = 1$ and $t_d = 2$. If the weight 7 codeword $X = 001\,00001\,1\,1\,01\,1\,1\,00 \in \mathbb{Z}_2^{16}$ is sent and $Y = \epsilon 01\,0\epsilon\epsilon 01\,1\,\mathbf{0}1\,\epsilon 1\,1\,1\,\mathbf{0}00 = 01001101111000 \in \mathbb{Z}_2^{14}$ is received (ϵ being the empty string), then the number of 0-insertion and 0-deletions in each of the $7+1 = 8$ 0-runs is less than $t_i = 1$ and $t_d = 2$. In this case, the receiver, from Y, can correct all the $1+2+1+1+1 = 6$ 0-errors in Y.

Here, we focus our attention on the theory and design of systematic (t_i,t_d)-LM0EC binary block codes. Such theory is based on the "max L_1 distance" defined below and it should not be confused with the theory on (t_i,t_d)-0EC codes, which are based on the "L_1 distance" [31].

***Definition 1** ((strict sense) systematic binary block code):* A binary block code $\mathcal{C}$ is systematic with $k \in \mathbb{N}$ information bits and $r \in \mathbb{N}$ check bits if, and only if, there exists a function $\mathcal{E} : \mathbb{Z}_2^k \to \mathcal{C} \subseteq \mathbb{Z}_2^{k+r}$ such that $\mathcal{E}(X) \stackrel{\text{def}}{=} X\ C(X)$, for all $X \in \mathbb{Z}_2^k$. The word $C(X) \subseteq \mathbb{Z}_2^r$ is called the check symbol associated with the information word X.

This paper gives some efficient systematic (t_i,t_d)-LM0EC code designs (according to Definition 1) which are redundancy optimal or close to optimal, as shown, for example, in Table I. The paper is organized as follows. Section II gives some notation and preliminary theory on 0-errors control codes. Section III contains a good lower bound on the minimal check symbol length required by any systematic (t_i,t_d)-LM0EC code. Section IV presents the proposed efficient systematic code designs. In Section V, some conclusions are drawn.

For clarity of theory development, as usual [18], [19], [22], [30]–[32], we assume no synchronization errors due to erroneous receptions of sequences of codewords (i. e., we assume that the receiver knows the length of the received word). However, by encoding the Hamming weight of the information word into the check word, our proposed codes can be made self synchronizing under limited magnitude 0-errors. This is because our codes decode the received check word first, are instantaneous in correcting the received 0-runs (see Section IV) and 1-errors are forbidden in the 0-error model. Furthermore, from the theory in [32] (see Theorem 1), any $(0, D-1)$-LM0EC code, $D \in \mathbb{N}$, is a (t_i,t_d)-LM0EC code, for all $t_i, t_d \stackrel{\text{def}}{=} D - t_i - 1 \in \mathbb{N}$, so (t_i,t_d)-LM0EC codes are also called D-LM0EC codes.

II. Notation and preliminaries

For $q \in \mathbb{N} \cup \{\infty\}$, let $\mathbb{Z}_q \stackrel{\text{def}}{=} \{0,1,\ldots,q-1\}$. Given $q,n \in \mathbb{N}$, the L_1 weight of a word $X = x_1x_2\ldots x_n \in \mathbb{Z}_q^n$ is the real sum $w(X) \stackrel{\text{def}}{=} w_{L_1}(X) \stackrel{\text{def}}{=} \sum_{i=1}^n x_i$. In this theory, the constant L_1 weight codes are important. So, for $n,w \in \mathbb{N}$ and a numeric set $\mathcal{B} \subseteq \mathbb{N}$, let $\mathcal{S}(\mathcal{B},n,w) \stackrel{\text{def}}{=} \{X \in \mathcal{B}^n : w_{L_1}(X) = w\}$ be the set of all n digit long words over $\mathcal{B}$ with constant L_1 weight w. Note that $\mathcal{S}(\mathcal{B},n,w) = \bigcup_{x\in\mathcal{B}} \mathcal{S}(\mathcal{B},n-1,w-x)x$, where the union is a disjoint union of sets. If $|S|$ indicates the cardinality of a set S then the general recurring formula, $|\mathcal{S}(\mathcal{B},n,w)| = \sum_{x\in\mathcal{B}} |\mathcal{S}(\mathcal{B},n-1,w-x)|$ holds for the cardinality of $\mathcal{S}(\mathcal{B},n,w)$. If $\mathcal{B} \stackrel{\text{def}}{=} \mathbb{Z}_q$ these "$\mathcal{B}$-nomial coefficients" become the "q-nomial coefficients", $|\mathcal{S}(\mathbb{Z}_q,n,w)| \stackrel{\text{def}}{=} \binom{n}{w}_q =$

*This work is supported by the National Science Foundation Grant CCF-2006571.

2024 IEEE International Symposium on Information Theory (ISIT) | 979-8-3503-8284-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISIT57864.2024.10619517

TABLE I

NUMBER $r \in \mathbb{N}$ OF CHECK BITS OF SYSTEMATIC (t_i, t_d)-LM0EC CODES FOR SOME VALUES OF INFORMATION BITS $k \in \mathbb{N}$ AND $D \stackrel{\text{def}}{=} (t_i + t_d) + 1 \in \mathbb{N}$. In the Table, for each value of $D = 2, 3, \dots, 128$, the LB labelled column gives a lower bound on r for **any** systematic D-LM0EC code; the CD column gives the value of r for the proposed D-LM0EC codes. The last column gives the value of r of the distinct weight codes in Subsubsection II-A1.

$\backslash D$	2		3		4		5		6		7		8		9		10		16		32		64		128		∞
$k\backslash$	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	Optimal CD
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	16	16	32	32	64	64	128	128	1
3	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10	11	11	17	17	33	33	65	65	129	129	4
4	3	4	6	6	7	8	9	9	10	10	11	11	12	12	13	13	14	14	20	20	36	36	68	68	132	132	11
5	4	5	7	8	9	9	10	11	12	13	14	14	15	16	17	17	18	18	24	24	40	40	72	72	136	136	26
6	5	6	8	9	11	12	13	13	14	15	16	17	18	19	20	21	22	22	32	33	50	50	82	82	146	146	57
7	6	7	10	11	12	13	15	16	17	18	19	19	20	21	22	23	24	25	35	36	64	64	97	97	161	161	120
8	7	8	11	12	14	15	17	18	19	20	22	23	24	25	26	27	28	29	39	40	70	71	129	129	196	196	247
9	8	9	13	14	16	17	19	20	22	23	24	25	27	28	29	30	32	33	44	45	75	76	136	137	252	252	502
10	9	10	14	15	18	19	22	23	25	26	27	28	30	31	33	34	35	36	50	51	82	83	144	145	269	270	1013
11	9	11	15	17	20	21	24	25	27	28	30	31	33	34	36	37	39	40	55	56	90	91	153	154	279	280	2036
12	10	12	17	18	22	23	26	27	30	31	33	34	36	37	39	40	42	43	59	60	101	102	166	167	293	294	4083
13	11	13	19	20	24	25	28	29	32	33	36	37	39	40	43	44	46	47	64	65	108	109	182	183	309	310	8178
14	12	14	20	21	26	27	31	32	35	36	39	40	43	44	46	47	50	51	70	71	116	117	203	204	334	335	16369
15	13	15	22	23	28	29	33	34	37	38	42	43	46	47	49	50	53	54	74	75	123	124	214	215	365	366	32752
16	14	16	23	24	30	31	35	36	40	41	45	46	49	50	53	54	57	58	80	81	133	134	225	226	403	404	65519
32	29	31	48	49	61	62	73	74	82	83	92	93	100	101	109	110	117	118	162	163	267	268	451	452	773	774	$2^{32}-33$
64	61	62	98	99	125	127	148	149	168	170	187	188	205	206	222	223	238	239	328	329	540	541	907	908	1553	1554	$2^{64}-65$
128	124	126	199	200	255	256	301	302	342	343	379	380	415	416	449	450	482	483	664	665	1089	1090	1826	1827	3123	3124	$2^{128}-129$
256	251	253	402	404	514	515	607	608	690	691	766	767	837	838	906	907	972	973	1339	1340	2192	2193	3673	3674	6276	6277	$2^{256}-257$
512	507	509	809	811	1034	1036	1222	1223	1387	1388	1540	1541	1684	1685	1821	1822	1954	1955	2690	2691	4402	4403	7373	7374	12593	12594	$2^{512}-513$
1024	1018	1020	1625	1626	2076	2077	2451	2452	2783	2785	3089	3090	3378	3379	3653	3654	3920	3921	5395	5396	8827	8828	14780	14781	25240	25241	$2^{1024}-1023$

$\sum_{x=0}^{q-1} \binom{n-1}{w-x}_q$, which, for $q = 2$, reduce to the usual binomial coefficients, $\binom{n}{w}$. If instead $\mathcal{B} \stackrel{\text{def}}{=} \mathbb{N} = \mathbb{Z}_\infty$ then $\mathcal{S}(\mathbb{N}, n, w)$ is the set of all compositions of the natural number w into n natural numbers and $|\mathcal{S}(\mathbb{N}, n, w)| \stackrel{\text{def}}{=} \binom{n}{w}_\infty = \binom{n+w-1}{w}$. In what follows, it is relevant to note that each set $\mathcal{S}(\mathbb{Z}_q, n, w)$ can be encoded in lexicographic order with the enumerative source encoding technique, say, with $O(n^2 \log q)$ bit operations by storing $O(n^2 q \log q)$ bits [4], [32]. Now, let $\mathcal{B}^*$ be the set of all finite length sequences over an alphabet $\mathcal{B}$. As in [18], [30], [32], consider the bijective map $A : \mathbb{Z}_2^* \to \mathbb{N}^*$ which associates any $X \in \mathcal{S}(\mathbb{Z}_2, m, u) \subseteq \mathbb{Z}_2^*$, $m, u \in \mathbb{N}$, with

$$A(X) \stackrel{\text{def}}{=} a_1 a_2 \dots a_u a_{u+1} \in \mathcal{S}(\mathbb{Z}_q, u+1, m-u) \subseteq \mathbb{Z}_q^*, \quad (1)$$

where, $q \stackrel{\text{def}}{=} q(m, u) \stackrel{\text{def}}{=} m - u + 1$ and, for all integers $h \in [1, u+1]$, $a_h \stackrel{\text{def}}{=} a_h(X) \in \mathbb{Z}_q \subseteq \mathbb{N}$ is the h-th 0-run length in the word X. For example, if $X = 010000111011100 \in \mathbb{Z}_2^*$ then $A(X) = 14001002 \in \mathbb{N}^*$. This mapping A defines a bijection from the set of all binary words of any finite length $m \in \mathbb{N}$ and Hamming weight u (= number of 1's of the binary words) into the set of words over $\mathbb{N}$ of length $u+1$ (= number of 0-runs of the binary words) and L_1 weight $m-u$ (= number of 0's of the binary words). For example, for $m = 6$, the mapping A acts on $\mathbb{Z}_2^6$ is as shown in Table II, where $\ell(X)$ indicates the length of any string X. Given $q \in \mathbb{N}$, let $|a-b| \stackrel{\text{def}}{=} \max\{a-b, b-a\} \in \mathbb{Z}_q \subseteq \mathbb{R}^+$, for all $a, b \in \mathbb{Z}_q$. The max L_1 distance plays an important role in the code design. This max L_1 distance, $D_\infty^{L_1} : \mathbb{Z}_q^* \times \mathbb{Z}_q^* :\to \mathbb{Z}_q \subseteq \mathbb{R}^+$, for any $A = a_1 a_2 \dots a_{\ell(A)}, B = b_1 b_2 \dots b_{\ell(B)} \in \mathbb{Z}_q^*$ is defined as

$$D_\infty^{L_1}(A, B) \stackrel{\text{def}}{=} \begin{cases} \max\limits_{i \in [1,n]} \{|a_i - b_i|\} & \text{if } \ell(A) = \ell(B), \\ \infty & \text{if } \ell(A) \neq \ell(B). \end{cases} \quad (2)$$

For example, if $A = 210$, $B = 021$ and $C = 0104 \in \mathbb{Z}_5^*$ then $D_\infty^{L_1}(A, B) = \max\{2, 1, 1\} = 2$ and $D_\infty^{L_1}(A, C) = D_\infty^{L_1}(B, C) = \infty$. Note that, for all $D, n \in \mathbb{N}$,

$$\forall A, B \in \mathbb{Z}_D^n, \quad D_\infty^{L_1}(A, B) < D. \quad (3)$$

With the one-to-one mapping A we can define a max distance metric in $\mathbb{Z}_2^*$ as

$$\forall X, Y \in \mathbb{Z}_2^*, \quad D_{0E}^{L_1}(X, Y) \stackrel{\text{def}}{=} D_\infty^{L_1}(A(X), A(Y)). \quad (4)$$

For example, if $X = 001011$, $Y = 100101$ and $Z = 10110000 \in \mathbb{Z}_2^*$ then $D_{0E}^{L_1}(X, Y) = \max\{2, 1, 1\} = 2$ and $D_{0E}^{L_1}(X, Z) = D_{0E}^{L_1}(Y, Z) = \infty$. As for the distinct L_1 distance based 0-error model in [30], [31], note that, with definition (4), the bijection A becomes also an isometry between the metric spaces $(\mathbb{N}^*, D_\infty^{L_1})$ and $(\mathbb{Z}_2^*, D_{0E}^{L_1})$. From the zero error capacity perspective [2], [3], [6], [8], [21], [24], the channel adjacency matrix [24] of the 0-error channel model with limited magnitude t is the graph adjacency matrix of $G^{(t)} \stackrel{\text{def}}{=} (\mathbb{Z}_2^*, E^{(t)})$ with edge set $E^{(t)} \stackrel{\text{def}}{=} \{(X, Y) \in \mathbb{Z}_2^* : D_{0E}^{L_1}(X, Y) \leq t\}$. Note that, since 1-errors are forbidden,

$$\forall X, Y \in \mathcal{C}, \quad w(X) \neq w(Y) \iff D_{0E}^{L_1}(X, Y) = \infty. \quad (5)$$

In [32], the following theorem is proved in terms of $D_\infty^{L_1}$.

Theorem 1 *(Comb. character. of (t_i, t_d)-LM0EC codes):* Let $m, t_i, t_d, D \stackrel{\text{def}}{=} t_i + t_d + 1 \in \mathbb{N}$ be given. A code $\mathcal{C} \subseteq \mathbb{Z}_2^m$ is (t_i, t_d)-LM0EC if, and only if,

$$\forall X, Y \in \mathcal{C}, \quad X \neq Y \implies D_{0E}^{L_1}(X, Y) \geq D. \quad (6)$$

It is also shown in [32] that the cardinality of the optimal non-systematic D-LM0EC code, $\mathcal{C}_{opt}$, of length m, satisfies

$$LB(m, D) \stackrel{\text{def}}{=} \sum_{u=0}^{m} \binom{u + \left\lfloor \frac{m-u}{D} \right\rfloor}{u} \leq |\mathcal{C}_{opt}| \leq \quad (7)$$
$$\sum_{u=0}^{m} \binom{u + 1 + \left\lfloor \frac{m-u}{D} \right\rfloor}{u+1} \stackrel{\text{def}}{=} UB(m, D);$$

TABLE II
THE MAPPING A ACTING ON $\mathbb{Z}_2^6$.

$w(X)$	X	$A(X)$	$\ell(A(X))$	$w_{L_1}(A(X))$
0	000000	6	1	6
1	000001	50	2	5
	000010	41		
	000100	32		
	001000	23		
	010000	14		
	100000	05		
2	000011	400	3	4
	000101	310		
	000110	301		
	001001	220		
	001010	211		
	001100	202		
	010001	130		
	010010	121		
	010100	112		
	011000	103		
	100001	040		
	100010	031		
	100100	022		
	101000	013		
	110000	004		
3	000111	3000	4	3
	001011	2100		
	001101	2010		
	001110	2001		
	010011	1200		
	010101	1110		
	010110	1101		
	011001	1020		
	011010	1011		
	011100	1002		
	100011	0300		
	100101	0210		
	100110	0201		
	101001	0120		
	101010	0111		
	101100	0102		
	110001	0030		
	110010	0021		
	110100	0012		
	111000	0003		
4	001111	20000	5	2
	010111	11000		
	011011	10100		
	011101	10010		
	011110	10001		
	100111	02000		
	101011	01100		
	101101	01010		
	101110	01001		
	110011	00200		
	110101	00110		
	110110	00101		
	111001	00020		
	111010	00011		
	111100	00002		
5	011111	100000	6	1
	101111	010000		
	110111	001000		
	111011	000100		
	111101	000010		
	111110	000001		
6	111111	0000000	7	0

for all $m, D \in \mathbb{N}$. In addition, it is shown in [32] that the following codes

$$\mathcal{C}_{m,D} \stackrel{\text{def}}{=} \{X \in \mathbb{Z}_2^m : a_h(X) = 0 \bmod D,\ \forall h \in [1, w(X)]\} \quad (8)$$

are D-LM0EC codes with $|\mathcal{C}_{m,D}| = LB(m, D)$ codewords. Also, if $\mathcal{C} \stackrel{\text{def}}{=} \mathcal{C}_{m,D} \subseteq \mathbb{Z}_2^m$ and $\mathcal{C}_u \stackrel{\text{def}}{=} \mathcal{C} \cap \mathcal{S}(\mathbb{Z}_2, m, u) \stackrel{\text{def}}{=} \mathcal{C}_{u,m,D}$, for all $u \in [0, m]$, then $\mathcal{C} = \bigcup_{u \in [0,m]} \mathcal{C}_u$ with $|\mathcal{C}_u| = \binom{u + \lfloor (m-u)/D \rfloor}{u}$.

***Example** 1 ($m = 6$ and $D = 3$): The code $\mathcal{C} = \mathcal{C}_{6,3}$ is given through the map A by*

$$A(\mathcal{C}_0) \stackrel{\text{def}}{=} \{6\}, n = 1, w = 6, q = 7,$$
$$A(\mathcal{C}_1) \stackrel{\text{def}}{=} \{32, 05\}, n = 2, w = 5, q = 6,$$
$$A(\mathcal{C}_2) \stackrel{\text{def}}{=} \{301, 031, 004\}, n = 3, w = 4, q = 5,$$
$$A(\mathcal{C}_3) \stackrel{\text{def}}{=} \{3000, 0300, 0030, 0003\}, n = 4, w = 3, q = 4,$$
$$A(\mathcal{C}_4) \stackrel{\text{def}}{=} \{00002\}, n = 5, w = 2, q = 3,$$
$$A(\mathcal{C}_5) \stackrel{\text{def}}{=} \{000001\}, n = 6, w = 1, q = 2,$$
$$A(\mathcal{C}_6) \stackrel{\text{def}}{=} \{0000000\}, n = 7, w = 0, q = 1;$$

TABLE III
THE CODE IN EXAMPLE 2

$\mathcal{C}_u$	$Y = X\,\boldsymbol{M_u} \in \mathcal{C}$	$A(Y)$	$\ell(A(\mathcal{C}_u))$	$w_{L_1}(A(\mathcal{C}_u))$
$\mathcal{C}_2$	000000 **11**	600	3	6
	000100 **01**	330		
	100000 **01**	060		
	000110 **00**	303		
	100010 **00**	033		
	110000 **00**	006		
$\mathcal{C}_5$	000111 **11**	300000	6	3
	100011 **11**	030000		
	110001 **11**	003000		
	111000 **11**	000300		
	111100 **01**	000030		
	111110 **00**	000003		
$\mathcal{C}_8$	111111 **11**	000000000	9	0

where $n \stackrel{\text{def}}{=} \ell(A(\mathcal{C}_u)) = u + 1$, $w \stackrel{\text{def}}{=} w_{L_1}(A(\mathcal{C}_u)) = m - u$ and $q \stackrel{\text{def}}{=} m - u + 1$. Since $D_{0E}^{L_1}(\mathcal{C}) = D = 3$, the code $\mathcal{C}$ can correct two 0-deletions (i. e., $t_i = 0$, $t_d = 2$) or two 0-insertions (i. e., $t_i = 2$, $t_d = 0$) or one 0-deletion and one 0-insertion (i. e., $t_i = 1$, $t_d = 1$). It has $|\mathcal{C}| = LB(6, 3) = \binom{2}{0} + \binom{2}{1} + \binom{3}{2} + \binom{4}{3} + \binom{4}{4} + \binom{5}{5} + \binom{6}{6} = 13$ codewords.

The codes in (8) are instantaneous (i. e., any received 0-run can be corrected as soon as it is received (please see [32, Subsection III.C]) but not self synchronizing under (t_i, t_d)-LM0EC. To recover synchronization at the receiver end, here we note that the string $M_u \stackrel{\text{def}}{=} 0^{(D-1)-\mu_u} 1^{\mu_u} \in \mathcal{S}(\mathbb{Z}_2, D-1, \mu_u)$, with $\mu_u \stackrel{\text{def}}{=} (m - 1 - u) \bmod D \in \mathbb{Z}_D$, can be appended to each constant weight u code $\mathcal{C}_{u,m,D}$. In this way, the instantaneous and self synchronizing code $\bar{\mathcal{C}}_{m,D} \stackrel{\text{def}}{=} \bigcup_{u=0}^{m} \mathcal{C}_{u,m,D} M_u$ is obtained with fixed length $r \stackrel{\text{def}}{=} m + D - 1$ which is exactly equal to

$$\begin{aligned} \bar{\mathcal{C}}_{m,D} &= \{X \in \mathbb{Z}_2^r : a_h(X) = 0 \bmod D,\ \forall h \in [1, w(X)+1]\} \\ &\subseteq \bigcup_{u \in [0,r] \cap (D \cdot \mathbb{Z} + r \bmod D)} \mathcal{S}(\mathbb{Z}_2, r, u). \end{aligned} \quad (9)$$

***Example** 2 ($m = 6$, $D = 3$ and $r = m + D - 1 = 8$): The code $\mathcal{C} = \bar{\mathcal{C}}_{6,3} = \mathcal{C}_2 \cup \mathcal{C}_5 \cup \mathcal{C}_8$ is given in Table III.*

The (t_i, t_d)-LM0EC on a received codeword sequence for the non systematic codes $\bar{\mathcal{C}}_{m,D}$ in (9) can be accomplished on the current received 0-run by first computing its length a' and then correcting a' by adding up to t_d or subtracting up to t_i so that a' becomes the nearest multiple of $D = t_i + t_d + 1$; this computation can be done instantaneously bit by bit until the length of the entire corrected bit sequence is exactly r so that the parsing point of the current received codeword is detected.

The check symbols in the systematic code design give in Section IV are codewords of the big enough instantaneous and self synchronizing codes in (9) and convey some information on the information word. The CD column in Table I gives exactly the length $r \in \mathbb{N}$ of such codes.

A. Some simple systematic code designs

Here, two interesting simple code designs are given.

1) Systematic Distinct Weight (DW) codes: In a distinct weight (DW) code no two codewords have the same Hamming weight. From (2), the minimum max L_1 distance of any distinct weight code is ∞ and so they can be used to correct any number of 0-errors in any 0-run. Indeed, as class of codes DW $\equiv$ ∞-0EC $\equiv$ ∞-LM0EC [31]. Now, a DW code of length $n \in \mathbb{N}$ contains $n + 1$ distinct codewords. Hence, an optimal DC code with $k \in \mathbb{N}$ information bits has length

TABLE IV
SYSTEMATIC $DW(k=3, n=7)$ CODE.

$d(X)$	X	Check of X	$w(\mathcal{E}(X))$
0	000	0000	0
1	001	0000	1
2	010	0001	2
3	011	0001	3
4	100	0111	4
5	101	0111	5
6	110	1111	6
7	111	1111	7

$n = 2^k - 1$. Let $d : \mathbb{Z}_2^k \to [0, 2^k - 1]$ be the one-to-one map which associates any X with the natural number $d(X)$ whose binary representation is X. An efficient optimal DW systematic encoding, $\mathcal{E} : \mathbb{Z}_2^k \to \mathbb{Z}_2^n$ for k information bits can be defined as $\mathcal{E}(X) \stackrel{\text{def}}{=} X\, 0^{n-k-[d(X)-w(X)]}\, 1^{d(X)-w(X)}$. Table IV gives an example with $k = 3$ information bits.

2) Repetition Codes: A D-repetition code is a code where each data symbol is repeated $D \in \mathbb{N}$ times. Since the minimum max L_1 distance of these codes is clearly D, they are D-LM0EC (and, even D-LM$\{0,1\}$EC) codes with $k \in \mathbb{N}$ data bits, length $n = Dk \in \mathbb{N}$ and $r = (D-1)k \in \mathbb{N}$ check bits. Even though they are not systematic in the strict sense considered in Definition 1 they are systematic in a wider sense [36, Section IV] and so it is interesting to consider them as touchstones.

III. REDUNDANCY LOWER BOUNDS FOR SYSTEMATIC D-LM0EC CODES

First, the following definition can be given in general for any class of binary systematic block error control codes.

***Definition** 2 (check-blocking set):* Given a class of systematic error control code with $k \in \mathbb{N}$ information bits let $C : \mathbb{Z}_2^k \to \mathbb{Z}_2^r$, $r \in \mathbb{N}$, be any function which associates every information word $X \in \mathbb{Z}_2^k$ with its check symbol $C(X) \in \mathbb{Z}_2^r$ so that the length $n \stackrel{\text{def}}{=} k + r \in \mathbb{N}$ code $\mathcal{C} \stackrel{\text{def}}{=} \{X\, C(X) : X \in \mathbb{Z}_2^k\}$ with r check bits is in the class. A set $\Delta \stackrel{\text{def}}{=} \Delta(k) \subseteq \mathbb{Z}_2^k$ of information words is called check-blocking if, and only if, for any check symbol assignment function C,

$$\forall X, Y \in \Delta, \quad X \neq Y \implies C(X) \neq C(Y).$$

Simply note that a lower bound on (the largest) $|\Delta|$ gives a lower bound on the number of check symbols (and hence, check bits) required by **any** systematic code design in the class. From (4) and (2), the D-LM0EC code class Δ data word set is check-blocking if, and only if, it satisfies the following condition.

$$\forall X, Y \in \Delta, \quad X \neq Y \implies D_{0E}^{L_1}(X, Y) < D. \tag{10}$$

We have the following theorems.

***Theorem** 2:* Given the class of D-LM0EC codes with k information bits, $k, D \in \mathbb{N}$, the largest check-blocking set of information words, $\Delta \stackrel{\text{def}}{=} \Delta(k, D) \subseteq \mathbb{Z}_2^k$, has a cardinality of

$$|\Delta| \geq \max_{v \geq 0} \binom{k - v + 1}{v}_D \stackrel{\text{def}}{=} NCLB(k, D). \tag{11}$$

Furthermore, if $D > \lfloor k/2 \rfloor$ then

$$|\Delta| \geq \binom{k}{\lceil k/2 \rceil}_2. \tag{12}$$

Proof: For all $v \in [0, k]$ let $u \stackrel{\text{def}}{=} k - v \in [0, k]$ and

$$\Delta_v \stackrel{\text{def}}{=} \{X \in \mathcal{S}(\mathbb{Z}_2, k, u) : A(X) \in \mathcal{S}(\mathbb{Z}_D, u+1, v)\} \subseteq \mathbb{Z}_2^k.$$

Note that if a data word $X \in \Delta_v$ then X contains v 0's and u 1's. For example, if $k = 6$ and $D = 3$ then $\Delta_2 = \mathcal{S}(\mathbb{Z}_2, 6, 4)$, but $\Delta_4 = A^{-1}(\{220, 211, 202, 121, 112, 022\}) \subsetneq \mathcal{S}(\mathbb{Z}_2, 6, 2)$ (see Table II). For all $v \in [0, k]$, Δ_v is a check-blocking set of Definition 2. In fact, if $X, Y \in \Delta_v$ then $A(X), A(Y) \in A(\Delta_v) \subseteq \mathbb{Z}_D^{u+1}$. So, from (4) and (3), $D_{0E}^{L_1}(X, Y) = D_\infty^{L_1}(A(X), A(Y)) < D$; i. e, Δ_v is check-blocking because of (10). Now, $|\Delta_v| = |\mathcal{S}(\mathbb{Z}_D, u+1, v)| = \binom{u+1}{v}_D$ because the map A is injective. So, (11) follows simply because $|\Delta| \geq |\Delta_v|$, for all $v \in [0, k]$. Finally, if $D > v$ then $\Delta_v = \mathcal{S}(\mathbb{Z}_2, k, u)$ and $|\Delta_v| = \binom{k}{v}$. So, (12) follows if $v = \lfloor k/2 \rfloor$. ■

For example, if $k = 6$ and $D = 3$ then the lower bound in (11) gives $|\Delta| \geq \max_{v \geq 0} \binom{7-v}{v}_3 = \binom{4}{3}_3 = |\mathcal{S}(\mathbb{Z}_3, 4, 3)| = 16$, as reported in Table V.

***Theorem** 3:* Let $k, D \in \mathbb{N}$ and $\Delta \stackrel{\text{def}}{=} \Delta(k, D) \subseteq \mathbb{Z}_2^k$ be given as in Theorem 2 and, for all $a \in [0, k-1]$, let $G_a \stackrel{\text{def}}{=} \{C(X10^a) : X10^a \in \Delta\} \subseteq \{C(X) : X \in \Delta\} \stackrel{\text{def}}{=} G$. Then $D_{0E}^{L_1}(G_a) \geq D$, $|G_a| \geq \max_{v \geq 0} \binom{k-v-a}{v}_D$, for all $a \in [0, k-1]$, and $|G| = \sum_{a \in [0, k-1]} |G_a|$.

Proof: For any $a \in [0, k-1]$, let $C(X10^a), C(Y10^a) \in G_a$ with $X10^a, Y10^a \in \Delta$. If $X \neq Y$ then $X10^a \neq Y10^a$ and $C_X \stackrel{\text{def}}{=} C(X10^a) \neq C(Y10^a) \stackrel{\text{def}}{=} C_Y$ because Δ is check-blocking. So, from (4), (2) and Definition 2,

$$\begin{aligned} &D_{0E}^{L_1}(X10^a\, C_X, Y10^a\, C_Y) = \\ &\quad \max\{D_\infty^{L_1}(A(X), A(Y)), D_\infty^{L_1}(C_X, C_Y)\} \geq D. \end{aligned} \tag{13}$$

But, from $X10^a, Y10^a \in \Delta$, relation (4) and (10) it follows $D_{0E}^{L_1}(X10^a, Y10^a) = D_\infty^{L_1}(A(X), A(Y)) < D$. So, from (4), $D_{0E}^{L_1}(C_X, C_Y)\} = D_\infty^{L_1}(C_X, C_Y)\} \geq D$. So, $D_{0E}^{L_1}(G_a) \geq D$. The relations on $|G_a|$ and $|G|$ follow as in Theorem 2 and because the G_a's are pairwise disjoint, respectively. ■

Let $\bar{\Gamma} \stackrel{\text{def}}{=} \{C(X) : X \in \mathbb{Z}_2^k\} \supseteq G$. Theorem 3 implies that, in the optimal case $\bar{\Gamma} = G$, if, say, $a = 0$ then $\max_{v \geq 0} \binom{k-v-a}{v}_D \leq |G_0| \leq |G| = |\bar{\Gamma}|$ and $G_0 \subseteq \bar{\Gamma}$ is a D-LM0EC code. Note that the case $a = 0$ would be equivalent to have a synchronizing "1" between X and $C(X)$. In Table I, for comparison purposes, we simply assumed $G_0 = \bar{\Gamma}$ and the existence of D-LM0EC length m codes with $UB(m, D)$ in (7) codewords, for all $m, D \in \mathbb{N}$. We computed the smallest length $m \stackrel{\text{def}}{=} m(K, D)$ such that $UB(m, D) \geq NCLB(k, D)$ in (11); and reported in the LB labelled columns the value of $r_{LB}(k, D) \stackrel{\text{def}}{=} m + D - 1$. We thought fair to add $D - 1$ check bits to m, in spite of the assumption $G_0 = \bar{\Gamma}$ and the synchronization non-guarantee between checks (codewords of a possibly existing D-LM0EC code) and data words.

IV. PROPOSED SYSTEMATIC CODE DESIGNS

Let $\langle x \rangle_b$ denote the $x \bmod b$ operation. The code design for $k \in \mathbb{N}$ data bits is defined as follows and relies on (5) and the max L_1 distance systematic q-ary code designs in [3, Subsection II.C], [6], where $q \stackrel{\text{def}}{=} q(X) \stackrel{\text{def}}{=} k - w(X) + 1$, for all data words $X \in \bigcup_{u \in [0,k]} \mathcal{S}(\mathbb{Z}_2, k, u) = \mathbb{Z}_2^k$. For any $u = w(X) \stackrel{\text{def}}{=} k - v$, the idea is to consider the vector $A(X) \in$

TABLE V

LOWER BOUND ON THE NUMBER OF CHECKS NEEDED IN **ANY** SYSTEMATIC CODE DESIGN AND THAT USED IN THE PROPOSED CODE DESIGN.

For some k and any $D\in\mathbb{N}$, the LB labelled columns give the lower bound in (11) and the CD columns the code design value in (14). If the entry is N_v then N is the number of distinct check symbols and v is a value where the maximum in (11) and (14) is obtained for the LB and CD columns, respectively.

$\backslash D$	2		3		4		5		6		7		8		9		$D\geq 10$	
$k\backslash$	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD	LB	CD
1	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0	1_0
2	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1	2_1
3	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1	3_1
4	4_1	4_1	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2	6_2
5	6_2	7_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2	10_2
6	10_2	11_2	16_3	17_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3	20_3
7	15_2	16_2	30_3	31_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3	35_3
8	21_2	26_3	50_3	51_3	65_4	66_4	70_4	70_4	70_4	70_4	70_4	70_4	70_4	70_4	70_4	70_4	70_4	70_4
9	35_3	42_3	90_4	96_4	120_4	121_4	126_4	126_4	126_4	126_4	126_4	126_4	126_4	126_4	126_4	126_4	126_4	126_4
10	56_3	64_3	161_4	168_4	216_5	222_5	246_5	247_5	252_5	252_5	252_5	252_5	252_5	252_5	252_5	252_5	252_5	252_5
11	84_3	99_4	266_4	294_5	413_5	420_5	455_5	456_5	462_5	462_5	462_5	462_5	462_5	462_5	462_5	462_5	462_5	462_5
12	126_4	163_4	504_5	540_5	728_5	756_6	875_6	882_6	917_6	918_6	924_6	924_6	924_6	924_6	924_6	924_6	924_6	924_6
13	210_4	256_4	882_5	927_5	1428_6	1464_6	1652_6	1660_6	1708_6	1709_6	1716_6	1716_6	1716_6	1716_6	1716_6	1716_6	1716_6	1716_6
14	330_4	386_4	1554_6	1711_6	2598_6	2643_6	3144_7	3180_7	3368_7	3376_7	3424_7	3425_7	3432_7	3432_7	3432_7	3432_7	3432_7	3432_7
15	495_4	638_5	2850_6	3061_6	4950_7	5115_7	6030_7	6075_7	6354_7	6363_7	6426_7	6427_7	6435_7	6435_7	6435_7	6435_7	6435_7	6435_7
16	792_5	1024_5	4917_6	5365_7	9240_7	9460_7	11385_8	11550_8	12465_8	12510_8	12789_8	12798_8	12861_8	12862_8	12870_8	12870_8	12870_8	12870_8
17	1287_5	1586_5	9042_7	9933_7	17205_8	17911_8	22110_8	22330_8	23760_8	23815_8	24210_8	24220_8	24300_8	24301_8	24310_8	24310_8	24310_8	24310_8
18	2002_5	2510_6	16236_7	17469_7	32802_8	33793_8	41470_9	42185_9	46420_9	46640_9	48070_9	48125_9	48520_9	48530_9	48610_9	48611_9	48620_9	48620_9
19	3003_5	4096_6	28314_8	31824_8	59950_9	62843_9	81367_9	82368_9	89232_9	89518_9	91652_9	91718_9	92257_9	92268_9	92367_9	92368_9	92378_9	92378_9

$\mathbb{Z}_q^{u+1}$ in (1) and encode the vector $\langle A(X)\rangle_D = A(X) \bmod D \in \mathbb{Z}_D^{u+1}$ into a check symbol $C(X)\in\Gamma$; where Γ is a self synchronizing non systematic D-LM0EC code in (9). The code Γ of length $r\in\mathbb{N}$ must be chosen big enough, but it can be independent of $u = w(X) = \ell(A(X)) - 1 \in [0,k]$ because of (5) and the assumption that the receiver knows the length of the received codeword. Specifically, the check symbol assignment map can be the union of $k+1$ distinct maps $C_u : \mathcal{S}(\mathbb{Z}_2,k,u)\to\Gamma$, each encoding $\langle A(X)\rangle_D \in \langle \mathcal{S}(\mathbb{Z}_q, u+1, v)\rangle_D \stackrel{\text{def}}{=} \mathcal{V}_u \subseteq \mathbb{Z}_D^{u+1}$; i. e., for all $u\in[0,k]$, for all $X\in\mathcal{S}(\mathbb{Z}_2,k,u)$ it must be

$$C(X) = C_u(X) = F_u(\langle A(X)\rangle_D);$$

with $F_u : \mathcal{V}_u \to \Gamma$ injective. Now, assume $\mathcal{E} = X\, C_{w(X)}(X) \in \mathbb{Z}_2^{k+r}$ is sent and $\hat{E} \stackrel{\text{def}}{=} \hat{X}\,\hat{C} \in \mathbb{Z}_2^*$ is received; where $\hat{X}$ and $\hat{C}$ are the erroneous version of $X\in\mathbb{Z}_2^k$ and $C_{w(X)}(X)\in\mathbb{Z}_2^r$, respectively. On receiving $\hat{E}$, the following (t_i,t_d)-LM0EC procedure is performed. From right to left, the receiver first corrects and parses $\hat{C}$ obtaining $C_{w(X)}(X)$ then it computes $w(\hat{X}) = w(X) = u$, and then it decodes $C(X) = C_u(X)$ to obtain $\langle A(X)\rangle_D \stackrel{\text{def}}{=} \alpha_1\alpha_2\ldots\alpha_{u+1}$. Then, it corrects the current received data part 0-run by computing its length, $\hat{a}_i \stackrel{\text{def}}{=} a_i(\hat{X})$, and then adding to $\hat{a}_i$ up to t_d or subtracting up to t_i so that $\hat{a}_i - \alpha_i$ becomes the nearest multiple of $D = t_i + t_d + 1$; this, until $\hat{X}\in(0^*10^*)^*$ has been parsed completely.

***Theorem** 4:* Let $\Gamma \stackrel{\text{def}}{=} \Gamma(k,D) \subseteq \mathbb{Z}_2^r$ be the set of check symbols required by the proposed D-LM0EC code design with k information bits, $k,D\in\mathbb{N}$. Then

$$|\Gamma| = \max_{v\geq 0} \sum_{i=0}^{\lfloor v/D\rfloor} \binom{k-v+1}{v-iD}_D. \tag{14}$$

Note that, if $D > \lfloor k/2\rfloor$ then $|\Gamma| = \binom{k}{\lceil k/2\rceil}_2$ is optimal since it reaches the lower bound in (12). In general, Γ is the shortest D-LM0EC code in (9) so that (14) holds and its length is given in the CD labelled column of Table I for some $k,D\in\mathbb{N}$.

Proof: For all $X\in\mathbb{Z}_2^k$, let $u \stackrel{\text{def}}{=} w(X)$ and $v \stackrel{\text{def}}{=} k-u$. Note that, $X\in\mathcal{S}(\mathbb{Z}_2,k,u) \implies A(X)\in\mathcal{S}(\mathbb{Z}_{v+1},u+1,v) \implies \langle A(X)\rangle_D \in \langle\mathcal{S}(\mathbb{Z}_{v+1},u+1,v)\rangle_D = \mathcal{V}_u$, where, really,

$$\mathcal{V}_u = \bigcup_{\nu\in[0,v]\cap(D\cdot\mathbb{Z}+\langle v\rangle_D)} \mathcal{S}(\mathbb{Z}_D, u+1, \nu).$$

Now, for all $u = k - v\in[0,k]$, the above map F_u must be injective (i. e., the check $C_u(X)$ must encode any possible value of $\langle A(X)\rangle_D \in \mathcal{V}_v$) so, the code Γ must have $|\Gamma| = \max_{v\geq 0}|\mathcal{V}_v| = \max_{v\geq 0}\sum_{i=0}^{\lfloor v/D\rfloor}\binom{u+1}{v-iD}_D$ elements. ■

***Example** 3: If $k=5$ and $D=3$ then $\max_{v\geq 0}|\mathcal{V}_v| = |\mathcal{V}_2| = |\mathcal{S}(\mathbb{Z}_3,4,2)| = \binom{4}{2}_3 = \binom{5}{3}_2 = 10$ and Γ can be chosen by picking, say, the first 10 words of the code $\bar{\mathcal{C}}_{6,3} = A^{-1}(3\cdot[\mathcal{S}(\mathbb{N},3,2)\cup\mathcal{S}(\mathbb{N},6,1)\cup\mathcal{S}(\mathbb{N},9,0)])$ in Table III. Its length $r=8$ is given in Table I.*

***Example** 4: If $k=6$ and $D=3$ then $\max_{v\geq 0}|\mathcal{V}_v| = |\mathcal{V}_3| = |\mathcal{S}(\mathbb{Z}_3,4,0)\cup\mathcal{S}(\mathbb{Z}_3,4,3)| = \binom{4}{3}_3 + \binom{4}{0}_3 = 16+1 = 17$ and Γ can be chosen by picking, say, the first 17 words of $\bar{\mathcal{C}}_{7,3} = A^{-1}(3\cdot[\mathcal{S}(\mathbb{N},1,3)\cup\mathcal{S}(\mathbb{N},4,2)\cup\mathcal{S}(\mathbb{N},7,1)\cup\mathcal{S}(\mathbb{N},10,0)])$. Its length $r=9$ is given in Table I.*

Because each $\mathcal{S}(\mathbb{Z}_q,n,w)$ can be indexed lexicographically as in [4], the entire coding process can be implemented with the indexing method as in [32] with $O(k^2\log D)$ bit operations by storing $O(k^2 D\log D)$ bits.

V. Concluding remarks

In this paper, some theory and design of optimal or close to optimal binary systematic D-LM0EC block codes are presented. Good lower and upper bounds are given for the number of distinct check symbols required by **any** systematic D-LM0EC block code in Theorem 2 and 4, respectively. In this respect, if $D > \lfloor k/2\rfloor$ then the number of distinct check symbols required by **any** code design is equal to the number of check symbols, $\binom{k}{\lceil k/2\rceil}$, used by the proposed codes. Remarkably, as D grows, the proposed codes outperform the wider sense systematic D-repetition codes given in Subsubsection II-A2 as shown, for example, in Table I.

REFERENCES

[1] K. A. S. Abdel-Ghaffar, F. Paluncic, H. C. Ferreira and W. A. Clarke, “On Helberg’s Generalization of the Levenshtein Code for Multiple Deletion/Insertion Error Correction”, *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1804–1808, March 2012.

[2] R. Ahlswede, H. Aydinian, L. H. Khachatrian, and L. M. G. M. Tolhuizen, “On q-ary codes correcting all unidirectional errors of a limited magnitude”, *Proceedings of Ninth International Workshop on Algebraic and Combinatorial Coding Theory*, Kranevo, Bulgaria, pp. 20–26, 2004.

[3] B. Bose, N. Elarief and L. G. Tallini, “On codes achieving zero error capacities in limited magnitude error channels”, *IEEE Transactions on Information Theory*, Vol.64, pp. 257–273, Jan. 2018.

[4] T. Cover, “Enumerative source encoding”, *IEEE Transactions on Information Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.

[5] L. Dolecek and V. Anantharam, “Repetition error correcting sets: Explicit constructions and prefixing methods”, *SIAM Journal on Discrete Mathematics*, vol. 23, no. 4, pp. 2120–2146, 2010.

[6] N. Elarief and B. Bose, “Optimal, Systematic, q-Ary Codes Correcting All Asymmetric and Symmetric Errors of Limited Magnitude”, *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 979–983, March 2010.

[7] H. C. Ferreira, W. A. Clarke, A. S. J. Helberg, K. A. S. Abdel-Ghaffar, and A. J. Han Vinck, “Insertion/Deletion Correction with Spectral Nulls”, *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 722–732, March 1997.

[8] L. Gargano, J. Körner and U. Vaccaro, “Capacities: From information theory to extremal set theory”, *Journal of Combinatorial Theory, Series A*, vol. 68, no. 2, pp. 296–316, Nov. 1994.

[9] V. Guruswami and J. Håstad. “Explicit two-deletion codes with redundancy matching the existential bound”, *32nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 21–32, Jan. 2021.

[10] A. S. J. Helberg and H. C. Ferreira, “On multiple insertion/deletion correcting codes”, *IEEE Transactions on Information Theory*, vol. 48, no. 1, pp. 305–308, Jan. 2002.

[11] S. Jain, F. Farnoud, M. Schwartz, and J. Bruck, “Duplication-Correcting Codes for Data Storage in the DNA of Living Organisms”, *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 4996–5010, Aug. 2017.

[12] W. Kautz, "Fibonacci codes for synchronization control", *IEEE Transactions on Information Theory*, vol. 11, no.2, pp. 284–292, April 1965.

[13] M. Kovacevic and V. T. F. Tan, “Asymptotically Optimal Codes Correcting Fixed-Length Duplication Errors in DNA Storage Systems”, *IEEE Communication Letters*, Vol. 12, no. 11, pp. 2094–2097, Nov. 2018.

[14] M. Kovačević, “Runlength-Limited Sequences and Shift-Correcting Codes: Asymptotic Analysis”, *IEEE Transactions on Information Theory*, Vol. 65, pp. 4804–4814, Aug. 2019.

[15] A. A. Kulkarni and N. Kiyavash, “Non-asymptotic Upper Bounds for Deletion Correcting Codes”, *IEEE Transactions on Information Theory*, Vol. 59, pp. 5115–5130, Aug. 2013.

[16] A. A. Kulkarni, “Insertion and deletion errors with a forbidden symbol”, *2014 IEEE Information Theory Workshop (ITW 2014)*, pp. 596–600, Nov. 2014.

[17] A. V. Kuznetsov and A. J. Han Vinck, “Single Peak-shift Correction In (d,k)-sequences”, *1991 IEEE International Symposium on Information Theory*, pp. 256–256, 1991.

[18] V. Levenshtein, “Binary codes with correction for deletions and insertions of the symbol 1’”, *Problems of Information Transmission*, vol. 1, no. 1, pp. 8–17, 1965.

[19] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions and reversals”, *Sov. Phys. Dokl.*, vol. 10, no. 8, pp. 707–710, 1966.

[20] V. I. Levenshtein and A. J. H. Vinck, “Perfect (d, k)-codes capable of correcting single peak-shifts”, *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 656–662, March 1993.

[21] L. Lovász, “On the Shannon Capacity of a Graph”, *IEEE Transactions on Information Theory*, vol. 25, no. 1, pp. 1–7, Jan. 1979.

[22] H. Mahdavifar and A. Vardy, “Asymptotically Optimal Sticky-Insertion-Correcting Codes with Efficient Encoding and Decoding”, *2017 IIEEE International Symposium on Information Theory*, pp. 2683–2687, June 2017.

[23] F. Palunčić, K. A. S. Abdel-Ghaffar, H. C. Ferreira, and W. A. Clarke, “A Multiple Insertion/Deletion Correcting Code for Run-Length Limited Sequences”, *IEEE Transactions on Information Theory*, vol. 58, pp. 1809–1824, March 2012.

[24] C. Shannon, “The zero error capacity of a noisy channel”, *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sept. 1956.

[25] J. Sima and J. Bruck, “Optimal k-Deletion Correcting Codes", *2019 IEEE International Symposium on Information Theory*, pp. 847–851, July 2019.

[26] J. Sima, N. Raviv and J. Bruck, “Two Deletion Correcting Codes From Indicator Vectors”, *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2375–2391, April 2020.

[27] J. Sima, R. Gabrys and J. Bruck, “Optimal Systematic t-Deletion Correcting Codes”, *2020 IEEE International Symposium on Information Theory*, pp. 769–774, 2020.

[28] J. Sima, R. Gabrys and J. Bruck, “Optimal Codes for the q-ary Deletion Channel”, *2020 IEEE International Symposium on Information Theory*, pp. 740–745, 2020.

[29] N. J. A. Sloane, “On single-deletion-correcting codes", in *Codes and Designs*, Ohio State University (Ray-Chaudhuri Festschrift), pp. 273–291, 2000. Online: https://arxiv.org/abs/math/0207197.

[30] L. G. Tallini, N. Alqwaifly and B. Bose, “Deletions and Insertions of the Symbol “0” and Asymmetric/Unidirectional Error Control Codes for the L_1 Metric”, *IEEE Transactions on Information Theory*, vol. 69, no. 1, pp. 86–106, Jan. 2023.

[31] L. G. Tallini, N. Alqwaifly and B. Bose, “Efficient Systematic Deletions/Insertions of 0’s Error Control Codes”, *2022 IEEE Information Theory Workshop (ITW)*, pp. 570-575, 2022. Extended version online: https://arxiv.org/abs/2302.06563

[32] L. G. Tallini, N. Alqwaifly and B. Bose, “Zero Deletion/Insertion Codes and Zero Error Capacity”, *2022 IEEE International Symposium on Information Theory*, pp. 986–991, 2022.

[33] L. G. Tallini, N. Alqwaifly and B. Bose, “On Deletion/Insertion of Zeros and Asymmetric Error Control Codes”, *2019 IEEE International Symposium on Information Theory*, pp. 2384–2388, 2019.

[34] L. G. Tallini and B. Bose, “On L_1-distance error control codes”, *2011 IEEE International Symposium on Information Theory*, pp. 1026–1030, July/Aug. 2011.

[35] L. Tallini, B, Bose and N. Elarief, “On efficient repetition error correcting codes”, *2010 IEEE International Symposium on Information Theory*, pp. 1012–1016, June 2010.

[36] L. G. Tallini and B. Bose, “On efficient high-order spectral-null codes”, *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2594–2601, Nov. 1999.