Deletions and Insertions of the Symbol "0" and Asymmetric/Unidirectional Error Control Codes for the L_1 Metric

Luca G. Tallini, Nawaf Alqwaifly, and Bella Bose, *Life Fellow, IEEE*

***Abstract*— This paper gives some theory and efficient design of binary block codes capable of controlling the deletions of the symbol "0" (referred to as 0-deletions) and/or the insertions of the symbol "0" (referred to as 0-insertions). This problem of controlling 0-deletions and/or 0-insertions (referred to as 0-errors) is shown to be equivalent to the efficient design of L_1 metric asymmetric error control codes over the natural alphabet, $\mathbb{N}$. In this way, it is shown that the t 0-insertion correcting codes are actually capable of controlling much more; namely, they can correct t 0-errors, detect $(t+1)$ 0-errors and, simultaneously, detect all occurrences of only 0-deletions or only 0-insertions in every received word (briefly, they are t-Symmetric 0-Error Correcting/$(t+1)$-Symmetric 0-Error Detecting/All Unidirectional 0-Error Detecting (t-Sy0EC/$(t+1)$-Sy0ED/AU0ED) codes). From the relations with the L_1 distance error control codes, new improved bounds are given for the optimal t 0-error correcting codes. Optimal non-systematic code designs are given. Decoding can be efficiently performed by algebraic means using the Extended Euclidean Algorithm (EEA).**

***Index Terms*— Deletion/insertion of zero errors, repetition/sticky errors, L_1 distance, asymmetric distance, elementary symmetric functions, constant weight codes.**

I. Introduction

LET A^* be the set of all finite length sequences over an alphabet A. In this paper, we are interested in the efficient design of binary block codes capable of correcting $t \in \mathbb{N}$ or less deletions and/or insertions of a fixed binary symbol, say, $0 \in \mathbb{Z}_2 \stackrel{\text{def}}{=} \{0,1\} \subseteq \mathbb{N}$. In this error model, if

$$X = 0100101000101110 \in \mathbb{Z}_2^{16} \tag{1}$$

is a transmitted binary sequence of length $n = 16$, then

$$\begin{aligned} Y &= 0\mathbf{0}10\lambda1\lambda100\mathbf{0}01\lambda11\mathbf{00}1\mathbf{0}0 \\ &= 001011000011100100 \qquad \in \mathbb{Z}_2^{18} \end{aligned} \tag{2}$$

is the received word obtained from X due to 3 deletions (λ represents the empty symbol) and 5 insertions of the symbol 0. The problem of designing efficient codes to control these types of 0-deletion and/or insertion errors (briefly, 0-errors) is an open research problem introduced by Levenshtein in [21] which is important for at least two reasons. From the application perspective, through the Gray mapping, correcting t deletions or insertions of 0's is equivalent to correcting t repetition errors [37] (or, sticky errors) which occur in high speed communication and data storage systems due to synchronization loss [11], [25], [37]. From the theoretical perspective, the design problem of t deletion and/or insertion of 0's Error Correcting (i. e., t-Symmetric 0-Error Correcting (t-Sy0EC)) codes is important because it is a particular instance of the general problem also introduced by Levenshtein in [22]. Even though the general problem of designing asymptotically optimal codes capable of correcting at most t deletions and/or insertions of any symbol appears to be very difficult [14], [15], [20], [22], [24], [31], [32], some efficient solutions have been given recently for the particular problems of correcting the 0-insertion errors (i. e., the insertion of 0's only) [11], [25] and the 0-errors (i. e., the deletion and/or insertion of 0's) [37]. In general many other insertion/deletion (edit) channel models have been considered in the literature [1], [2], [13], [16], [26]. Note that solution to any restricted insertion/deletion channel models may give hints on how to solve the general problem.

With regard to the 0-error problem, for all $X, Y \in \mathbb{Z}_2^*$, let

$$\begin{aligned} d_{0\text{-}D/I}(X,Y) \stackrel{\text{def}}{=} &\text{the minimum number of deletions} \\ &\text{and/or insertions of 0's needed to} \\ &\text{transform the binary word } X \text{ to } Y. \end{aligned} \tag{3}$$

For example, if X and Y are the words given in (1) and (2) respectively, then $d_{0\text{-}D/I}(X,Y) = 8$. The above function introduced in [21] is a distance (called here the deletion/insertion of 0's distance or 0-error distance). In fact, it is a graph distance defined in the graph (N, E) where the set of nodes is $N \stackrel{\text{def}}{=} \mathbb{Z}_2^*$ and the set of edges is

$$E \stackrel{\text{def}}{=} \left\{(X,Y) \in N^2 : \ d_{0\text{-}D/I}(X,Y) = 1\right\}.$$

Synchronization errors due to 0-errors can be controlled by inserting a marker or synchronization sequence between consecutive codewords in the sequences that are sent [13], [21]. Thus, we assume no synchronization errors due to erroneous

Manuscript received 10 December 2021; revised 19 June 2022; accepted 12 August 2022. Date of publication 2 September 2022; date of current version 22 December 2022. This work was supported by the NSF under Grant CCF-2006571. An earlier version of this paper was presented in part at the 2019 IEEE International Symposium on Information Theory [DOI: 10.1109/ISIT.2019.8849470]. *(Corresponding author: Luca G. Tallini.)*

Luca G. Tallini is with the Facoltà di Scienze della Comunicazione, Università degli Studi di Teramo, Campus di Coste Sant'Agostino, 64100 Teramo, Italy (e-mail: ltallini@unite.it).

Nawaf Alqwaifly is with the College of Engineering, Qassim University, Buraydah 52571, Saudi Arabia (e-mail: alqwaifly@qec.edu.sa).

Bella Bose is with the School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331 USA (e-mail: bella.bose@oregonstate.edu).

Communicated by R. Gabrys, Associate Editor for Coding and Decoding.

Digital Object Identifier 10.1109/TIT.2022.3203594

receptions of sequences of codewords (i. e., we assume that the receiver knows the length of the received word). In this case, since 1-errors are forbidden in our error model,

$$w_H(X) \neq w_H(Y) \iff d_{0\text{-}D/I}(X,Y) = \infty; \qquad (4)$$

where $w_H(Z) \in \mathbb{N}$ denotes the Hamming weight of any $Z \in \mathbb{N}^*$. In this way, the metric space $(\mathbb{Z}_2^*, d_{0\text{-}D/I})$ or its associated graph (N, E) remains partitioned into many distinct connected components, one for each possible Hamming weight, $w = w_H(X)$, of words $X \in \mathbb{Z}_2^*$. The major contributions of the paper are as follows:

1) In Section II, an isometry is explicitly defined in (15) which shows that the design of t-Sy0EC codes is equivalent to the design of the L_1 error correcting codes. In this way, many combinatorial characterizations are derived for t-Sy0EC codes. In particular, it is shown that t-Sy0EC codes are equivalent to the disjoint union of some L_1 metric t-SyEC constant weight codes over $\mathbb{N}$ of distinct weights and lengths. These last codes can correct up to t symmetric L_1 metric errors, detect up to $(t+1)$ symmetric L_1 metric errors and, simultaneously, detect all occurrences of only negative or only positive L_1 metric errors in every received word (i. e., they are t Symmetric Error Correcting, $(t+1)$ Symmetric Error Detecting and All Unidirectional Error Detecting codes; or briefly, t-SyEC/$(t+1)$-SyED/AUED codes [4]). For these reasons, the combinatorial equivalence holds among t 0-deletion error correcting codes, t 0-insertion error correcting codes, t-Sy0EC codes (already proved by Levenshtein in [21]) and the more powerful t-Sy0EC/$(t+1)$-Sy0ED/AU0ED codes capable of correcting up to t symmetric 0-errors, detecting up to $(t+1)$ symmetric 0-errors and, simultaneously, detecting all occurrences of only 0-deletions or only 0-insertions in every received word.
2) In Section III, the general Algorithm 3.1 is defined in Subsection III-B which efficiently reduces the t-SyEC/$(t+1)$-SyED/AUED decoding design problem for constant weight codes to the less powerful (τ_-, τ_+)-EC decoding design problem for the L_1 metric. This implies that any efficient (τ_-, τ_+)-EC scheme gives an efficient t-SyEC/$(t+1)$-SyED/AUED scheme which, in turn, gives efficient t-Sy0EC/$(t+1)$-Sy0ED/AU0ED codes because of the isometry discussed in Section II. In this way, based on the σ-code theory in [35], [36], [37], [38], [39], [40], and [41], some non-systematic t-Sy0EC/$(t+1)$-Sy0ED/AU0ED codes together with their efficient t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding algorithms are designed. Note that, in [11] and [25], the authors have given codes and decoding algorithms for t sticky-insertion error correcting codes which can correct at most t insertions of a repeated symbol [37] and are equivalent, through the Gray mapping, to t 0-insertion (only) error correcting codes. Such codes are constructed over prime fields and over the Lee metric. Here, it is shown that we can use the simpler L_1 metric and prime power fields to design them and because of these reasons the proposed codes give better information rates and error control performances (of t-Sy0EC/$(t+1)$-Sy0ED/AU0ED) than those in [11] and [25]. In [37], for fixed $t_-, t_+ \in \mathbb{N}$, codes with decoding algorithms are given which can (only) simultaneously correct t_- sticky-deletions and t_+ sticky-insertions (named (t_-, t_+)-Insertion/Deletion Of Repeated Symbol Error Correcting codes). Such codes are designed by 1) Gray map reducing the t_- sticky-deletion and t_+ sticky-insertion error correction problem to the t_- 0-deletion and t_+ 0-insertion error correction problem which, in turn, is 2) reduced into the t_- negative and t_+ positive error correction problem for the L_1 distance over $\mathbb{N}$ by using the "bucket of 0's mapping" defined here in (15). In [16], and [21], some code design are given for t-Sy0EC. However, only totally asymmetric error correcting algorithm (i. e., correcting only 0-deletions or only 0-insertions) are shown and neither explicit nor practical algorithms are defined which can perform (at least) t-Sy0EC [16]. Here, based on the σ-code theory and the above two mentioned reductions, t-Sy0EC code designs are explicitly defined together with their efficient algebraic t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding algorithm. In this way, the code's maximal error control capabilities are developed.
3) Section IV focuses on obtaining new non-asymptotic bounds for the largest cardinality, $D(n,t)$, of a t-Sy0EC/$(t+1)$-Sy0ED/AU0ED binary code of length n and what are their consequences in terms of the asymptotic bounds. To our knowledge, no non-asymptotic bounds are given for this coding problem and this perspective is new. In this way, we improve/generalize known asymptotic bounds. In general, thanks to the reduction of the design problem of t-Sy0EC codes to the design problem of L_1 metric t-SyEC constant weight codes over $\mathbb{N}$, **any** bound on the largest cardinality, $CW(\mathbb{N}, n, w, t) \in \mathbb{N}$, of a **constant** weight w code of length n over the alphabet $\mathbb{N}$ with minimum symmetric L_1 distance greater than $2t$ gives a bound to $D(n,t)$. In this way, new non-asymptotic and asymptotic lower and upper bounds are given here for $D(n,t)$. The lower bounds follow from the general σ-code based design. For t fixed, the Sidon set based codes in [16] give very good asymptotic bounds which improve the asymptotic bounds in [11], and [21]. For all $n, t \in \mathbb{N}$, the cardinality of our σ-code theory based codes is slightly bigger than the Sidon set based codes in [16]. The non-asymptotic upper bound (which follows from a simple sphere packing argument for L_1 metric t-SyEC constant weight codes over $\mathbb{N}$) allows to derive an interesting upper bound, plotted in Figure 1, on the asymptotic information rate of any infinite family of t-Sy0EC codes of length n. Noticeably, such upper bound improves on the general case upper bound in [20] for the values of $\tau = t/n$ which roughly belong to the real set $[0.28, 0.35] \cup [0.43, 1)$.

Some concluding remarks are given in Section V.

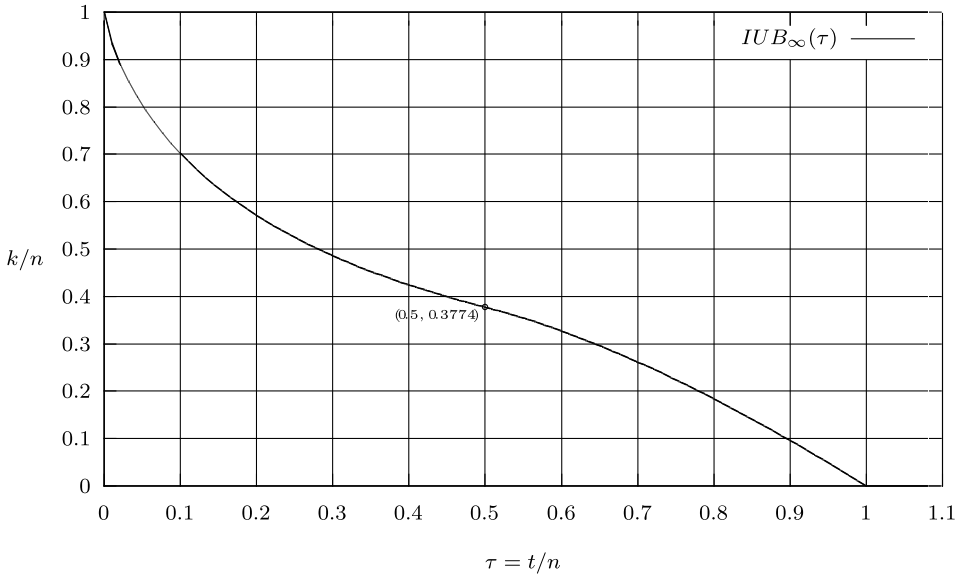

Fig. 1. Upper bound on the asymptotic information rate of t-Sy0EC codes. The asymptotic information rate of any code design is below the above curve.

II. 0-Deletion/Insertion Errors and the L_1 Metric

In this section, it is shown that the design problem of t-Sy0EC codes is equivalent to the design problem of some L_1 metric asymmetric error control codes over the natural alphabet, $\mathbb{N}$. Before describing this result, some background materials are given first.

For $m \in \mathbb{N} \cup \{\infty\}$ let

$$\mathbb{Z}_m \stackrel{\text{def}}{=} \{0, 1, \ldots, m-1\} \subseteq \mathbb{N} \stackrel{\text{def}}{=} \mathbb{Z}_\infty.$$

Also, for $x, y \in \mathbb{Z}_m$, define the natural subtraction as $x \dot{-} y = \max\{0, x - y\}$. For example, if $x = 2$ and $y = 0$ then $x \dot{-} y = 2$ and $y \dot{-} x = 0$. Given any two words $X, Y \in \mathbb{Z}_m^n$ of length $n \in \mathbb{N}$, the operations $X \cap Y \in \mathbb{Z}_m^n$, $X \cup Y \in \mathbb{Z}_m^n$, $X + Y \in \mathbb{N}^n$, and $X \dot{-} Y \in \mathbb{Z}_m^n$ are defined as the digit by digit $\min$, $\max$, integer addition and $\dot{-}$ operation between X and Y, respectively. For example, if $m = 3$, $n = 9$, $X = 012012012$ and $Y = 000111222$ then $X \cap Y = 000011012$, $X \cup Y = 012112222$, $X + Y = 012123234$, $X \dot{-} Y = 012001000$ and $Y \dot{-} X = 000100210$. In addition, the support of a word $X = x_1 x_2 \ldots x_n \in \mathbb{Z}_m^n$ is $\partial X = s_1 s_2 \ldots s_n \in \mathbb{Z}_2^n$ where $s_i = 1$ if $x_i \neq 0$ and $s_i = 0$ otherwise. For example $\partial(42101) = (11101)$. Given a support ∂S as an index set, say $\partial S = [1, n]$, every word in $X = x_1 x_2 \ldots x_n \in \mathbb{Z}_m^n$ can be regarded as a multiset over the index set ∂S where each component, x_i of X defines the multiplicity of $i \in \partial S$ as an element of X. In this way, there is a one-to-one correspondence between m-ary words and multisets; and the above operations can be regarded as multisets operations too. So, in the following, we will identify m-ary words of length n with multisets over an index set containing n distinct elements (which, for code construction purposes, will be contained in a field). The cardinality of a word/multiset $X = x_1 x_2 \ldots x_n \in \mathbb{Z}_m^n$ is the L_1 weight of X and is naturally defined as the real sum

$$|X| \stackrel{\text{def}}{=} w_{L_1}(X) \stackrel{\text{def}}{=} \sum_{i \in \partial S} x_i.$$

For example, $|01232| = w_{L_1}(01232) = 8$. Note that for $m = 2$ the L_1 weight and the Hamming weight coincide. So, when this creates no confusion we will indicate the weight of X as $w(X)$.

To better describe the error control properties of codes for the L_1 metric, the following distances between m-ary words $X, Y \in \mathbb{Z}_m^n$ are considered in [38] and [40] (the "+" sign below indicates an integer sum).

$$\begin{aligned} \text{symmetric } L_1\text{: } d_{L_1}^{sy}(X,Y) &\stackrel{\text{def}}{=} |Y \dot{-} X| + |X \dot{-} Y|, \qquad (5)\\ \text{asymmetric } L_1\text{: } d_{L_1}^{as}(X,Y) &\stackrel{\text{def}}{=} \max\{|Y \dot{-} X|, |X \dot{-} Y|\},\\ \text{Hamming: } d_H(X,Y) &\stackrel{\text{def}}{=} |\partial(Y \dot{-} X)| + |\partial(X \dot{-} Y)|. \end{aligned}$$

For example, if $m = 5$, $n = 5$, $X = 01423$, $Y = 43213$ then $|X \dot{-} Y| = 3$, $|Y \dot{-} X| = 6$, $|\partial(X \dot{-} Y)| = 2$, $|\partial(Y \dot{-} X)| = 2$ and $d_{L_1}^{sy}(X,Y) = 3 + 6 = 9$, $d_{L_1}^{as}(X,Y) = \max\{6, 3\} = 6$ and $d_H(X,Y) = 2 + 2 = 4$. From the error control perspective, if X is the transmitted word and Y is the received word then $Y \dot{-} X$ and $X \dot{-} Y$ give the increasing and decreasing error vectors, respectively. Thus,

$$X = Y - (Y \dot{-} X) + (X \dot{-} Y).$$

Note that,

$$\text{for all } X, Y \in \mathbb{Z}_m^n, \quad d_H(X,Y) \leq d_{L_1}^{sy}(X,Y) \qquad (6)$$

because $w_H(X) = |\partial X| \leq |X| \stackrel{\text{def}}{=} w_{L_1}(X)$, for all $X \in \mathbb{Z}_m^n$.

Constant weight codes play an important role in what follows. Thus, given $n, w \in \mathbb{N}$ and any numeric set $A \subseteq \mathbb{N}$ as alphabet, let

$$\mathcal{S}(A, n, w) \stackrel{\text{def}}{=} \{X \in A^n : \ w_{L_1}(X) = |X| = w\} \qquad (7)$$

be the set of all words over A of length n and constant weight w. We readily note, from (7), that

$$\mathcal{S}(A,n,w)=\bigcup_{x\in A}\mathcal{S}(A,n-1,w-x)x; \tag{8}$$

where the above union is a disjoint union of sets and $\mathcal{S}x\subseteq A^n$ indicates the set of words obtained concatenating every word in the set $\mathcal{S}\subseteq A^{n-1}$ with $x\in A$. Hence, the general recurring formula,

$$|\mathcal{S}(A,n,w)|=\sum_{x\in A}|\mathcal{S}(A,n-1,w-x)|, \tag{9}$$

holds for, say, the "A-nominal coefficient n choose w", $|\mathcal{S}(A,n,w)|$. If $A=\mathbb{Z}\!\!\mathbb{Z}_m$ then the cardinality of the above set is the m-nominal coefficient n choose w

$$|\mathcal{S}(\mathbb{Z}\!\!\mathbb{Z}_m,n,w)|=\binom{n}{w}_m=\sum_{v=0}^{m-1}\binom{n-1}{w-v}_m, \tag{10}$$

for all integers $m\in\mathbb{I}\!\mathbb{N}$. The quantity $\binom{n}{w}_m$ is the coefficient of the monomial z^w in the standard form of the polynomial $[1+z+\ldots+z^{(m-1)}]^n$ which, for $m=2$, reduces to the usual binomial coefficient (i. e., $\binom{n}{w}_2=\binom{n}{w}$). The m-nomial coefficient sequence has been studied in the ambit of m-ary unordered codes and share many properties with the binomial coefficient sequence obtained for $m=2$ [28]. If instead, $A=\mathbb{Z}\!\!\mathbb{Z}_\infty=\mathbb{I}\!\mathbb{N}$ then we can define

$$\binom{n}{w}_\infty\stackrel{\text{def}}{=}|\mathcal{S}(\mathbb{I}\!\mathbb{N},n,w)|$$

and note that the cardinality of $\mathcal{S}(\mathbb{I}\!\mathbb{N},n,w)$ is the composition of a natural number w into n natural numbers. In this way,

$$|\mathcal{S}(\mathbb{I}\!\mathbb{N},n,w)|\stackrel{\text{def}}{=}\binom{n}{w}_\infty=\binom{n+w-1}{n-1}=\binom{n+w-1}{w}. \tag{11}$$

In this case, the recursive formula (9) becomes

$$|\mathcal{S}(\mathbb{I}\!\mathbb{N},n,w)|=\binom{n}{w}_\infty=\sum_{v=0}^{w}\binom{n-1}{v}_\infty= \sum_{v=0}^{w}\binom{n+v-2}{n-2}=\binom{n+w-1}{n-1} \tag{12}$$

because $x\geq 0$ and $(w-x)\geq 0$ ($\iff x,(w-x)\in\mathbb{I}\!\mathbb{N}$).

Now, if $X\in\mathbb{Z}\!\!\mathbb{Z}_2^*$ then X can be uniquely written as [21], [23],

$$X=0^{v_1}10^{v_2}10\ldots010^{v_w}10^{v_{w+1}} \tag{13}$$

where $l=l(X)\in\mathbb{I}\!\mathbb{N}$ indicates the length of any $X\in A^*$, $w=w_H(X)\in[0,l(X)]$ is the Hamming weight of X and, for all integers $i\in[1,w+1]$, $v_i\stackrel{\text{def}}{=}v_i(X)\in\mathbb{Z}\!\!\mathbb{Z}_{l-w+1}\subseteq\mathbb{I}\!\mathbb{N}$ is the i-th run length of 0's in the word X. Note that

$$v_{w+1}=(l(X)-w(X))-\sum_{i=1}^{w}v_i. \tag{14}$$

Given the above representation, consider the following bijective function (which we call here the bucket of 0's mapping)

$$V:\mathbb{Z}\!\!\mathbb{Z}_2^*\to\mathbb{Z}\!\!\mathbb{Z}_\infty^*=\mathbb{I}\!\mathbb{N}^* \tag{15}$$

TABLE I

THE MAPPING V ACTING ON $\mathbb{Z}\!\!\mathbb{Z}_2^4$. IN THE TABLE $v_{w(X)+1}$ IS IN BOLDFACE AND $l(X)$ INDICATES THE LENGTH OF ANY $X\in A^*$

$l(X)=n$	$w(X)$	X	$V(X)=\hat{V}(X)\boldsymbol{v_{w+1}}$	$l(V(X))$	$w(V(X))$
4	0	0000	**4**	1	4
4	1	0001	3**0**	2	3
		0010	2**1**		
		0100	1**2**		
		1000	0**3**		
4	2	0011	2**0**0	3	2
		0101	1**0**0		
		0110	1**0**1		
		1001	0**2**0		
		1010	0**1**1		
		1100	0**0**2		
4	3	0111	1**0**0**0**	4	1
		1011	0**1**0**0**		
		1101	0**0**1**0**		
		1110	0**0**0**1**		
4	4	1111	0000**0**	5	0

which associates any $X\in\mathbb{Z}\!\!\mathbb{Z}_2^*$ represented as in (13) with

$$V(X)\stackrel{\text{def}}{=}(v_1,v_2,\ldots,v_w,v_{w+1})\in\mathbb{I}\!\mathbb{N}^*.$$

For example, if

$$X=01\,001\,01\,0001\,01\,1\,1\,0000000\in\mathbb{Z}\!\!\mathbb{Z}_2^*$$

then

$$V(X)=(1,2,1,3,1,0,0,7)\in\mathbb{I}\!\mathbb{N}^*.$$

The mapping V in (15), already considered by Levenshtein in [21], defines a bijection from the set of all binary words of any finite length $n\in\mathbb{I}\!\mathbb{N}$ and Hamming weight w (=number of 1's of the binary words) into the words over $\mathbb{I}\!\mathbb{N}$ of length $w+1$ (=number of buckets defined by the w 1's of the binary words) and L_1 weight $n-w$ (= number of 0's of the binary words). Except for the rightmost "1" which is dropped, the function

$$V^{-1}:\mathbb{Z}\!\!\mathbb{Z}_\infty^*=\mathbb{I}\!\mathbb{N}^*\to\mathbb{Z}\!\!\mathbb{Z}_2^*$$

is nothing but the prefix free unary representation of a sequence of integer numbers. Hence, both V and V^{-1} are one-to-one mappings such that

$$V(\mathcal{S}(\mathbb{Z}\!\!\mathbb{Z}_2,n,w))=\mathcal{S}(\mathbb{I}\!\mathbb{N},w+1,n-w),$$

and

$$\mathcal{S}(\mathbb{Z}\!\!\mathbb{Z}_2,n,w)=V^{-1}(\mathcal{S}(\mathbb{I}\!\mathbb{N},w+1,n-w)).$$

For example, for $n=4$, the mapping V acts on $\mathbb{Z}\!\!\mathbb{Z}_2^4$ is as reported in Table I. Let

$$\hat{V}:\mathbb{Z}\!\!\mathbb{Z}_2^*\to\mathbb{I}\!\mathbb{N}^* \tag{16}$$

be the function obtained from V by dropping the last component; $\hat{V}$ associates any $X\in\mathbb{Z}\!\!\mathbb{Z}_2^*$ represented as in (13) with

$$\hat{V}(X)\stackrel{\text{def}}{=}(v_1,v_2,\ldots,v_w)\in\mathbb{I}\!\mathbb{N}^*.$$

Obviously, since V is a one-to-one function, it is possible to reconstruct X from $V(X)$; likewise, even though $\hat{V}$ is not one-to-one (for example, $\hat{V}(0110) = \hat{V}(011000) = (1,0)$), it is possible to reconstruct X from $\hat{V}(X)$ and $n = l(X)$ because of (14). In this case, v_{w+1} can be considered as a parity digit which makes the L_1 weight $w_{L_1}(V(X)) = n-w$. Both functions V and $\hat{V}$ play important roles in our code designs and analysis. Consider the following example words

$$\begin{aligned} X &= 01\,001\,01\,0001\,01\,1\,1\,0 \quad \in \mathbb{Z}_2^{16}, \\ Y &= 001\,001\,1\,00001\,1\,1\,001\,00 \in \mathbb{Z}_2^{19}, \\ Y' &= 001\,001\,01\,0001\,01\,00 \quad \in \mathbb{Z}_2^{16}. \end{aligned}$$

Then their associated V values are

$$\begin{aligned} V(X) &= (1,2,1,3,1,0,0,\mathbf{1}) \in \mathbb{N}^8, \\ V(Y) &= (2,2,0,4,0,0,2,\mathbf{2}) \in \mathbb{N}^8, \\ V(Y') &= (2,2,1,3,1,\mathbf{2}) \quad \in \mathbb{N}^6. \end{aligned}$$

Note that if X is sent, Y' can never be received because $7 = w(X) \neq w(Y') = 5$ and 1-errors are forbidden in our channel model; whereas, Y can erroneously be received and the number of 0-deletions ($= 2$) plus the number of 0-insertions ($= 5$) from X to Y is equal to the L_1 distance between $V(X)$ and $V(Y)$, $d_{L_1}^{sy}(V(X),V(Y)) = 2+5 = 7$. In fact, in general, a sequence $Y \in \mathbb{Z}_2^*$ is obtained from the sequence $X \in \mathbb{Z}_2^*$ due to t_- deletions and t_+ insertions of the symbol 0 if, and only if, $w(Y) = w(X)$ and $d_{L_1}^{sy}(V(Y),V(X)) = t_- + t_+$; that is, $V(Y)$ is obtained from $V(X)$ due to a negative error pattern of magnitude t_- and a positive error pattern of magnitude t_+. Hence, the bucket of 0's mapping $X \to V(X)$ reduces the t_- 0-deletion and t_+ 0-insertion error correction problem into the t_- negative and t_+ positive error correction problem for the L_1 distance over $\mathbb{N}$.

Theorem 2.1 (Isometry Between $(\mathbb{Z}_2^*, d_{0\text{-}D/I})$ and $(\mathbb{N}^*, d_{L_1}^{sy})$): For all $X, Y \in \mathbb{Z}_2^*$,

$$d_{0\text{-}D/I}(X,Y) = \begin{cases} d_{L_1}^{sy}(V(X),V(Y)) & \text{if } w(X)=w(Y), \\ \infty & \text{if } w(X)\neq w(Y). \end{cases} \tag{17}$$

Relation (17) implies that $d_{0\text{-}D/I}(X,Y) < \infty$ if, and only if, $w(X) = w(Y)$. So, if we extend the domain of $d_{L_1}^{sy}$ from $\mathbb{N}^l \times \mathbb{N}^l$, $l \in \mathbb{N}$, to $\mathbb{N}^* \times \mathbb{N}^*$ by letting $d_{L_1}^{sy}(U,V) = \infty$ whenever $l(U) \neq l(V)$ then,

$$\text{for all } X, Y \in \mathbb{Z}_2^*, \quad d_{0\text{-}D/I}(X,Y) = d_{L_1}^{sy}(V(X),V(Y)).$$

This implies that the mapping V in (15) is an isometry between the metric spaces $(\mathbb{Z}_2^*, d_{0\text{-}D/I})$ and $(\mathbb{N}^*, d_{L_1}^{sy})$.

Proof: In order to prove (17), assume $X, Y \in \mathbb{Z}_2^*$ with $w \stackrel{\text{def}}{=} w(X) = w(Y)$ and recall the definition (3) of $d_{0\text{-}D/I}$. Let $e \in \mathbb{N}$ be the number of 0-deletions and $f \in \mathbb{N}$ be the number of 0-insertions needed to transform the binary word X to Y in such a way that $d_{0\text{-}D/I}(X,Y) = e+f$. In particular, let

$$e \stackrel{\text{def}}{=} e_1 + e_2 + \ldots + e_{w+1}$$

and

$$f \stackrel{\text{def}}{=} f_1 + f_2 + \ldots + f_{w+1};$$

where $e_i \in \mathbb{N}$ is the number of 0-deletions occurred in the i-th run of 0's of X and $f_i \in \mathbb{N}$ is the number of 0-insertions occurred in the i-th run of 0's of X, for all integers $i \in [1, w+1]$. In this way,

$$d_{0\text{-}D/I}(X,Y) = e + f = \sum_{i=1}^{w+1} (e_i + f_i) \tag{18}$$

and $e_i + f_i \in \mathbb{N}$ is the number of 0-operations (i. e., deletions and/or insertions of the symbol 0) in the i-th run of 0's to transform X to Y, for all $i \in [1, w+1]$. From the minimality requirement in (3), since $d_{0\text{-}D/I}(X,Y) = e+f$, the quantity $e_i + f_i$ is minimum, for all $i \in [1, w+1]$. Note that

$$v_i(X) = v_i(Y) + e_i - f_i, \quad \text{for all integers } i \in [1, w+1]. \tag{19}$$

Now, let $i \in [1, w+1]$ be given. If $e_i \geq f_i \geq 0$ then $f_i = 0$ because $e_i + f_i$ is minimum number of 0-operations in the i-th run of 0's to transform X to Y. Hence, from (19), the absolute value,

$$|v_i(X) - v_i(Y)| = v_i(X) - v_i(Y) = e_i \geq 0.$$

if instead, $f_i \geq e_i \geq 0$ then $e_i = 0$ because $e_i + f_i$ is the minimum number of 0-operations in the i-th run of 0's to transform X to Y. Hence, from (19), $v_i(Y) = v_i(X) + f_i - e_i$, and

$$|v_i(X) - v_i(Y)| = v_i(Y) - v_i(X) = f_i \geq 0.$$

In any case, the absolute value satisfies

$$|v_i(X) - v_i(Y)| = e_i + f_i,$$

for any given $i \in [1, w+1]$. Hence,

$$d_{L_1}^{sy}(V(X),V(Y)) = \sum_{i=1}^{w+1} |v_i(X) - v_i(Y)| = e + f = d_{0\text{-}D/I}(X,Y).$$

because of (18). ∎

In general, the isometry V in (15) reduces the design problem of error control codes for the insertion/deletion of 0's problem to the design problem of error control codes under the L_1 metric. In particular, for all $w \in [0,n]$, the one-to-one function V transforms any word $X \in \mathcal{S}(\mathbb{Z}_2, n, w) \subseteq \mathbb{Z}_2^n$ into a word $V(X) = (v_1, v_2, \ldots, v_{w+1}) \in \mathcal{S}(\mathbb{N}, w+1, n-w) \subseteq \mathbb{Z}_{n-w+1}^{w+1}$. Furthermore, any fixed length $n \in \mathbb{N}$ binary code, $\mathcal{C} \subseteq \mathbb{Z}_2^n$, is union of block (i. e., constant) length $n \in \mathbb{N}$ constant weight $w \in [0,n]$ codes, where the union is over w; say, $\mathcal{C} = \bigcup_{w \in [0,n]} \mathcal{C}_w$, with $\mathcal{C}_w \subseteq \mathcal{S}(\mathbb{Z}_2, n, w)$. So, the image of $\mathcal{C}$ through the isometry V is equal to

$$V(\mathcal{C}) = V\left(\bigcup_{w \in [0,n]} \mathcal{C}_w\right) = \bigcup_{w \in [0,n]} V(\mathcal{C}_w) \subseteq \mathbb{N}^n$$

with $\mathcal{A}_w \stackrel{\text{def}}{=} V(\mathcal{C}_w) \subseteq \mathcal{S}(\mathbb{N}, w+1, n-w)$, for all $w \in [0,n]$. Since the $d_{0\text{-}D/I}$ distance between binary words of distinct weight is ∞, the insertion/deletion of 0's code design problem is reduced to the proper design of the L_1 metric constant weight error control codes $\mathcal{A}_w$, for all $w \in [0,n]$.

Thus, in general, any L_1 distance error control property of codes over $\mathbb{N}$ reflects into the analogous $d_{0\text{-}D/I}$ distance error control property of codes over $\mathbb{Z}_2$ because of Theorem 2.1. So, from the L_1 metric asymmetric/unidirectional coding theory [4], [6], [38], [40], [43] and Theorem 2.1, the following theorem holds which gives only some (maximal) error correction capabilities of t-Sy0EC codes. Following the classical asymmetric/unidirectional coding theory notation [4], in the theorem below, t-SyXC/d-SyXD/AUXD indicates the class of codes capable of correcting t symmetric errors, detecting d symmetric errors and, simultaneously detecting all unidirectional errors; where the errors are of type X defined as follows. If $X =$"0E" then the codes are in the binary sequences domain of the function V in (15) and the errors are 0-errors, if instead $X =$"E" then then the codes are in the integer sequences codomain of the function V and the errors are L_1 distance errors. Analogously, (t_-, t_+)-0EC indicates the class of codes capable of correcting t_- deletions of 0's and, simultaneously, t_+ insertions of 0's; and (t_-, t_+)-EC indicates the class of codes capable of correcting t_- negative errors and, simultaneously, t_+ positive errors in the L_1 metric [35], [36], [37], [38], [39], [40], [41].

Theorem 2.2 (Error Control Capabilities and Combinatorial Characterizations of t-Sy0EC): Let $t, t_-, t_+, \tau \in \mathbb{N}$ be any numbers such that $t_- + t_+ = t$ and $\tau \in [0, t]$. If

$$\mathcal{C} = \bigcup_{w \in [0,n]} \mathcal{C}_w \subseteq \mathbb{Z}_2^n$$

is a binary code of length $n \in \mathbb{N}$ and $\mathcal{C}_w \stackrel{\text{def}}{=} \mathcal{C} \cap \mathcal{S}(\mathbb{Z}_2, n, w)$, for all integer $w \in [0, n]$, then $V(\mathcal{C}_w) \subseteq \mathcal{S}(\mathbb{N}, w+1, n-w)$, for all $w \in [0, n]$; and the following statements are equivalent:

1) $\mathcal{C}$ is a t-Sy0EC code (i. e., $\mathcal{C}$ is a t-Symmetric 0-error Correcting Code);
2) $\mathcal{C}$ is a $(t = t_-, 0)$-0EC code (i. e., $\mathcal{C}$ is a t deletion of 0's error correcting code);
3) $\mathcal{C}$ is a $(0, t = t_+)$-0EC code (i. e., $\mathcal{C}$ is a t insertion of 0's error correcting code);
4) $d_{0\text{-}D/I}(\mathcal{C}) > 2t$;
5) for all $w \in [0, n]$, $d_{0\text{-}D/I}(\mathcal{C}_w) > 2t$ ($\iff$ $\mathcal{C}_w$ is a t-Sy0EC);
6) for all $w \in [0, n]$, $d^{sy}_{L_1}(V(\mathcal{C}_w)) \geq 2(t+1)$ ($\iff$ $V(\mathcal{C}_w)$ is a τ-SyEC/$(2t - \tau + 1)$-SyED/AUED code over $\mathbb{N}$);
7) for all $w \in [0, n]$, $d^{as}_{L_1}(\hat{V}(\mathcal{C}_w)) \geq t + 1$ ($\iff$ $\hat{V}(\mathcal{C}_w)$ is a (t_-, t_+)-EC code over $\mathbb{N}$);
9) $d_{0\text{-}D/I}(\mathcal{C}) > 2t + 1$;
10) $\mathcal{C}$ is a τ-Sy0EC/$(2t - \tau + 1)$-Sy0ED/AU0ED code.

Proof: The equivalences among 1), 2), 3) and 4) come from [21]. The equivalence between 4) and 5) comes from Theorem 2.1 or [21] because if $w(X) \neq w(Y)$ then $d_{0\text{-}D/I}(X, Y) = \infty$. The equivalence between 5) and 6) comes because of Theorem 2.1, because $\mathcal{A}_w \stackrel{\text{def}}{=} V(\mathcal{C}_w)$ is a constant weigh $n - w$ code of length $w + 1$ over $\mathbb{N}$ and because

$$d^{sy}_{L_1}(\mathcal{A}) > 2t \iff d^{sy}_{L_1}(\mathcal{A}) \geq 2(t+1) \tag{20}$$

is valid or any constant weight code $\mathcal{A}$ over $\mathbb{N}$. Following the word/multiset notation at the beginning of this Section, relation (20) holds true because, for all $U_1, U_2 \in \mathbb{N}^n$,

$$\begin{aligned} |U_1| = |U_1 \cap U_2| + |U_1 \dot{-} U_2| = \\ |U_1 \cap U_2| + |U_2 \dot{-} U_1| = |U_2| \iff \\ |U_1 \dot{-} U_2| = |U_2 \dot{-} U_1|; \end{aligned} \tag{21}$$

and $d^{sy}_{L_1}(U_1, U_2) = |U_1 \dot{-} U_2| + |U_2 \dot{-} U_1|$. Note that the combinatorial characterization for any code, $\mathcal{A} \subseteq \mathbb{N}^n$ to be τ-SyEC/$(2t - \tau + 1)$-SyED/AUED is that for all $U_1, U_2 \in \mathcal{A}$, $|U_1 \dot{-} U_2| \geq t + 1$ (see, for example [4], [6], [43]). So, the combinatorial characterization for any constant weight w code, $\mathcal{A} \subseteq \mathcal{S}(\mathbb{N}, n, w)$ to be τ-SyEC/$(2t - \tau + 1)$-SyED/AUED is $d^{sy}_{L_1}(\mathcal{A}) \geq 2(t+1)$ because for any two codewords U_1 and U_2 in a constant weight code, $|U_1 \dot{-} U_2| = |U_2 \dot{-} U_1|$ (from (21)). Analogously, the equivalence between 6) and 7) comes from the distance definitions as follows. Let $\hat{\mathcal{A}}$ be the code obtained by shortening the code $\mathcal{A}$ in one, say the last, component. From (5), it is readily seen that if $\mathcal{A} \subseteq \mathcal{S}(\mathbb{N}, n, w)$ then

$$d^{sy}_{L_1}(\mathcal{A}) \geq 2(t+1) \iff d^{as}_{L_1}(\mathcal{A}) \geq t + 1.$$

So, this equivalence follows because $\hat{\mathcal{A}_w} \stackrel{\text{def}}{=} \hat{V}(\mathcal{C}_w)$ is obtained by shortening the constant weight code $\mathcal{A}_w \stackrel{\text{def}}{=} V(\mathcal{C}_w))$ in the last (i. e., the $(w+1)$-th) component. Note that the combinatorial characterization for any code, $\mathcal{A} \subseteq \mathbb{N}^n$ to be (t_-, t_+)-EC code is that $d^{as}_{L_1}(\mathcal{A}) \geq t_- - t_+ + 1 = t + 1$ (see, for example, Theorem 1 in [40] with $d_- = d_+ = 0$). The relation in 6) is equivalent to the relation in 9) because, from Theorem 2.1, if $w(X) \neq w(Y)$ then $d_{0\text{-}D/I}(X, Y) = \infty$. The relation in 6) implies 10) because the τ-Sy0EC/$(2t - \tau + 1)$-Sy0ED/AU0ED decoding algorithm design problem for $\mathcal{C}$ can be reduced to the τ-SyEC/$(2t - \tau + 1)$-SyED/AUED decoding algorithms for $V(\mathcal{C}_w)$, for all $w \in [0, n]$; it is a matter, for the receiver to compute the number, w, of received 1's (recall that 1-errors are forbidden) of the received word, $R \in \mathbb{Z}_2^*$ and apply **any** τ-SyEC/$(2t - \tau + 1)$-SyED/AUED decoding algorithm for $\mathcal{A} \stackrel{\text{def}}{=} \mathcal{A}_w = V(\mathcal{C}_w) \subseteq \mathcal{S}(\mathbb{N}, w+1, n-w)$ with input word $V(R) \in \mathbb{N}^{w+1}$. In fact, this is the decoding strategy described in Sub-section III-B for $\tau = t$. On the other hand, clearly 10) implies 1) and, so, 10) implies 6) because of the above. ∎

III. Non Systematic Code Design

A. σ-Code Based Design

Our code design is based on the L_1 metric error control σ-codes over $\mathbb{Z}_m$ defined in [38] and [40]. The σ-code theory is based on the sigma polynomials of a word defined below. Let $m \in \mathbb{N} \cup \{\infty\}$, $\mathbb{F}$ be **any** field and $\partial S \subseteq \mathbb{F}$ be a set of $n \in \mathbb{N}$ distinct elements in $\mathbb{F}$ used as index set. The σ-polynomial associated with a word $X \in \mathbb{Z}_m^n$ is defined as [38],

$$\begin{aligned} \sigma_X(z) \stackrel{\text{def}}{=} z^{x_0} \prod_{a \in \partial S - \{0\}} (1 - az)^{x_a} = \\ \sigma_0(X) + \sigma_1(X) z + \sigma_2(X) z^2 + \ldots \in \mathbb{F}[z]. \end{aligned} \tag{22}$$

For example, if $n=8$, $\partial S=\{0,a_1,a_2,a_3,a_4,a_5,a_6,a_7\}\subseteq\mathbb{F}$ and $X=23021000=\{0,0,a_1,a_1,a_1,a_3,a_3,a_4\}$ then

$$\begin{aligned}\sigma_X(z)=&z^2(1-a_1z)^3(1-a_3z)^2(1-a_4z)=\\&1\cdot z^2-(3a_1+2a_3+a_4)z^3+\\&(3a_1^2+6a_1a_3+3a_1a_4+a_3^2+2a_3a_4)z^4+\ldots\\&-(a_1^3a_3^2a_4)z^9.\end{aligned}$$

Note that $\sigma_X(z)$ is a polynomial of degree $\deg(\sigma_X)=w_{L_1}(X)=|X|$ having $w_H(X)=|\partial X|$ distinct roots in $\mathbb{F}$, each with multiplicity x_a, for $a\in\partial S\subseteq\mathbb{F}$. In particular, X coincides with the multiset of all the inverses of the roots of $\sigma_X(z)$, where we let $1/0\stackrel{\text{def}}{=}0$. Hence, its coefficient sequence is given by the elementary symmetric functions, 1, $\sigma_1(X-\{0\})$, $\sigma_2(X-\{0\})$, $\ldots\in\mathbb{F}$, of the elements in the multiset $X-\{0\}$ ordered in increasing order of their degrees, and eventually right shifted by $x_0\in\mathbb{Z}_m\subseteq\mathbb{N}$ if $0\in\partial S\subseteq\mathbb{F}$. At this point, the general definition of σ-code is the following. For all polynomials $g(z),\sigma(z)\in\mathbb{F}[z]$, the m-ary σ-code of length n associated with g and σ is defined as

$$\mathcal{C}_{g,\sigma}\stackrel{\text{def}}{=}\mathcal{C}_{g,\sigma}(\mathbb{Z}_m,n)\stackrel{\text{def}}{=}\left\{X\in\mathbb{Z}_m^n\;\middle|\;\begin{array}{l}\sigma_X(z)=c_X\sigma(z)\bmod g(z),\\ \text{with } c_X\in\mathbb{F}-\{0\}\end{array}\right\}.\tag{23}$$

From the σ-code theory in [35], [36], [37], [38], [39], [40], and [41], the following relation holds (for example, see Theorem 3.2 in [36] or Theorem 5 in [38]).

$$\gcd(\sigma,g)=1\;\;\Rightarrow\;\;d_{L_1}^{as}(\mathcal{C}_{g,\sigma})\geq\deg(g).\tag{24}$$

At this point, the code design idea is as follows. For simplicity, we choose $g(z)=z^{t+1},\sigma(z)=1+\hat{\sigma}(z)\in K[z]$ and $\partial S\subseteq\mathbb{F}-\{0\}$. In this way, $\gcd(\sigma,g)=1$, $d_{L_1}^{as}(\mathcal{C}_{g,\sigma})\geq t+1$ because of (24), and the non empty σ-codes (23) simplify as follows because $\sigma_0(X)=1\in K$:

$$\mathcal{C}_{z^{t+1},\sigma}\stackrel{\text{def}}{=}\mathcal{C}_{z^{t+1},\sigma}(\mathbb{N},n)\stackrel{\text{def}}{=}\{X\in\mathbb{N}^n\mid\sigma_X(z)=\sigma(z)\bmod z^{t+1}\}.\tag{25}$$

In this case, to define a t-Sy0EC code $\mathcal{C}\subseteq\mathbb{Z}_2^n$, the σ-codes in (25) are used in the function $\hat{V}$ codomain where $\hat{V}$ is given in (16). So, $X\in\mathcal{C}$ if, and only if $\sigma_{\hat{V}(X)}(z)=\sigma(z)\bmod z^{t+1}$, where $\sigma(z)$ is a monic polynomial of degree t. Note that under the mapping $X\to\sigma_{\hat{V}(X)}(z)\bmod z^{t+1}$, the set of constant weight w vectors of length n over $\mathbb{Z}_2$ (and in fact, the set $\mathcal{S}(\mathbb{N},w+1,n-w)$) is partitioned into $|\mathbb{F}|^t$ classes, $\mathcal{D}_1,\mathcal{D}_2,\ldots,\mathcal{D}_{|\mathbb{F}|^t}$, where, X and Y are in $\mathcal{D}_i$ if, and only if, $\sigma_{\hat{V}(X)}(z)=\sigma_{\hat{V}(Y)}(z)\bmod z^{t+1}$. Now, we prove that each of the $\hat{V}(\mathcal{D}_i)$'s is an asymmetric L_1 distance $t+1$ code. Suppose $X,Y\in\mathcal{D}_i$, let $\hat{V}\stackrel{\text{def}}{=}\hat{V}(X)$ and $\hat{U}\stackrel{\text{def}}{=}\hat{V}(Y)$. Then, $\sigma_{\hat{V}}(z)=\sigma_{\hat{U}}(z)\bmod z^{t+1}$ and this implies

$$\sigma_{\hat{V}\dot{-}\hat{U}}(z)=\sigma_{\hat{U}\dot{-}\hat{V}}(z)\bmod z^{t+1}\tag{26}$$

because

$$\text{for all } A,B\in\mathbb{N}^n,\;\sigma_A(z)\sigma_{B\dot{-}A}(z)=\sigma_B(z)\sigma_{A\dot{-}B}(z),\tag{27}$$

and $\gcd\left(\sigma_{\partial S}(z),g(z)=z^{t+1}\right)=1$. Now, if the asymmetric L_1 distance between $\hat{V}$ and $\hat{U}$ is $s<t+1$ then the degrees of $\sigma_{\hat{V}\dot{-}\hat{U}}(z)$ and $\sigma_{\hat{U}\dot{-}\hat{V}}(z)$ are $s<t+1$ and thus, $\sigma_{\hat{V}\dot{-}\hat{U}}(z)=\sigma_{\hat{U}\dot{-}\hat{V}}(z)$ because of (26). This means, $\sigma_{\hat{V}\dot{-}\hat{U}}(z)$ has $2s$ roots (i. e., the s roots of $\sigma_{\hat{V}\dot{-}\hat{U}}(z)$ and the s roots of $\sigma_{\hat{U}\dot{-}\hat{V}}(z)$), which gives a contradiction. Therefore, the minimum asymmetric L_1 distance of the code $\hat{V}(\mathcal{D}_i)$ is at least $t+1$; for all $i\in[1,|\mathbb{F}|^t]$. So, under the mapping $X\to\sigma_{\hat{V}(X)}(z)\bmod z^{t+1}$, the set $\mathcal{S}(\mathbb{Z}_2,n,w)$ is partitioned into the $|\mathbb{F}|^t$ classes $\mathcal{D}_i$'s. Thus, by pigeon-hole principle, one of the classes, say $\tilde{\mathcal{D}}(\mathbb{F};n,w)$ should have at least $\binom{n}{w}/|\mathbb{F}|^t$ codewords and minimum 0-error distance $2(t+1)$ because 6) is equivalent to 7) in Theorem 2.2. So, from Theorem 2.2, the t-Sy0EC code, $\mathcal{C}$, can be simply defined by letting for all $w\in[0,w]$, $\mathcal{C}_w\stackrel{\text{def}}{=}\tilde{\mathcal{D}}(\mathbb{F};n,w)\subseteq\mathcal{S}(\mathbb{Z}_2,n,w)$, where to maximize $|\mathcal{C}|$, the algebraic structure $\mathbb{F}$ is chosen to be the smallest possible field if $t>1$ or the smallest group if $t=1$. In this way, the number of codewords is

$$|\mathcal{C}|\geq\sum_{w=0}^{n}\left\lceil\binom{n}{w}\Big/|\mathbb{F}_w|^t\right\rceil.\tag{28}$$

where $\mathbb{F}_w$ is the smallest field, $\mathbb{F}$, whose cardinality is $|\mathbb{F}|>w$, when $t>1$ and $\mathbb{F}_w=(\mathbb{Z}_{w+1},+\bmod(w+1))$ when $t=1$. Note that if $t=1$ then the simpler group-theoretic code construction for single asymmetric error correcting codes [18] can be used; in this way, $|\mathbb{F}_w|=w+1$ and $|\mathcal{C}|\geq(2^{n+1}-1)/(n+1)$ [11], [21]. In Subsection IV-A, the lower bound in (28) is refined as in (62); where, recall that, $D(n,t)$ is the largest cardinality t-Sy0EC/$(t+1)$-Sy0ED/AU0ED binary code of length n. Table II shows a non-systematic code obtained with the construction defined by the lower bound in (28) for $n=7$ and $t=2$. Table III shows an interesting improvement of the code in Table II which is possibly optimal and where known asymptotic lower bounds fail.

We want to mention that if $t>1$ then the lower bound given here improves the lower bound given in [11] for two major reasons. First, unlike the analysis in [11], our analysis shows no restrictions on the design parameters, but the restriction given by the definition of the integer sequence $\{|\mathbb{F}_w|:w\in\mathbb{N}\}$ alone, which, as a set, is equal to the prime power sequence (given below in increasing order of its elements)

$$\mathbb{PP}\stackrel{\text{def}}{=}\{2,3,\mathbf{4},5,7,\mathbf{8},\mathbf{9},11,13,\mathbf{16},17,\ldots\}\stackrel{\text{def}}{=}\{q_1,q_2,q_3,\ldots\}.\tag{29}$$

Second, the prime sequence

$$\mathbb{P}\stackrel{\text{def}}{=}\{2,3,5,7,11,13,17,\ldots\}$$

on which the analysis in [11] is based, contains bigger gaps between consecutive elements than the prime power sequence because $\mathbb{P}\subsetneq\mathbb{PP}$. In Table IV, the lower bound on the number of codewords given by our proposed code is compared with those of [11]. In all cases, the proposed codes give either more number of codewords or the same number of codewords.

However, other choices of $g(z)$ and ∂S give better lower bounds; and these are discussed in Subsection IV-A. The considerations and code designs are exactly the same except that the relations are taken $\bmod g(z)$ instead of $\bmod z^{t+1}$.

TABLE II

NON-SYSTEMATIC CODE PARAMETERS WITH $n = 7$ AND $t = 2$. HERE, THE LOWER BOUND IN (28) GIVES 14 BUT THE ACTUAL CODE DEFINING THE LOWER BOUND IN (28) HAS $|\mathcal{C}| = 16$ CODEWORDS. THE LOWER BOUND IN (62) GIVES 15 AND THE LOWER BOUND IN SECTION 4 OF [11] (SEE TABLE IV) GIVES 13. ALSO, THE UPPER BOUND VALUE OBTAINED WITH (67) IS 29. IN (62) AND (67), THE FUNCTION $D(n,t)$ IS THE LARGEST CARDINALITY t-SY0EC/$(t+1)$-SY0ED/AU0ED BINARY CODE OF LENGTH n

$l(X)=n$	$w(X)$	X	$V(X)=\hat{V}(X)\boldsymbol{v}_{\boldsymbol{w}+1}$	$l(V(X))=w(X)+1$	$w(V(X))$	$\mathbb{F}_w$	$\sigma_{\hat{V}(X)}(z) \bmod z^3$	$\lvert\mathcal{C}\rvert=16>14=$ LB in (28)
7	0	0000000	**7**	1	7	$GF(2)$	1	$1 \geq \lceil 1/2^2 \rceil = 1$
7	1	1000000	0**6**	2	6	$GF(2)$	1	$2 \geq \lceil 7/2^2 \rceil = 2$
		0000100	4**2**					
7	2	0101000	11**3**	3	5	$GF(3)$	$1+2z^2$	$3 \geq \lceil 21/3^2 \rceil = 3$
		0100001	14**0**					
		0000101	41**0**					
7	3	1110000	000**4**	4	4	$GF\left(2^2\right)$	1	$5 \geq \lceil 35/4^2 \rceil = 3$
		0101010	111**1**					
		1100001	004**0**					
		1000011	040**0**					
		0000111	400**0**					
7	4	1110100	0001**2**	5	3	$GF(5)$	$1+z$	$2 \geq \lceil 35/5^2 \rceil = 2$
		0010111	2100**0**					
7	5	1111100	00000**2**	6	2	$GF(7)$	1	$1 \geq \lceil 21/7^2 \rceil = 1$
7	6	1111110	0000001	7	1	$GF(7)$	1	$1 \geq \lceil 7/7^2 \rceil = 1$
7	7	1111111	0000000	8	0	$GF\left(2^3\right)$	1	$1 \geq \lceil 1/8^2 \rceil = 1$

TABLE III

IMPROVED CODE DESIGN PARAMETERS FOR $n = 7$ AND $t = 2$. HERE, WE GIVE A CODE WITH CARDINALITY $|\mathcal{C}'| = 21 > 16 = |\mathcal{C}|$, WHICH IS BIGGER THAN THE CODE $\mathcal{C}$ IN TABLE II. WE CONJECTURE $\mathcal{C}'$ TO BE OPTIMAL. NOTE THAT, FOR THE VALUES $n = 7$ AND $t = 2$, THE ASYMPTOTIC LEVENSHTEIN UPPER BOUND (3) IN [21] AND THE ASYMPTOTIC UPPER BOUND (22) IN [16] BOTH FAIL AND GIVE $20 < 21$

$l(X)=n$	$w(X)$	X	$V(X)=\hat{V}(X)\boldsymbol{v}_{\boldsymbol{w}+1}$	$l(V(X))=w(X)+1$	$w(V(X))$	$\lvert\mathcal{C}'\rvert=21$
7	0	0000000	**7**	1	7	1
7	1	1000000	0**6**	2	6	3
		0001000	3**3**			
		0000001	6**0**			
7	2	1100000	00**5**	3	5	4
		0100100	12**2**			
		1000001	05**0**			
		0000011	50**0**			
7	3	1110000	000**4**	4	4	5
		0101010	111**1**			
		1100001	004**0**			
		1000011	040**0**			
		0000111	400**0**			
7	4	1111000	0000**3**	5	3	5
		1110001	0003**0**			
		1100011	0030**0**			
		1000111	0300**0**			
		0001111	3000**0**			
7	5	1111100	00000**2**	6	2	1
7	6	1111110	0000001	7	1	1
7	7	1111111	0000000	8	0	1

B. t-Sy0EC/$(t+1)$-Sy0ED/AU0ED Decoding Algorithm for t-Sy0EC Codes

Let $\mathcal{C} = \bigcup_{w\in[0,n]} \mathcal{C}_w$ be a t-Sy0EC code of length n, where $\mathcal{C}_w \stackrel{\text{def}}{=} \mathcal{C} \cap \mathcal{S}(\mathbf{Z}\!\mathbf{Z}_2, n, w)$. From Theorem 2.2 with $\tau = t$, $\mathcal{C}$ is actually a t-Sy0EC/$(t+1)$-Sy0ED/AU0ED code and here we give an efficient t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding algorithm for $\mathcal{C}$ exploiting, in this way, its maximum error control capabilities. Such algorithm is as follows. If $C \in \mathcal{C} \subseteq \mathbf{Z}\!\mathbf{Z}_2^n$ is sent and $R \in \mathbf{Z}\!\mathbf{Z}_2^*$ is received, the decoder computes $w \stackrel{\text{def}}{=} w(R) \in [0,n]$ and applies Algorithm 3.1 below with input 1) the constant L_1 weight $\omega \stackrel{\text{def}}{=} n - w$ code of length $\nu \stackrel{\text{def}}{=} w+1$ over the alphabet $\mathbb{N}$,

$$\mathcal{A} \stackrel{\text{def}}{=} \mathcal{A}(\mathbb{N}, \nu, \omega) \stackrel{\text{def}}{=} V(\mathcal{C}_w) \subseteq \mathcal{S}(\mathbb{N}, w+1, n-w) \quad (30)$$

and 2) the word $Y = V(R) \in \mathbb{N}^\nu$. On getting as output the word $X' \in \mathbb{N}^\nu$ the decoder computes $C' \stackrel{\text{def}}{=} V^{-1}(X')$

TABLE IV

A COMPARISON ON THE NUMBER OF CODEWORDS BETWEEN THE PROPOSED METHOD TO THAT IN SECTION 4. OF [11] (IN PARENTHESES). ALL THE VALUES ARE OBTAINED FROM (28); BUT THOSE FROM [11] ARE OBTAINED WITH THE RESTRICTION THAT $|\mathbb{F}_w|$ IS A PRIME NUMBER (I. E., $|\mathbb{F}_w| \in \mathbb{P}$). THE CODE LENGTH IS INDICATED WITH n AND ERROR CORRECTING CAPABILITY WITH t

$n\backslash t$	2	3	4	5	6	7	8
1	2(2)	2(2)	2(2)	2(2)	2(2)	2(2)	2(2)
2	3(3)	3(3)	3(3)	3(3)	3(3)	3(3)	3(3)
3	4(4)	4(4)	4(4)	4(4)	4(4)	4(4)	4(4)
4	5(5)	5(5)	5(5)	5(5)	5(5)	5(5)	5(5)
5	8(8)	6(6)	6(6)	6(6)	6(6)	6(6)	6(6)
6	10(9)	7(7)	7(7)	7(7)	7(7)	7(7)	7(7)
7	14(13)	8(8)	8(8)	8(8)	8(8)	8(8)	8(8)
8	19(18)	10(10)	9(9)	9(9)	9(9)	9(9)	9(9)
9	28(26)	14(13)	10(10)	10(10)	10(10)	10(10)	10(10)
10	42(38)	15(14)	11(11)	11(11)	11(11)	11(11)	11(11)
11	68(60)	21(20)	12(12)	12(12)	12(12)	12(12)	12(12)
12	107(94)	27(24)	13(13)	13(13)	13(13)	13(13)	13(13)
13	179(156)	38(33)	16(15)	14(14)	14(14)	14(14)	14(14)
14	307(261)	56(47)	19(18)	15(15)	15(15)	15(15)	15(15)
15	529(446)	84(68)	24(22)	16(16)	16(16)	16(16)	16(16)
16	924(775)	133(107)	29(24)	17(17)	17(17)	17(17)	17(17)
17	1635(1376)	217(174)	41(35)	18(18)	18(18)	18(18)	18(18)
18	2906(2469)	354(280)	57(46)	20(20)	19(19)	19(19)	19(19)
19	5206(4485)	589(472)	85(68)	24(22)	20(20)	20(20)	20(20)
20	9367(8198)	993(809)	127(99)	31(26)	21(21)	21(21)	21(21)
21	16940(15057)	1693(1401)	198(155)	37(31)	22(22)	22(22)	22(22)
22	30775(27739)	2911(2456)	309(245)	50(41)	23(23)	23(23)	23(23)
23	56173(51233)	5040(4329)	502(405)	70(57)	24(24)	24(24)	24(24)
24	102988(94823)	8799(7680)	822(676)	99(80)	29(27)	25(25)	25(25)
25	189666(175897)	15467(13677)	1365(1142)	148(119)	33(30)	26(26)	26(26)
26	350742(327071)	27361(24452)	2289(1944)	222(180)	38(34)	27(27)	27(27)
27	651081(609761)	48698(43894)	3885(3346)	350(288)	52(43)	28(28)	28(28)
28	1212671(1139852)	87153(79105)	6641(5786)	558(466)	69(58)	29(29)	29(29)
29	2265130(2136335)	156748(143113)	11452(10073)	907(766)	97(82)	31(31)	30(30)
30	4241040(4013493)	283106(259870)	19871(17627)	1497(1278)	141(119)	34(33)	31(31)
31	7956075(7555602)	513158(473476)	34689(31005)	2499(2155)	213(179)	41(38)	32(32)
32	14949764(14247760)	932969(865220)	60863(54779)	4204(3659)	328(277)	51(46)	33(33)
33	28131692(26902762)	1700592(1584988)	107245(97172)	7126(6256)	524(445)	64(56)	34(34)
34	53009586(50848760)	3106921(2909318)	189659(172941)	12152(10757)	844(725)	84(74)	35(35)
35	100029016(96180476)	5688586(5348587)	336508(308630)	20832(18583)	1378(1194)	120(104)	37(37)
36	189041792(182028906)	10438266(9845096)	598893(552020)	35884(32233)	2280(1992)	177(154)	41(40)
37	357858744(344666584)	19197637(18139555)	1069087(989121)	62073(56081)	3802(3343)	268(231)	47(45)
38	678652915(652912787)	35394109(33450055)	1914333(1774962)	107802(97826)	6375(5645)	424(367)	57(53)
39	1289471483(1237458799)	65425462(61732235)	3438899(3189221)	187963(171007)	10755(9571)	672(585)	69(65)
40	2454854787(2346795440)	121268674(114023746)	6198496(5737240)	329053(299485)	18244(16286)	1087(952)	95(86)
41	4682570390(4454101119)	225404144(210817704)	11211673(10333798)	578437(525380)	31110(27789)	1779(1559)	135(121)
42	8948582405(8461990103)	420124150(390246164)	20351513(18638792)	1021132(923224)	53318(47541)	2936(2578)	195(173)
43	17130848687(16095622815)	785148313(723442191)	37072583(33672617)	1810358(1625288)	91848(81517)	4892(4287)	304(265)
44	32846089219(30659137728)	1470985850(1343471455)	67762062(60949036)	3223173(2867089)	159014(140106)	8204(7156)	472(410)
45	63064193961(58494107149)	2762154375(2499987313)	124253656(110569059)	5762117(5069592)	276661(241408)	13846(11989)	755(653)
46	121223228302(111796305517)	5197012977(4662765181)	228508108(201105822)	10340963(8988381)	483669(417129)	23522(20172)	1226(1050)

as the estimate of the sent codeword C. The output signal $cor \in \{0,1\}$ is such that if $cor = 1$ then 0-errors are corrected.

The following Algorithm 3.1 is a general efficient error control algorithm for any m-ary constant weight w code, $\mathcal{A}$, of length n with minimum L_1 distance $d^{sy}_{L_1}(\mathcal{A}) \geq 2(t+1)$. Note that Algorithm 3.1 efficiently reduces the t-SyEC/$(t+1)$-SyED/AUED decoding design problem for constant weight codes to the less powerful (τ_-, τ_+)-EC decoding design problem; proving that the two problems are indeed equivalent.

Algorithm 3.1 (General t-SyEC/$(t+1)$-SyED/AUED Decoding Algorithm for Constant Weight Codes):

Input:

1) The constant weight code $\mathcal{A} \stackrel{\text{def}}{=} \hat{\mathcal{A}}\, x_n \subseteq \mathcal{S}(\mathbf{Z}\!\mathbf{Z}_m, n, w)$, where

$$x_n \stackrel{\text{def}}{=} w - w_{L_1}(\hat{X}) \in \mathbf{Z}\!\mathbf{Z}_m, \quad \hat{X} \in \hat{\mathcal{A}},$$

is the parity digit; together with a set, $\mathcal{D}ec(\hat{\mathcal{A}})$, of **any** (possibly efficient) (τ_-, τ_+)-EC decoding algorithms, say $\mathcal{D}ec(\hat{\mathcal{A}}, \tau_-, \tau_+)$, for the shortened code $\hat{\mathcal{A}}$, for all $\tau_-, \tau_+ \in \mathbb{N}$ such that $\tau_- + \tau_+ = t < d^{as}_{L_1}(\hat{\mathcal{A}})$; and,

2) the (received) word $Y = \hat{Y}\, y_n \in \mathbb{Z}_m^n$ with $\hat{Y} \in \mathbb{Z}_m^{n-1}$ and $y_n \in \mathbb{Z}_m$.

Output:

1) A word $X' = \hat{X}'\, x'_n \in \mathbb{Z}_m^n$, where $\hat{X}' \in \mathbb{Z}_m^{n-1}$ and $x'_n \in \mathbb{Z}_m$ (the word X' represents the estimate of the sent codeword $X \stackrel{\text{def}}{=} \hat{X}\, x_n \in \mathcal{A}$); and,
2) a signal $cor \in \{0,1\}$ such that if $cor = 1$ then errors are corrected; i. e., $X' = X$.

Execute the following steps.

S1: Compute

$$\Delta(X,Y) \stackrel{\text{def}}{=} |Y| - w = |Y \dot{-} X| - |X \dot{-} Y|. \tag{31}$$

S2: If $|\Delta(X,Y)| \geq t+1$ then set $cor = 0$, set X' to be any word, output cor, output X' and **exit**.

S3: Otherwise, if $|\Delta(X,Y)| \leq t$ then execute the following steps.

S3.1: Compute

$$\tau_- \stackrel{\text{def}}{=} \left\lfloor \frac{t - \Delta(X,Y)}{2} \right\rfloor \text{ and } \tau_+ \stackrel{\text{def}}{=} \left\lfloor \frac{t + \Delta(X,Y)}{2} \right\rfloor. \tag{32}$$

Note that $0 \leq \tau_-, \tau_+ \leq t$ (because $|\Delta(X,Y)| \leq t$) and

$$\tau_- + \tau_+ \leq \frac{t - \Delta(X,Y)}{2} + \frac{t + \Delta(X,Y)}{2} = t. \tag{33}$$

S3.2: With the word $\hat{Y} \in \mathbb{Z}_m^{n-1}$ as input, execute the algorithm $\mathcal{D}ec(\hat{\mathcal{A}}, \tau_-, t - \tau_-)$ for $\hat{\mathcal{A}}$. Let $\hat{X}' \in \mathbb{Z}_m^{n-1}$ be its output word.

S3.3: Set $X' = \hat{X}'\, x'_n \in \mathcal{A}$ if $\hat{X}' \in \hat{\mathcal{A}}$, and $X' =$ any word if $\hat{X}' \notin \hat{\mathcal{A}}$; where

$$x'_n = w - w_{L_1}(\hat{X}') \tag{34}$$

is the parity digit of $\hat{X}'$.

S3.4: Set

$$cor = \begin{cases} 1 \text{if } X' \in \mathcal{A} \text{ and } d_{L_1}^{sy}(X', Y) \leq t, \\ 0 \text{otherwise} \end{cases} \tag{35}$$

S3.5: Output X', output cor and **exit**.

Theorem 3.1 (Correctness of Algorithm 3.1): Given $m \in \mathbb{IN} \cup \{\infty\}$ and $n, w, t \in \mathbb{IN}$, let $\mathcal{A}$ be any m-ary constant weight w code of length n with minimum L_1 distance

$$d_{L_1}^{sy}(\mathcal{A}) \geq 2t+2 \iff d_{L_1}^{as}(\hat{\mathcal{A}}) \geq t+1. \tag{36}$$

If for all (sent codeword) $X \in \mathcal{A}$ and (received word) $Y \in \mathbb{Z}_m^n$,

$$\begin{cases} \text{either} \delta(X,Y) \stackrel{\text{def}}{=} \min\{|Y \dot{-} X|, |X \dot{-} Y|\} = 0, \\ \text{or} \quad d_{L_1}^{sy}(X,Y) \leq t+1, \end{cases} \tag{37}$$

then Algorithm 3.1 gives the correct output as a t-SyEC/$(t+1)$-SyED/AUED decoding algorithm for $\mathcal{A}$; that is, by definition of t-SyEC/$(t+1)$-SyED/AUED decoding,

C1) if (37) holds and $cor = 1$ then $X' = X$; and,
C2) if $d_{L_1}^{sy}(X,Y) \leq t$ then $cor = 1$ (and hence, $X' = X$).

Proof: Let $X \in \mathcal{A}$, $Y \in \mathbb{Z}_m^n$ and assume (37) holds. First, let us prove that if $cor = 1$ then $X' = X$. Note that $cor = 1$ if, and only if, step S3.4 is executed and (35) evaluates to 1. In particular, if $cor = 1$ then $X \in \mathcal{A}$, $|\Delta(X,Y)| \leq t$, $X' \in \mathcal{A}$ and $d_{L_1}^{sy}(X',Y) \leq t$. And so, $X \in \mathcal{A}$, $d_{L_1}^{sy}(X,Y) \leq t+1$, $X' \in \mathcal{A}$ and $d_{L_1}^{sy}(X',Y) \leq t$ because of (37) and $d_{L_1}^{sy}(X,Y) = |\Delta(X,Y)| + 2\delta(X,Y)$. Hence, $X \in \mathcal{A}$, $X' \in \mathcal{A}$ and $d_{L_1}^{sy}(X,X') \leq d_{L_1}^{sy}(X,Y) + d_{L_1}^{sy}(Y,X') \leq t+1+t < 2t+2$. This implies $X' = X$ because $d_{L_1}^{sy}(\mathcal{A}) \geq 2(t+1)$. So, condition C1) of the theorem is satisfied. Now we prove that if $d_{L_1}^{sy}(X,Y) \leq t$ then $cor = 1$. First note that, from (5) and (31), the following relations hold for any $X \stackrel{\text{def}}{=} \hat{X}x_n, Y \stackrel{\text{def}}{=} \hat{Y}y_n \in \mathbb{IN}_m^{n-1} \times \mathbb{IN}_m$:

$$|\hat{X} \dot{-} \hat{Y}| \leq |X \dot{-} Y| = \frac{d_{L_1}^{sy}(X,Y) - \Delta(X,Y)}{2},$$
$$|\hat{Y} \dot{-} \hat{X}| \leq |Y \dot{-} X| = \frac{d_{L_1}^{sy}(X,Y) + \Delta(X,Y)}{2}. \tag{38}$$

Now, if $d_{L_1}^{sy}(X,Y) \leq t$ then $|\Delta(X,Y)| \leq d_{L_1}^{sy}(X,Y) \leq t$ and so, step S3 is executed. In this case, from the relations in (38), $d_{L_1}^{sy}(X,Y) \leq t$, (32) and (33), it follows,

$$|\hat{X} \dot{-} \hat{Y}| \leq \tau_-, \quad \text{and} \quad |\hat{Y} \dot{-} \hat{X}| \leq \tau_+ \leq t - \tau_-. \tag{39}$$

From the hypothesis (36), $d_{L_1}^{as}(\hat{\mathcal{A}}) \geq t+1$, and so, from (39), decoding algorithm $\mathcal{D}ec(\hat{\mathcal{A}}, \tau_-, t-\tau_-)$ will give the correct output in step S3.2. Hence, $\hat{X}' = \hat{X} \in \hat{\mathcal{A}}$, and so, from (34), $X' = X \in \mathcal{A}$ and $d_{L_1}^{sy}(X',Y) = d_{L_1}^{sy}(X,Y) \leq t$. This implies that $cor = 1$ is set in (35). In this way, also condition C2) of the theorem is satisfied. ∎

In the case of the σ-codes in (23), the efficient (τ_-, τ_+)-EC decoding algorithm, $\mathcal{D}ec(\hat{\mathcal{A}}, \tau_-, \tau_+)$, for the code $\hat{\mathcal{A}}$ is based on the key equation [35], [36], [37], [38], [39], [40], [41],

$$\sigma_X(z)\sigma_{Y \dot{-} X}(z) = \sigma_Y(z)\sigma_{X \dot{-} Y}(z), \text{ for all } X, Y \in \mathbb{Z}_m^n, \tag{40}$$

relating the σ-polynomials (22). Again, for simplicity, assume $\partial S \subseteq \mathbb{F} - \{0\}$ with $|\partial S| = n-1$, $g(z) = z^{t+1}$ with $\gcd\{z^{t+1}, \tilde{\sigma}(z)\} = 1$, so that

$$\hat{\mathcal{A}} \stackrel{\text{def}}{=} \mathcal{C}_{z^{t+1},\tilde{\sigma}}(\mathbb{Z}_m, n-1) = \left\{ \hat{X} \in \mathbb{Z}_m^{n-1} \;\middle|\; \begin{array}{l} \sigma_1(\hat{X}) = \tilde{\sigma}_1, \\ \sigma_2(\hat{X}) = \tilde{\sigma}_2, \ldots, \\ \sigma_t(\hat{X}) = \tilde{\sigma}_t \end{array} \right\} \tag{41}$$

and, hence,

$$\mathcal{A} \stackrel{\text{def}}{=} \mathcal{A}_{z^{t+1},\tilde{\sigma}}(\mathbb{Z}_m, n, w) \stackrel{\text{def}}{=} \left\{ X \in \mathbb{Z}_m^n \;\middle|\; \begin{array}{l} X = \hat{X}\, x_n \text{ with} \\ x_n = w - w_{L_1}(\hat{X}), \\ \sigma_1(\hat{X}) = \tilde{\sigma}_1, \\ \sigma_2(\hat{X}) = \tilde{\sigma}_2, \ldots, \\ \sigma_t(\hat{X}) = \tilde{\sigma}_t \end{array} \right\}. \tag{42}$$

If $X = \hat{X}\, x_n \in \mathcal{A} = \hat{\mathcal{A}}\, x_n$ is sent and $Y = \hat{Y}\, y_n \in \mathbb{IN}^n$ is received then, from (40),

$$\text{for all } \hat{X} \in \hat{\mathcal{A}} \subseteq \mathbb{Z}_m^{n-1} \text{ and } \hat{Y} \in \mathbb{IN}^{n-1},$$
$$\sigma_{\hat{Y} \dot{-} \hat{X}}(z) = [\sigma_{\hat{Y}}(z)/\sigma(z)]\sigma_{\hat{X} \dot{-} \hat{Y}}(z) \bmod z^{t+1} \tag{43}$$

where $\sigma_{\hat{Y} \dot{-} \hat{X}}(z)$ and $\sigma_{\hat{X} \dot{-} \hat{Y}}(z)$ are unknown and $[\sigma_{\hat{Y}}(z)/\sigma(z)]$ is known to the receiver. In this way, algorithm $\mathcal{D}ec(\hat{\mathcal{A}}, \tau_-, \tau_+)$ consists in solving the equation (43) with the constrains $\deg(\sigma_{\hat{Y} \dot{-} \hat{X}}) \leq t_+$ and $\deg(\sigma_{\hat{X} \dot{-} \hat{Y}}) \leq t_-$ required by (τ_-, τ_+)-EC decoding. This can be efficiently performed with the Extended Euclidean Algorithm. Note, however, that Algorithm 3.1 is of general type and can be efficiently applied to any constant weight code, $\mathcal{A}$, with minimum distance

$2t+2$ having efficient (τ_-,τ_+)-EC decoding algorithms for its shortened code, $\hat{\mathcal{A}}$. So, in general, efficiently decodable t-Sy0EC based σ-codes can be defined by choosing as $\hat{\mathcal{A}}$ the general codes in (23). In this case, the t-Sy0EC codes are, $\mathcal{C} =$

$$\bigcup_{w\in[0,n]} \left\{X\in\mathbb{Z}_2^n : V(X)\in\mathcal{A}_{g_{w,t},\tilde{\sigma}_{w,t}}(\mathbb{Z}_m, w+1, n-w)\right\};$$

where, for all $w\in[0,n]$, the field $\mathbb{F}_{w,t}$, the code support $\partial S_{w,t}\subseteq\mathbb{F}_w$ and the polynomial $g_{w,t}(z)\in\mathbb{F}_w[z]$ define a given triplet, $(\mathbb{F}_w, \partial S_w, g_w(z))$, of set $\Gamma(w,t)$ in (50); and $\tilde{\sigma}_{w,t}(z)\in\mathbb{F}_{w,t}[z]$ is a given polynomial such that $\gcd(\tilde{\sigma}_{w,t}, g_{w,t})=1$.

Note that, if the receiver knows the check information, $C \stackrel{\text{def}}{=} \sigma_{V(X)}(z) \bmod g(z)$, of any sent word $X\in\mathbb{Z}_2^n$ then it is capable of decoding the corresponding received word, $Y\in\mathbb{Z}_2^*$. Following and improving the fixed length recursive code design idea in [37], in [42], we have given systematic code designs whose strategy is to recursively send a $(t-1)$-Sy0EC encoding of C to the receiver; strangely enough, the $(t-1)$-Sy0EC capability is enough for the recursive t-Sy0EC design to be well defined. This comes from the combinatorial properties of the constant weight σ-codes. So, for these recursive codes, the challenging problem is to give a well defined t-Sy0EC/$(t+1)$-Sy0ED/AU0ED error control algorithm by keeping the redundancy below the optimal value of $t\log_2 k + o(t\log n)$ given by Theorem 4.1 of Section IV. In fact, [42] gives fixed length $n\in\mathbb{IN}$ systematic recursive σ-code based asymptotically optimal codes to efficiently encode k information bits. These codes have efficient t-Sy0EC/$(t+1)$-Sy0ED/AU0ED error control algorithms and redundancy $n-k\le t\log_2 k + o(t\log n)$ bits, for all $k,t\in\mathbb{IN}$.

IV. Bounds on the Cardinality of the Optimal t-Sy0EC Codes

For all $n,t\in\mathbb{IN}$, let $D(n,t)$ be the largest cardinality of a t-Sy0EC/$(t+1)$-Sy0ED/AU0ED binary code of length n. In this section we are mainly interested in finding lower and upper bounds on $D(n,t)$ which depend on n and t where both are considered as variables. In particular, we find bounds which hold true except at most a finite number of couples $(n,t)\in\mathbb{IN}^2$; that is, bounds which are asymptotic in, say, $s\stackrel{\text{def}}{=} n+t$. We are not aware of such a perspective in the literature, where t is always assumed to be a constant. Note that, a priori, for this coding problem, n could be fixed and t could go to infinity. In this setting, asymptotically optimal codes can be defined as follows.

Definition 4.1 (Asymptotically Optimal Codes): A family of t-Sy0EC/$(t+1)$-Sy0ED/AU0ED binary codes of length n, $\mathcal{C}(n,t)\subseteq\mathbb{Z}_2^n$, $n,t\in\mathbb{IN}$, is asymptotically optimal if, and only if, the ratio between the redundancy of $\mathcal{C}(n,t)$ and the optimal redundancy approaches 1 as $s\stackrel{\text{def}}{=} n+t$ grows large; i. e.,

$$\lim_{s\to\infty}\frac{n-\log_2|\mathcal{C}(n,t)|}{n-\log_2|D(n,t)|}=1.$$

Thanks to Theorem 2.2 and this perspective, many bounds can be reproved, improved and generalized. In particular, the following theorem will be shown in this section.

Theorem 4.1 (On the Optimal Redundancy): Let $n,t\stackrel{\text{def}}{=} t(n)\in\mathbb{IN}$. If

$$\log_2 t = o(\log n) \iff t = 2^{o(\log n)}$$

(for example, $t=2^{\sqrt{\log_2 n}}=2^{o(\log n)}$) then, the optimal redundancy of the t-Sy0EC/$(t+1)$-Sy0ED/AU0ED binary codes is

$$n-\log_2|D(n,t)| = t\log_2 n + o(t\log n). \tag{44}$$

So, any family of t-Sy0EC/$(t+1)$-Sy0ED/AU0ED binary codes whose redundancy is $t\log_2 n + o(t\log n)$ and $t=2^{o(\log n)}$ is asymptotically optimal according to Definition 4.1. For t constant, relation (44) was noticed in [25] for the sticky-insertion error correcting codes. Note, on the other hand, if $t\ge n-1$ then the optimal redundancy is

$$n-\log_2|D(n,t)| = n-\log_2(n+1),$$

as implied by Theorem 4.3 below. The upper bound side of (44) will be proved after Theorem 4.4 and the lower bound side of (44) will be proved after Theorem 4.7.

Now, a tight relation between $D(n,t)$ and the L_1 distance codes is given below in Theorem 4.2. Given any numeric alphabet $A\subseteq\mathbb{IN}$, $n,w\in\mathbb{IN}$, a **constant** weight w code of length n over the alphabet A is a block code of length n over A where every codeword weight is exactly w. A **constrained** weight w code of length n over the alphabet A is a block code of length n over A where every codeword weight belongs in the integer interval $[(w-\max_{a\in A} a), w]$. Furthermore, let

1) $CW(A,n,w,t)\in\mathbb{IN}$ be the largest cardinality of a **constant** weight w code of length n over the alphabet A with minimum symmetric L_1 distance greater than $2t$; and,
2) $LW(A,n,w,t)\in\mathbb{IN}$ be the largest cardinality of a **constrained** weight w code of length n over the alphabet A with minimum asymmetric L_1 distance greater than t.

Theorem 4.2: First, for any $n,w,t\in\mathbb{IN}$,

$$CW(\mathbb{Z}_m, n+1, w, t) = LW(\mathbb{Z}_m, n, w, t), \quad \text{for all } m\in\mathbb{IN}\cup\{\infty\}. \tag{45}$$

Then, **for all** $n,t\in\mathbb{IN}$,

$$D(n,t) = \sum_{w\in[0,n]} CW(\mathbb{IN}, w+1, n-w, t) = \sum_{w\in[0,n]} LW(\mathbb{IN}, w, n-w, t). \tag{46}$$

So, **any** lower or upper bound on the function $LW(\mathbb{IN},n,w,t)$ (or, equivalently, $CW(\mathbb{IN},n+1,w,t)$) gives bounds on $D(n,t)$.

Proof: First, relation (45) holds because if $\mathcal{A}\stackrel{\text{def}}{=}\hat{\mathcal{A}}v_{n+1}$ is any constant weight w code of length n over $\mathbb{Z}_m$ then the parity digit $v_{n+1}\in[w-(m-1),w]\cap\mathbb{IN}$ if, and only if, $w-v_{n+1}\in[0,m-1]\cap\mathbb{IN}=\mathbb{Z}_m$ and because of the minimum distance relation,

$$d_{L_1}^{sy}(\mathcal{A}) = 2d_{L_1}^{as}(\hat{\mathcal{A}}).$$

At this point, relation (46) is a direct consequence of the isometry V and Theorem 2.2. ∎

To derive the bounds on $D(n,t)$, first, let us use some easily computable values of $LW(\mathbb{IN}, w, n-w, t)$ in (46). Let

$$CW(\nu,\omega,t) \stackrel{\text{def}}{=} CW(\mathbb{IN},\nu,\omega,t)$$

and

$$LW(\nu,\omega,t) \stackrel{\text{def}}{=} LW(\mathbb{IN},\nu,\omega,t),$$

for all $\nu,\omega \in \mathbb{IN}$. We have

Theorem 4.3: For all, $n,t \in \mathbb{IN}$,

$$D(n,t) = \begin{cases} n+1 & \text{if } t \geq n-1, \\ 2+\left\lceil \dfrac{n}{t+1} \right\rceil + \displaystyle\sum_{w=2}^{n-t-1} LW(w,n-w,t)+t & \\ & \text{if } t < n-1. \end{cases} \tag{47}$$

In particular,

$$D(n,t) = \begin{cases} D(n,\infty) = n+1 \text{if } t \geq n-1, \\ n+2 \quad \text{if } t = n-2, \\ n+4 \quad \text{if } t = n-3. \end{cases} \tag{48}$$

Proof: From (46), we have

$$\begin{aligned} D(n,t) = {} & LW(0,n,t) + LW(1,n-1,t) + \\ & LW(2,n-2,t) + \ldots + LW(n-t-1,t+1,t) + \\ & LW(n-t,t,t) + LW(n-t+1,t-1,t) + \ldots + \\ & LW(n,0,t) = \\ & \sum_{\omega=0}^{n} LW(n-\omega,\omega,t). \end{aligned} \tag{49}$$

With regard to the $t+1$ terms in the third line of (49), if $\nu \in \mathbb{IN}$ is any and $\omega \in [0,t]$ then $LW(\nu,\omega,t)=1$. In fact, if $X,Y \in \mathbb{IN}^{\nu}$ and $0 \leq |X|,|Y| \leq t$ then

$$\begin{aligned} |X| &= |X \cap Y| + |X \dot{-} Y|, \\ |Y| &= |Y \cap X| + |Y \dot{-} X| \end{aligned}$$

and

$$\begin{aligned} d_{L_1}^{as}(X,Y) = {} & \max\{|X \dot{-} Y|, |Y \dot{-} X|\} = \\ & \max\{|X|,|Y|\} - |X \cap Y| \leq \max\{|X|,|Y|\} \leq t; \end{aligned}$$

That is, no minimum asymmetric distance $t+1$ constrained $\omega \in [0,t]$ code exists with 2 codewords; i. e., $1 \geq LW(\nu,\omega,t)$. On the other hand, any code with only one codeword has minimum asymmetric distance $t+1$; i. e., $LW(\nu,\omega,t) \geq 1$; and so, $LW(\nu,\omega,t) = 1$ for all $\omega \leq t$. With regard to the two terms in the first line of (49), note that obviously, $LW(0,n,t) = CW(\mathbb{IN},1,n,t) = 1$, whereas,

$$LW(1,n-1,t) = \left\lceil \frac{n}{t+1} \right\rceil.$$

The above equality comes because with one symbol in $\mathbb{Z}_n$, the code

$$\{0,(t+1),2(t+1),\ldots,(\lceil n/(t+1) \rceil - 1)(t+1)\}$$

is optimal (see also Theorem 2.4 in [8]). All this implies (47). With regard to (48), if $t \geq n-1$ then $D(n,t) = n+1$ because of (47). However, in this case, no two distinct codewords of any code can have the same weight so that the minimum $d_{0\text{-}D/I}$ distance of the code is ∞ (i. e., it can correct any number of 0-errors) and any optimal code contains exactly one codeword for any of the $n+1$ distinct weights. In this case, the optimal code is a zero error capacity code [8] with strictly positive information rate given by $\log_2(n+1)/n$. The remaining cases, $t = n-2$ and $n-3$, of (48) come from (47) and the equality $LW(2,n-2,n-3) = 3$. ∎

Given (47), in the following we assume $t+1 < n$ because, otherwise, $D(n,t) = n+1$ and it is completely determined. In this case, note that $s = n+t \to \infty$ if, and only if $n \to \infty$; so, asymptotic relations can be intended as the length n grows large. Now, the remaining terms $CW(\mathbb{IN}, w+1, n-w, t) = LW(w,n-w,t)$ of the sum in (46) will be bounded as described in Subsections IV-A and IV-B.

A. Lower Bounds and General σ-Code Based Design

With regard to the non-asymptotic lower bound, the code designs are based on the σ-code theory developed here in Section III and in [35], [36], [37], [38], [39], [40], and [41]. More precisely, for all $w,t \in \mathbb{IN}$, let

$$\Gamma(w,t) \stackrel{\text{def}}{=} \left\{ (\mathbb{F},\partial S,g) \;\middle|\; \begin{array}{l} \mathbb{F} \text{ is a finite field containing a} \\ \text{set } \partial S \subseteq \mathbb{F} \text{ with } |\partial S| = w, \\ g(z) \in \mathbb{F}[z] \text{ is monic with} \\ \deg(g) = t+1 \text{ and} \\ \partial S \cap \{\alpha \in \mathbb{F} : g(\alpha) = 0\} = \emptyset \end{array} \right\}, \tag{50}$$

Also, for all finite field $\mathbb{F}$ and monic polynomial $g(z) \in \mathbb{F}[z]$, let

$$\Phi(\mathbb{F},g) \stackrel{\text{def}}{=} \frac{\phi(g)}{|\mathbb{F}|-1}; \tag{51}$$

where $\phi(g)$ indicates the number of polynomials in $\mathbb{F}$ of degree less than $\deg(g)$ which are co-prime with $g(z)$. If $g(z) \in \mathbb{F}[z]$ is monic and has the following factorization in $\mathbb{F}[z]$,

$$g(z) = \prod_{i=1}^{h} [p_i(z)]^{m_i},$$

with $p_i(z) \in \mathbb{F}[z]$ distinct irreducible polynomials and $m_i \in \mathbb{IN}$, for all $i = 1,2,\ldots,h$, then $\phi(g)$ can be easily computed as [35], [36],

$$\phi(g) = \prod_{i=1}^{h} \left(|\mathbb{F}|^{\deg(p_i)} - 1 \right) |\mathbb{F}|^{(m_i-1)\deg(p_i)}.$$

By using the pigeon principle as in Subsection III-A and the results in [35] and [36], if $(\mathbb{F}_{w,t}, S_{w,t}, g_{w,t}) \in \Gamma(w,t)$, with $w \in \mathbb{IN}$, then

$$\left\lceil \binom{n}{w} \middle/ \Phi(\mathbb{F}_{w,t}, g_{w,t}) \right\rceil \leq LW(w,n-w,t);$$

for all $n \in \mathbb{IN}$. Hence, from (47), the following non-asymptotic lower bounds on $D(n,t)$ hold **for all** $n,t \in \mathbb{IN}$. If

$$(\mathbb{F}_{w,t}, S_{w,t}, g_{w,t}) \in \Gamma(w,t),$$

with $w\in\mathbb{IN}$, is any triplet sequence then

$$2+\left\lceil\frac{n}{t+1}\right\rceil+\sum_{w=2}^{n-t-1}\left\lceil\binom{n}{w}\Big/\Phi(\mathbb{F}_{w,t},g_{w,t})\right\rceil+t\le D(n,t). \tag{52}$$

Now, note that the triplets $(\mathbb{F}_{w,t},\partial S_{w,t},g_{w,t})\in\Gamma(w,t)$ which give the minimum value of $\Phi(\mathbb{F}_{w,t},g_{w,t})\in\mathbb{IN}$, will depend on the number of irreducible polynomials of any given degree and may be computed easily; for all $w\in\mathbb{IN}$. However, finding in general the minimizing triplet sequences is beyond the scope of this paper [10]; but, some relevant choices of these triplets can be given refining an example in [35] and [36] as follows. Assume $t>1$. In this case, for all $w\in\mathbb{IN}$, let $\hat{\mathbb{F}}_w$ be the smallest field, $\mathbb{F}$, whose cardinality is $|\mathbb{F}|\ge w$, $q_w\stackrel{\text{def}}{=}|\hat{\mathbb{F}}_w|\in\mathbb{IN}$, q_w prime power and $\delta_w\stackrel{\text{def}}{=}q_w-w\in\mathbb{IN}$. If we let $\mathbb{F}\stackrel{\text{def}}{=}\mathbb{F}_{w,t}\stackrel{\text{def}}{=}\hat{\mathbb{F}}_w$, $\partial S\stackrel{\text{def}}{=}\partial S_{w,t}\subseteq\hat{\mathbb{F}}_w$ be any subset such that $|\partial S_{w,t}|=w\le q_w$, the number $d\in\mathbb{IN}$ be

$$d\stackrel{\text{def}}{=}t+1-\delta_w\ge 2$$

and the degree $t+1$ polynomial $g(z)\in\hat{\mathbb{F}}_w[z]$ be

$$g(z)\stackrel{\text{def}}{=}g_{w,t}(z)\stackrel{\text{def}}{=}\prod_{\alpha\in\hat{\mathbb{F}}_w-S_{w,t}}(z-\alpha)\cdot[a(z)]^{\lfloor d/2\rfloor-(d\bmod 2)}\cdot[b(z)]^{(d\bmod 2)} \tag{53}$$

with $a(z),b(z)\in\hat{\mathbb{F}}_w[z]$ irreducible polynomials in $\hat{\mathbb{F}}_w$ such that $\deg(a)=2$ and $\deg(b)=3$ (note that there exists at least one irreducible polynomial for each degree), then, from (51) and $|\mathbb{F}|=|\hat{\mathbb{F}}_w|=q_w$,

$$\hat{\Phi}(w,t)\stackrel{\text{def}}{=}\Phi\left(\hat{\mathbb{F}}_w,g\right)\stackrel{\text{def}}{=}\frac{(q_w-1)^{\delta_w}\left(q_w^2-1\right)q_w^{d-3\cdot(d\bmod 2)-2}\left(q_w^3-1\right)^{(d\bmod 2)}}{q_w-1}=(q_w-1)^{\delta_w}(q_w+1)\,q_w^{d-3\cdot(d\bmod 2)-2}\left(q_w^3-1\right)^{(d\bmod 2)}; \tag{54}$$

and the lower bound in (52) will give,

$$D(n,t)\ge 2+t+\left\lceil\frac{n}{t+1}\right\rceil+\sum_{w=2}^{n-t-1}\left\lceil\frac{\binom{n}{w}}{(q_w-1)^{\delta_w}(q_w+1)q_w^{d-3\cdot(d\bmod 2)-2}(q_w^3-1)^{(d\bmod 2)}}\right\rceil. \tag{55}$$

In [16], the authors proposed Sidon sets to design good t-Sy0EC codes. However, as the authors in [16] mention, in general, the Sidon set code design to be practical should satisfy various requisites; among which, it should have an efficient decoding algorithm. In general, note that, for Sidon set based codes computing the error pattern from the syndrome of the received word may be difficult as t grows large because the only solution may be a table look-up method. It turns out that our σ-codes (23) are Sidon set based codes (see Section III in [10]) and, as shown in Subsection III-B, they also have an efficient t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding algorithm which exploits the maximum error correcting capabilities of the codes. Sidon set σ-codes (23) can also be good in term of redundancy, In fact, Bose-Chowla Sidon sets [5] are proposed in [16] to design large t-Sy0EC codes and, hence, derive good asymptotic lower bounds of t-Sy0EC codes, say as

$$B_{[16]}\stackrel{\text{def}}{=}B_{[16]}(n,t)\stackrel{\text{def}}{=}\sum_{w=0}^{n}\binom{n}{w}\Big/\Phi_{[16]}(w,t)\le D(n,t). \tag{56}$$

Note that the Bose-Chowla designs of Sidon sets work well when $\delta_w\stackrel{\text{def}}{=}q_w-w\in\{0,1\}$, so for a fair comparison with the lower bound in (55) given as,

$$\hat{B}\stackrel{\text{def}}{=}\hat{B}(n,t)\stackrel{\text{def}}{=}\sum_{w=0}^{n}\binom{n}{w}\Big/\hat{\Phi}(w,t)\le D(n,t), \tag{57}$$

analogously to Theorem 6.1 in [18], we let

$$\Phi_{[16]}(w,t)\stackrel{\text{def}}{=}\begin{cases}q_w^t+q_w^{t-1}+\ldots+1 & \text{if } \delta_w=0 \text{ (B-C design)},\\ q_w^t-1 & \text{if } \delta_w=1 \text{ (B-C design)},\\ q_w^t-q_w^{t-1} & \text{if } \delta_w>1.\end{cases} \tag{58}$$

Even with this choice, comparing (54) with (58) it can be seen that, for all $w,t\in\mathbb{IN}$, with $t>2$,

$$\frac{\Phi_{[16]}(w,t)}{\hat{\Phi}(w,t)}\ge 1+\frac{1}{q_w^2};$$

and so,

$$\begin{aligned}\hat{B}(n,t)=\sum_{w=0}^{n}\frac{\binom{n}{w}}{\hat{\Phi}(w,t)}&=\sum_{w=0}^{n}\left[\frac{\Phi_{[16]}(w,t)}{\hat{\Phi}(w,t)}\right]\cdot\frac{\binom{n}{w}}{\Phi_{[16]}(w,t)}\ge\\ &\sum_{w=0}^{n}\left(1+\frac{1}{q_w^2}\right)\cdot\frac{\binom{n}{w}}{\Phi_{[16]}(w,t)}=\\ &B_{[16]}(n,t)+\sum_{w=0}^{n}\frac{1}{q_w^2}\cdot\frac{\binom{n}{w}}{\Phi_{[16]}(w,t)}\ge\\ &B_{[16]}(n,t)+\frac{1}{q_n^2}\sum_{w=0}^{n}\frac{\binom{n}{w}}{\Phi_{[16]}(w,t)}=\\ &B_{[16]}(n,t)+\frac{1}{q_n^2}B_{[16]}(n,t)=\left(1+\frac{1}{q_n^2}\right)B_{[16]}(n,t).\end{aligned}$$

Bertrand's postulate [3] states that for any integer $c>1$ there is always at least one prime p such that $c<p<2c$; and so, the following non-asymptotic lower bound holds.

$$\hat{B}(n,t)\ge B_{[16]}(n,t)+\frac{1}{4n^2}B_{[16]}(n,t). \tag{59}$$

If $t=2$ then an analogous non-asymptotic bound can be obtained. This and relation (59) show that the simple choice of $g(z)\in\hat{\mathbb{F}}_w[z]$ in (53) gives some small lower bound improvement with respect to the Bose-Chowla Sidon sets choice; for all $n,t\in\mathbb{IN}$. We just mention that using the minimum value of (51) found in [10] for $t+1<q_w\lesssim n$, some even larger Sidon set σ-codes can be obtained. In any case, all these mean that Sidon set σ-codes (23) are good codes in terms of both redundancy and decoding complexity. In this paper we are mainly focusing on efficient decoding of good codes and these σ-code based codes, in general, may be difficult to encode if their cardinality is big. However, if n is small, say $n\le 46$ as in Table IV, converting information words to

codewords and viceversa can be implemented with table look-ups. With small table look-ups, various efficient and practical schemes can be designed as the following simple example.

Example 4.1 (Efficient Systematic t-Sy0EC/$(t+1)$-Sy0ED/AU0ED Coding Scheme): Let $k \in \mathbb{N}$ be the number of information bits that need to be encoded in a systematic t-Sy0EC code. Given such k, let $\mathbb{F} = \mathbb{F}_k$ be the smallest field such that $k < |\mathbb{F}|$ and $\partial S \subseteq \mathbb{F} - \{0\}$ be the index set with $k = |\partial S| \simeq |\mathbb{F}|$ distinct non-zero field elements. For an information word $X \in \mathbb{Z}_2^k$, let

$$\hat{V}_X \stackrel{\text{def}}{=} \hat{V}(X)0^{k-l(\hat{V}(X))} \in \mathbb{N}^k.$$

From (22), the σ-polynomial of $\hat{V}_X$ is

$$\sigma_{\hat{V}_X}(z) = 1 + \sigma_1(\hat{V}_X)z + \sigma_2(\hat{V}_X)z^2 + \ldots \in \mathbb{F}[z]$$

and so,

$$\sigma_{\hat{V}_X}(z) \bmod z^{t+1} = 1 + \sigma_1(\hat{V}_X)z + \sigma_2(\hat{V}_X)z^2 + \cdots + \sigma_t(\hat{V}_X)z^t.$$

Encoding: let

$$\mathcal{E}_0^{(t)} : \mathbb{F} \to \mathcal{C}_0^{(t)} \subseteq \mathcal{S}(\mathbb{Z}_2, r, \lfloor r/2 \rfloor) \tag{60}$$

be any redundancy efficient encoding to a t-Sy0EC balanced code of fixed length $r \in \mathbb{N}$, say for example, $\mathcal{C}_0^{(t)} = V^{-1}(\mathcal{A})$ with $\mathcal{A}$ given in (42) defined with the field $\mathbb{F}_{\lfloor r/2 \rfloor}$. Note that, since

$$k = |\partial S| \simeq |\mathbb{F}| \simeq |\mathcal{C}_0^{(t)}|$$

is small we have that such encodings can be efficiently precomputed and implemented with small table look-ups. At this point, the systematic t-Sy0EC/$(t+1)$-Sy0ED/AU0ED encoding, say

$$\mathcal{E}^{(t)} : \mathbb{Z}_2^k \to \mathcal{C}^{(t)} \subseteq \mathbb{Z}_2^{k+rt},$$

is defined as

$$\mathcal{E}^{(t)}(X) \stackrel{\text{def}}{=} X \;\; C_1 \;\; C_2 \ldots C_t \stackrel{\text{def}}{=}$$
$$X \;\; \mathcal{E}_0^{(t)}(\sigma_1(\hat{V}_X)) \;\; \mathcal{E}_0^{(t)}(\sigma_2(\hat{V}_X)) \ldots \mathcal{E}_0^{(t)}(\sigma_t(\hat{V}_X)).$$

Decoding: the t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding procedure is simple. Assume $\mathcal{E}^{(t)}(X) \in \mathcal{C}^{(t)}$ is sent and $F \in \mathbb{Z}_2^*$ is received in such a way that F is affected by either $(t+1)$ 0-errors or only 0-deletions or only 0-insertions. Upon receiving the binary sequence

$$F \stackrel{\text{def}}{=} Y \;\; D_1 \;\; D_2 \ldots D_t,$$

where $D_i \in \mathbb{Z}_2^*$ is the received version of $C_i = \mathcal{E}_0^{(t)}(\sigma_i(\hat{V}_X))$, for all $i = 1, 2, \ldots, t$, the decoder, from right to left parses F to compute D_t, D_{t-1}, ..., D_1 and, hence, Y. It can do so by counting the number of received bits and the number of received bits equal to 1 in F and putting a "cutting comma" in the sequence exactly between the c-th and $(c+1)$-th received bit where

$$c \stackrel{\text{def}}{=} \begin{cases} i_{\lfloor r/2 \rfloor} & \text{if } r \le i_{\lfloor r/2 \rfloor}, \\ r & \text{if } i_{\lfloor r/2 \rfloor} < r < i_{\lfloor r/2 \rfloor + 1}, \\ i_{\lfloor r/2 \rfloor + 1} - 1 & \text{if } i_{\lfloor r/2 \rfloor + 1} \le r; \end{cases}$$

and $i_w \in \mathbb{N}$ indicates the number of received bits just after the reception of the w-th bit equal to 1 in a binary sequence, for all $w \in \mathbb{N}$. For $i = t, t-1, \ldots, 1$, as soon as the receiver has parsed D_i, it decodes it with the t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding algorithm described in Section III. If for some $i \in [1, t]$, the decoding algorithm for D_i detects an error then the receiver detects an error. Otherwise, if for all $i \in [1, t]$, the decoding algorithm for D_i corrects the errors, the receiver assumes $D_i = C_i$ and computes

$$\sigma_i(\hat{V}_X) = \left[\mathcal{E}_0^{(t)}\right]^{-1}(D_i) \in \mathbb{F}.$$

Knowing, $\tilde{\sigma}(z) \stackrel{\text{def}}{=} \sigma_{\hat{V}_X}(z) \bmod z^{t+1}$, as described at the end of Section III, the receiver can apply the t-Sy0EC/$(t+1)$-Sy0ED/AU0ED decoding Algorithm 3.1 with input the constant weight code

$$\mathcal{A} = \mathcal{A}_{z^{t+1}, \tilde{\sigma}}(\mathbb{Z}_m, w(Y)+1, k - w(Y)) \subseteq$$
$$V(\mathcal{S}(\mathbb{Z}_2, k, w(Y))) \subseteq \mathcal{S}(\mathbb{N}, w(Y)+1, k - w(Y))$$

given in (42) and $V(Y) \in \mathbb{N}^{w(Y)+1}$. If Algorithm 3.1 detects errors (i. e., $cor = 0$) then the receiver detects errors and outputs a guessed codeword; if it corrects errors (i. e., $cor = 1$) then the receiver outputs the word $E' = V^{-1}(V') \in \mathbb{Z}_2^k$; $V' \in \mathbb{N}$ being the output word from Algorithm 3.1. Note that in any of the 4 possible cases: 1) F is affected by at most t 0-errors, 2) F is affected by exactly $(t+1)$ 0-errors, 3) F is affected only by 0-deletions and 4) F is affected only by 0-insertions; this t-Sy0EC/$(t+1)$-Sy0ED/AU0ED procedure works; namely, in case 1) $cor = 1$ and so $E' = \mathcal{E}^{(t)}(X)$; whereas, in the other three cases, if (by any chance) $cor = 1$ then $E' = \mathcal{E}^{(t)}(X)$.

Redundancy Analysis: The overall code length is $N \stackrel{\text{def}}{=} k + tr$. From the base code design choice (60), it follows,

$$|\mathcal{C}_0^{(t)}| \ge \frac{\binom{r}{\lfloor r/2 \rfloor}}{\lceil r/2 \rceil^t} \simeq 2^{r + t - t\log_2 r - (1/2)\log_2 r - 0.326}$$

because of the pigeon principle, the Bertrand's postulate [3] and the Stirling's approximation. So, choose $r \in \mathbb{N}$ as the smallest integer such that $|\mathbb{F}| \le |\mathcal{C}_0^{(t)}|$ in such a way that $\mathcal{E}_0^{(t)}$ be well defined as an encoding. This and $k = |\partial S| \simeq |\mathbb{F}| \simeq |\mathcal{C}_0^{(t)}|$, implies,

$$\log_2 |\mathbb{F}| \simeq r + t - (t + 1/2)\log_2 r \;\Rightarrow$$
$$r \simeq \log_2 |\mathbb{F}| + (t + 1/2)\log_2 \log_2 |\mathbb{F}| - t =$$
$$\log_2 k + (t + 1/2)\log_2 \log_2 k - t;$$

and so, the overall code redundancy is $N - k = t\log_2 k + t(t+1/2)\log_2 \log_2 k - t^2$. Since $k \le N$, this implies that,

$$\frac{N - k - t\log_2 N}{t\log_2 N} \le \frac{N - k - t\log_2 k}{t\log_2 N} =$$
$$\frac{\Theta(t^2 \log\log k)}{t\log_2 N} \le \frac{\Theta(t\log\log N)}{\log_2 N} = \Theta\left(\frac{t\log\log N}{\log N}\right).$$

Hence, if $t = t(N) = o(\log N / \log\log N)$ then

$$\frac{N - k - t\log_2 N}{t\log_2 N} = \Theta\left(\frac{t\log\log N}{\log N}\right) \to 0;$$

and this systematic t-Sy0EC/$(t+1)$-Sy0ED/AU0ED coding scheme example is asymptotically optimal according to

Definition 4.1 because of Theorem 4.1. For example, if $t = 2$ and $r = 18$ then $\binom{r}{\lfloor r/2 \rfloor} / \lceil r/2 \rceil^t = 601$ and so this simple design gives a systematic 2-Sy0EC/3-Sy0ED/AU0ED coding scheme with $k = 601$ information bits and length

$$N = k + 2t = 601 + 2 \cdot 18 = 637.$$

Note that, if $t = 2$ and $r \geq 18$ then $0 < 2t \log_2 N - tr \to +\infty$. ∎

To reach the asymptotical optimal redundancy, Theorem 4.1 tells us that t can be as big as $t = t(n) = 2^{o(\log n)}$ showing some room for improvement. The above Example 4.1 shows that the challenging design problem for reaching the asymptotical optimal redundancy is essentially a matter of appropriately governing synchronization within the various parts of a codeword. Note that in designing systematic codes there are at least two parts: the information part and the check part. As mentioned before, [42] fills this gap and more by giving fixed length $n \in \mathbb{IN}$ systematic recursive σ-code based codes with efficient t-Sy0EC/$(t+1)$-Sy0ED/AU0ED error control algorithms and redundancy $n - k \leq t \log_2 k + o(t \log n)$ bits; and all this for any $k, t \in \mathbb{IN}$.

If $t = 1$ then the multiplication operation of a field is not needed and the single asymmetric error correcting group-theoretic code [6], [7], [18] based design can be used. In this case, for all $w \in \mathbb{IN}$, $\hat{\mathbb{F}}_{w,1}$ can be any Abelian group with $w + 1$ distinct elemants, $\partial S_{w,1} = \hat{\mathbb{F}}_{w,1} - \{0\}$ and we obtain,

$$\left\lceil \frac{2^{n+1} - 1}{n+1} \right\rceil \leq \sum_{w=0}^{n} \left\lceil \binom{n}{w} \Big/ (w+1) \right\rceil \leq D(n,1); \quad (61)$$

which is the lower bound given in [11]. So, let $\mathbb{F}_w$ be the smallest field, $\mathbb{F}$, whose cardinality is $|\mathbb{F}| > w$, when $t > 1$; and $\mathbb{F}_w$ be any Abelian group of cardinality no less than $w+1$ (such as the cyclic group $(\mathbb{Z}_{w+1}, + \bmod (w+1))$) when $t = 1$. Note that $|\hat{\mathbb{F}}_w| \leq |\mathbb{F}_w|$. In Subsection III-A, for a given $t \in \mathbb{IN}$, we let

$$(\mathbb{F}_{w,t}, \partial S_{w,t}, g_{w,t}) = (\mathbb{F}_w, \mathbb{F}_w - \{0\}, z^{t+1}), \text{ for all } w \in \mathbb{IN}.$$

With this simplifying choice of triplet sequence, the lower bound in (52) becomes,

$$2 + t + \left\lceil \frac{n}{t+1} \right\rceil + \sum_{w=2}^{n-t-1} \left\lceil \binom{n}{w} \Big/ |\mathbb{F}_w|^t \right\rceil \leq D(n,t); \quad (62)$$

which is worse than the lower bounds in (52) and (55), but still improves the non-asymptotic lower bounds found in the literature.

From (28) a new non-asymptotic lower bound can be derived which relates $D(n,t)$ with the prime power gap sequence.

Theorem 4.4: Let $\mathbb{PP}$ the prime power sequence as in (29), $q_0 \stackrel{\text{def}}{=} 1$ and,

$$\Gamma_w \stackrel{\text{def}}{=} \max_{q_i \in \mathbb{PP} \cap [0, |\mathbb{F}_w|]} (q_i - q_{i-1}), \quad \text{for all } w \in \mathbb{IN} - \{0\}.$$

If $n \in \mathbb{IN}$ and $t \stackrel{\text{def}}{=} t(n) \in \mathbb{IN}$ is any then

$$D(n,t) \geq \sum_{w=0}^{n} \left\lceil \binom{n}{w} \Big/ |\mathbb{F}_w|^t \right\rceil \geq \quad (63)$$
$$\frac{1}{2\left[1 + \frac{2\Gamma_{\lceil n/2 \rceil} + (n \bmod 2)}{n}\right]^t} \cdot \frac{2^{n+t}}{n^t} \geq$$
$$\frac{2^{n+t}}{2\left(1 + \frac{2\Gamma_{\lceil n/2 \rceil} + 1}{n}\right)^t n^t}.$$

Considering primes only, Cramér's conjecture asymptotically quantifies the gap value between two consecutive primes to be as small as $O\left(\log^2 p\right)$; p being the smallest between the two prime values [9]. If this conjecture is true then $\Gamma_n = O\left(\log^2 n\right)$ as n gets large. In any case, Bertrand's postulate [3] states that for any integer $c > 1$ there is always at least one prime p such that $c < p < 2c$. This implies, $\Gamma_n \leq n$.

Proof: For all $v \in \mathbb{IN}$, let

$$\delta_v \stackrel{\text{def}}{=} |\mathbb{F}_v| - (v+1) \geq 0 \quad (64)$$

and $i(v) \in \mathbb{IN}$ be the index such that $|\mathbb{F}_v| = q_{i(v)} \in \mathbb{PP}$. From the minimality of $|\mathbb{F}_v|$ in the definition of $\mathbb{F}_v$, we have $v + 1 > q_{i(v)-1} \in \mathbb{PP}$, and so,

$$0 \leq \delta_v = |\mathbb{F}_v| - (v+1) =$$
$$q_{i(v)} - (v+1) < q_{i(v)} - q_{i(v)-1}.$$

So, for all $w \in \mathbb{IN}$ and for all integer $u \in [0, w]$,

$$0 \leq \delta_u \leq \max_{v \in [0,w]} \delta_v < \max_{v \in [0,w]} (q_{i(v)} - q_{i(v)-1}) =$$
$$\max_{q_i \in \mathbb{PP} \cap [0, |\mathbb{F}_w|]} (q_i - q_{i-1}) = \Gamma_w \quad \Rightarrow$$
$$0 \leq \delta_u \leq \Gamma_w - 1, \quad \text{for all integer } u \in [0, w]. \quad (65)$$

Now, from the definition of $D(n,t)$, (28), (64), (65) and $\sum_{w=0}^{\lceil n/2 \rceil} \binom{n}{w} \geq 2^{n-1}$, it follows,

$$D(n,t) \geq \sum_{w=0}^{n} \left\lceil \binom{n}{w} \Big/ |\mathbb{F}_w|^t \right\rceil \geq \sum_{w=0}^{n} \binom{n}{w} \Big/ |\mathbb{F}_w|^t =$$
$$\sum_{w=0}^{n} \binom{n}{w} \Big/ (w + 1 + \delta_w)^t \geq$$
$$\sum_{w=0}^{n} \binom{n}{w} \Big/ (w + \Gamma_w)^t \geq$$
$$\sum_{w=0}^{\lceil n/2 \rceil} \binom{n}{w} \Big/ (w + \Gamma_w)^t \geq$$
$$\sum_{w=0}^{\lceil n/2 \rceil} \binom{n}{w} \Big/ \left(\lceil n/2 \rceil + \Gamma_{\lceil n/2 \rceil}\right)^t \geq$$
$$\frac{2^{n-1}}{\left(\lceil n/2 \rceil + \Gamma_{\lceil n/2 \rceil}\right)^t} =$$
$$\frac{2^{n-1}}{(n/2)^t \left(1 + (n \bmod 2)/n + 2\Gamma_{\lceil n/2 \rceil}/n\right)^t} =$$
$$\frac{2^{n+t}}{n^t} \cdot \frac{1}{2\left\{1 + \left[2\Gamma_{\lceil n/2 \rceil} + (n \bmod 2)\right]/n\right\}^t}.$$

This implies (63); hence, the Theorem. ∎

TABLE V

CODE EXAMPLE WITH $n = 5$ AND $t = 1$. EXAMPLE OF A 1-SY0EC/2-SY0ED/AU0ED CODE $\mathcal{C}$ OF LENGTH 5 WITH $|\mathcal{C}| = 14$ CODEWORDS. FOR $n = 5$ AND $t = 1$ THE NON-ASYMPTOTIC LOWER BOUND IN (61) GIVES 13, THE NON-ASYMPTOTIC UPPER BOUND IN (68) GIVES 16; HOWEVER, THE ASYMPTOTIC **Upper** BOUND IN [11], [16], [21] GIVES (INCORRECTLY) $\lfloor 64/5 \rfloor = 12$. WE BELIEVE $\mathcal{C}$ TO BE OPTIMAL

$l(X) = n$	$w(X)$	X	$V(X) = \hat{V}(X)\boldsymbol{v}_{w+1}$	$l(V(X)) = w(X) + 1$	$w(V(X))$	$\lvert\mathcal{C}\rvert = 14$
5	0	00000	**5**	1	5	1
5	1	10000 00100 00001	0**4** 2**2** 4**0**	2	4	3
5	2	11000 01010 10001 00011	00**3** 11**1** 03**0** 30**0**	3	3	4
5	3	11100 11001 10011 00111	000**2** 002**0** 020**0** 200**0**	4	2	4
5	4	11110	0000**1**	5	1	1
5	5	11111	00000**0**	6	0	1

From Theorem 4.4, the relation $\Gamma_{\lceil n/2 \rceil} \leq n/2 - 2$ (which is valid for $n \geq 6$), if $n \in \mathbb{N}$, with $n \geq 6$, and $t = t(n) \in \mathbb{N}$ is any then

$$n - \log_2 |D(n,t)| \leq t \log_2 n - t + t \log_2 \left(1 + \frac{2\Gamma_{\lceil n/2 \rceil} + 1}{n}\right) + 1 \leq t \log_2 n + 1 = t \log_2 n + o(t \log n);$$

which proves the upper bound side of (44).

Relation (63) is non-asymptotic. However, if $t = t(n)$ is small with respect to n and, in particular, $t = t(n) = o(n/\Gamma_n)$ then (63) gives the following asymptotic lower bound,

$$D(n,t) \gtrsim \frac{1}{2} \cdot \frac{2^{n+t}}{n^t}; \tag{66}$$

which is only half the asymptotic lower bound proved in [11] and [16] for fixed t. But something more can be done. In particular, if $\lambda \in (0, 1/2)$ is a real constant,

$$v \stackrel{\text{def}}{=} v(n) \stackrel{\text{def}}{=} \lambda n + o(n)$$

and

$$t = t(n) = o(\min\{n/\log n, n/\Gamma_n\})$$

then, as n grows,

$$D(n,t) \geq \sum_{w=0}^{n} \left\lceil \binom{n}{w} \Big/ |\mathbb{F}_w|^t \right\rceil \geq \sum_{w=v}^{n} \left\lceil \binom{n}{w} \Big/ |\mathbb{F}_w|^t \right\rceil \gtrsim \sum_{w=0}^{n} \binom{n}{w} \Big/ (w+1)^t \stackrel{\text{def}}{=} C(n,t).$$

In this case, by integrating t times $(1+x)^n = \sum_{w=0}^{n} x^w$, letting $x = 0$ to find the t integration constants and then letting $x = 1$, it follows

$$C(n,t) \gtrsim \frac{2^{n+t}}{(n+1)(n+2) \cdot \ldots \cdot (n+t)} = \frac{2^{n+t}}{t! \binom{n+t}{t}}$$

if $s = n + t$ grows large. Furthermore, if $t = o(n)$ then $2^{n+t}/t!\binom{n+t}{t} \gtrsim 2^{n+t}/n^t$. All these imply the following theorem which generalizes the asymptotic lower bound result in [11] and [16].

Theorem 4.5: If $t = t(n) = o(\min\{n/\log n, n/\Gamma_n\})$ then $D(n,t) \gtrsim 2^{n+t}/n^t$.

We mention that the bound in (63) and $C(n,t)$ are bigger than the hypergraph based non constructive non asymptotic lower bound given in [19].

B. Upper Bounds

With regard to the non-asymptotic upper bound, using a sphere packing argument applied to constrained weight $n - w$ codes of length w over the alphabet $\mathbb{N}$ with minimum asymmetric L_1 distance greater than t, we will prove here that for $t \leq n$,

$$LW(w, n-w, t) \leq \left\lfloor \binom{n+2t}{w+t} \Big/ \binom{n+2t}{t} \right\rfloor.$$

So, from the equality (47), the following simple explicit non-asymptotic upper bound holds for $D(n,t)$.

$$D(n,t) \leq \sum_{w=0}^{n-t-1} \left\lfloor \binom{n+2t}{w+t} \Big/ \binom{n+2t}{t} \right\rfloor + t + 1. \tag{67}$$

Note that the above bound can be used to prove (48) and, most importantly, to get the interesting big picture information represented in Figure 1. We are not aware of any explicit non-asymptotic upper bound for $D(n,t)$. When $t = 1$, from (67), we obtain the following new non-asymptotic upper bound for $D(n,1)$,

$$D(n,1) \leq \sum_{w=0}^{n-2} \left\lfloor \binom{n+2}{w+1} \Big/ (n+2) \right\rfloor + 2 \leq \left\lfloor \frac{2^{n+2} - (n+2)(n-1)/2 - 2}{n+2} \right\rfloor. \tag{68}$$

Note that, if $n = 5$ then the asymptotic upper bound in [11], [16], and [21] for $t = 1$ gives $D(n,1) \lesssim 2^{n+1}/n \leq 64/5 = 12.8$, however the code example in Table V contains

14 codewords (may be, it is optimal). On the other hand, the above upper bound (correctly) gives $D(5,1) \leq 16$.

Recall the general definitions of the values $CW(A,n,w,t)$ and $LW(A,n,w,t)$ given at the beginning of this Section IV. The following theorems give a simple upper bound on the cardinality of t-asymmetric/unidirectional EC codes over $\mathbb{Z}_m$, $m \in \mathbb{N} \cup \{\infty\}$.

Theorem 4.6: For all $m \in \mathbb{N} \cup \{\infty\}$ and $n,w,t \in \mathbb{N}$ the following relation holds.

$$CW(\mathbb{Z}_m, n+1, w, t) = LW(\mathbb{Z}_m, n, w, t) \leq \left\lfloor \sum_{v=w-(m-1)}^{w+t} \binom{n}{v}_{m+\mu} \Big/ \sum_{\tau=0}^{t} \binom{n}{\tau}_m \right\rfloor; \tag{69}$$

where

$$\mu \stackrel{\text{def}}{=} \mu(m,t) \stackrel{\text{def}}{=} \min\{m-1, t\}.$$

In particular, if $m = \infty$ then

$$LW(\mathbb{N}, n, w, t) \leq \left\lfloor \binom{n+w+t}{n} \Big/ \binom{n+t}{n} \right\rfloor, \tag{70}$$

Proof: See the Appendix. ∎

From Theorem 4.2 and Theorem 4.6, the non-asymptotic upper bound on t-Sy0EC codes can be derived easily in the following theorem.

Theorem 4.7 (Explicit Upper bound on t-Sy0EC): If $\mathcal{C} \subseteq \mathbb{Z}_2^n$ is any t-Sy0EC binary code of length n then

$$|\mathcal{C}| \leq D(n,t) = \sum_{w=0}^{n-t-1} LW(\mathbb{N}, w, n-w, t) + t + 1 \leq \sum_{w=0}^{n-t-1} \left\lfloor \binom{n+2t}{w+t} \Big/ \binom{n+2t}{t} \right\rfloor + t + 1 \leq \tag{71}$$

$$\frac{2^{n+2t} - \sum_{\tau=0}^{2t} \binom{n+2t}{\tau} - \sum_{\tau=0}^{t-1} \binom{n+2t}{\tau}}{\binom{n+2t}{t}} + t + 1 \leq \frac{2^{n+2t} - \sum_{\tau=0}^{2t} \binom{n+2t}{\tau}}{\binom{n+2t}{t}} + t + 1 \leq 2^{n+2t} \Big/ \binom{n+2t}{t}.$$

where the above inequalities hold for any $n,t \in \mathbb{N}$.

Proof: The leftmost relations in the first line of (71) come from Theorem 4.2 and Theorem 4.3. The remaining relations follow from (70) and

$$LW(\mathbb{N}, w, n-w, t) \leq \left\lfloor \binom{w+(n-w)+t}{w} \Big/ \binom{w+t}{w} \right\rfloor = \left\lfloor \binom{n+t}{w} \Big/ \binom{w+t}{w} \right\rfloor = \left\lfloor \frac{(n+t)!}{w!(n+t-w)!} \cdot \frac{w!t!}{(w+t)!} \right\rfloor = \left\lfloor \frac{(n+2t)!}{(n+t-w)!(w+t)!} \cdot \frac{(n+t)!t!}{(n+2t)!} \right\rfloor = \left\lfloor \binom{n+2t}{w+t} \Big/ \binom{n+2t}{t} \right\rfloor \leq \binom{n+2t}{w+t} \Big/ \binom{n+2t}{t}.$$

So, the theorem is proved. ∎

In Lemma 2 of [21], Levenshtein proved that if t is fixed and $n \to \infty$ then $D(n,t) \lesssim t!2^{n+t}/n^t$. This asymptotic upper bound was generalized and improved in [16] by a factor of $\binom{t}{t/3}/2^{t/3} = 2^{[h(1/3)-1/3]t} = 2^{0.585 \cdot t}$. Note that even though the upper bound in Theorem 4.7 is roughly only $2^{\theta(t)}$ times bigger than these asymptotic upper bounds, the bounds in (71) hold true for any value of $t,n \in \mathbb{N}$ (and not only for fixed constant t). In particular, we note that Levenshtein's argument in proving his upper bound can be carried out only if t is roughly less than $n/4$, in which case, however, if

$$\mu \stackrel{\text{def}}{=} \left\lfloor \left(n - \sqrt{2tn \log_e n}\right)/4 \right\rfloor$$

and $t = o(n/\log n)$ then,

$$D(n,t) = D'_\mu(n,t) + D''_\mu(n,t) < 2^{n-t} \Big/ \binom{\mu}{t} + 2\sum_{v=0}^{2\mu} \binom{n}{v} \lesssim 2^{n-t} \Big/ \binom{\mu}{t} + \frac{2^{n+1}}{n^{2t}}.$$

In fact, let $h(x)$ be the binary entropy function,

$$h(x) \stackrel{\text{def}}{=} -[x \log_2 x + (1-x)\log_2(1-x)], \text{ with } x \in [0,1].$$

Using the well known bound

$$\sum_{v=0}^{w} \binom{n}{w} \leq 2^{h(w/n) \cdot n}, \quad \text{for } n,w \in \mathbb{N} \text{ and } w \in [0, n/2], \tag{72}$$

by letting

$$w \stackrel{\text{def}}{=} 2\mu \simeq \frac{n}{2} - \frac{\sqrt{2tn \log_e n}}{2} \stackrel{\text{def}}{=} \frac{n}{2} - \Delta,$$

we have that if

$$\frac{\Delta}{n} = \frac{\sqrt{2tn \log_e n}}{2n} = \sqrt{\frac{t \log_e n}{2n}} \to 0$$

then

$$-h\left(\frac{w}{n}\right) \cdot n = w \log_2 \frac{w}{n} + (n-w)\log_2 \frac{n-w}{n} = \left(\frac{n}{2} - \Delta\right)\log_2\left(\frac{1}{2} - \frac{\Delta}{n}\right) + \left(\frac{n}{2} + \Delta\right)\log_2\left(\frac{1}{2} + \frac{\Delta}{n}\right) = -n + \left(\frac{n}{2} - \Delta\right)\log_2\left(1 - \frac{2\Delta}{n}\right) + \left(\frac{n}{2} + \Delta\right)\log_2\left(1 + \frac{2\Delta}{n}\right) \simeq -n + \left[\left(\frac{n}{2} - \Delta\right)\left(-\frac{2\Delta}{n}\right) + \left(\frac{n}{2} + \Delta\right)\left(\frac{2\Delta}{n}\right)\right] \cdot \log_2 e = -n + \frac{4\Delta^2}{n} \cdot \log_2 e = -n + 2t \log_2 n;$$

and so,

$$\sum_{v=0}^{2\mu}\binom{n}{v} \le 2^{h(2\mu/n)\cdot n} \simeq 2^{n-2t\log_2 n} = \frac{2^n}{(n^t)^2} = \frac{2^n}{n^{2t}}.$$

However, if $t = o(\sqrt{n}) = o(n/\log n)$ then

$$D(n,t) \lesssim 2^{n-t}\left/\binom{\mu}{t} + \frac{2^{n+1}}{n^{2t}} \lesssim 2^t t!\frac{2^n}{n^t} + \frac{2^{n+1}}{n^{2t}} \lesssim 2^t t!\frac{2^n}{n^t};\right.$$

as $n\to\infty$. In this way, Levenshtein's asymptotic upper bound can be generalized as follows.

Theorem 4.8: If $t = t(n) = o(\sqrt{n})$ then $D(n,t) \lesssim 2^t t!\frac{2^n}{n^t}$.

The above theorem is more than enough to prove the lower bound side of (44).

Since $t = t(n) \in \mathbb{N}$ can be any in (71), some interesting big picture considerations can be derived from (71). In fact, the following theorem holds.

Theorem 4.9: For any $n, t \in \mathbb{N}$ and $s \stackrel{\text{def}}{=} n + 2t \in \mathbb{N}$,

$$D(n,t) \le \begin{cases} 2^{[1-h(t/s)]s+(\log_2 t)/2+3/2} & \\ \qquad\qquad \text{if } t\in(0,(n-2)/2), & \\ 2^{[h((2t+1)/s)-h(t/s)]s+(\log_2 t)/2+3/2}+t+1 & \\ \qquad\qquad \text{if } t\in[(n-2)/2,n-1), & \\ n+1 \qquad\qquad \text{if } t\in[n-1,+\infty); & \end{cases}$$

where, note

$$t\in\left(0,\frac{n-2}{2}\right) \iff \frac{t}{s} = \frac{t}{n+2t} \in \left(0,\frac{1}{4}\cdot\frac{1-2/n}{1-1/n}\right) \simeq \left(0,\frac{1}{4}\right)$$

and

$$t\in\left[\frac{n-2}{2},n-1\right) \iff \frac{t}{s}\in\left[\frac{1}{4}\cdot\frac{1-2/n}{1-1/n},\frac{1}{3}\cdot\frac{1-1/n}{1-2/(3n)}\right) \simeq \left[\frac{1}{4},\frac{1}{3}\right) \iff \frac{2t+1}{s} = \frac{2t+1}{n+2t} \in \left[\frac{1}{2},\frac{2}{3}\cdot\frac{1-1/(2n)}{1-2/(3n)}\right) \simeq \left[\frac{1}{2},\frac{2}{3}\right).$$

Proof: From the following well known approximation

$$\sqrt{2\pi n}n^n e^{-n}e^{1/(12n+1)} < n! < \sqrt{2\pi n}n^n e^{-n}e^{1/(12n)};$$

the following lower bound holds

$$\binom{n}{w} \ge \frac{2^{h(w/n)\cdot n}}{\sqrt{8n(w/n)(1-w/n)}} = \frac{2^{h(w/n)\cdot n}}{2^{h(w/n)\cdot n-(1/2)[\log_2 w+3+\log_2(1-w/n)]}} \qquad \text{for } n,w\in\mathbb{N} \text{ and } w\in[1,n-1]. \tag{73}$$

So, the theorem follows from (72), (73), $h(1-x) = h(x)$ and, depending on the cases $t\in(0,(n-2)/2)$ and $t\in[(n-2)/2,n-1)]$, from the two upper bounds in the last line of (71). Note that $D(n,t) = n+1$ if $t\in[n-1,+\infty)$ because of Theorem 4.3. ∎

A direct consequence of Theorem 4.9 is the following theorem.

Theorem 4.10 (Upper Bound on the Asymptotic Information Rate of t-Sy0EC Codes): For any $n, t\in\mathbb{N}$ let $\tau \stackrel{\text{def}}{=} t/n \in \mathbb{R}$ and

$$IUB_\infty(\tau) \stackrel{\text{def}}{=} \begin{cases} \left[1-h\left(\frac{\tau}{1+2\tau}\right)\right](1+2\tau) & \\ \qquad\qquad \text{if } \tau\in(0,1/2), & \\ \left[h\left(\frac{2\tau}{1+2\tau}\right)-h\left(\frac{\tau}{1+2\tau}\right)\right](1+2\tau) & \\ \qquad\qquad \text{if } \tau\in[1/2,1), & \\ 0 \qquad\qquad \text{if } \tau\in[1,+\infty). & \end{cases}$$

The asymptotic information rate of any infinite family of t-Sy0EC codes of length n with $k\in\mathbb{R}$ information bits, $n,t\in\mathbb{N}$, satisfies the following relations,

$$\lim_{n\to\infty}\frac{k}{n} \le \lim_{n\to\infty}\frac{\log_2 D(n,t)}{n} \le IUB_\infty(\tau).$$

Proof: First note that $k \le \log_2 D(n,t)$, for all $n,t\in\mathbb{N}$; and so, $\lim_{n\to\infty} k/n \le \lim_{n\to\infty}\log_2 D(n,t)/n$. Now, since

$$t\in(0,n-1) \;\Rightarrow\; \lim_{n\to\infty}\frac{(\log_2 t)/2+3/2}{n} = 0,$$

it follows,

$$s = n+2t \text{ and } t\in\left[\frac{n-2}{2},n-1\right) \;\Rightarrow\; h\left(\frac{2t+1}{s}\right)-h\left(\frac{t}{s}\right)\ge 0 \;\Rightarrow\; \left[h\left(\frac{2t+1}{s}\right)-h\left(\frac{t}{s}\right)\right]s+\frac{\log_2 t}{2}+\frac{3}{2}\ge 0 \;\Rightarrow\; \lim_{n\to\infty}\frac{1}{n}\log_2\left(1+\frac{t+1}{2^{[h((2t+1)/s)-h(t/s)]s+(\log_2 t)/2+3/2}}\right) = 0,$$

and

$$\frac{t}{s} = \frac{t}{n+2t} = \frac{(t/n)}{1+2(t/n)} \iff \frac{t}{n} = \frac{(t/s)}{1-2(t/s)}.$$

So, by letting $\tau \stackrel{\text{def}}{=} t/n$, from Theorem 4.9, it follows,

$$\lim_{n\to\infty}\frac{\log_2 D(n,t)}{n} \le \begin{cases} (1-h(t/s))(s/n) & \text{if } t/s\in(0,1/4), \\ (h(2t/s)-h(t/s))(s/n) & \text{if } t/s\in[1/4,1/3), \\ 0 & \text{if } t/s\in[1/3,+\infty); \end{cases} = IUB_\infty(\tau);$$

and the theorem is proved. ∎

Figure 1 gives the plot of $IUB_\infty(\tau)$. Note that the quantity $IUB_\infty(\tau)$ is an upper bound for the general (deletion) case channel model because of Lemma 1 in [22]. Comparing this plot with the analogous plot in [20] for the general case it can be noticed that $IUB_\infty(\tau)$ improves on the general case upper bound in [20] for the values of $\tau = t/n$ which roughly belong to the real set $[0.28,0.35]\cup[0.43,1)$.

V. Concluding Remarks

Some theory and efficient design of binary block codes capable of controlling the deletions and/or insertions of the symbol "0" (i. e., the 0-errors) are given. It is shown that the design of codes for insertion and/or deletion of zeros is equivalent to the design of the L_1 metric error control codes. Some close to optimal non-systematic codes for correcting these errors are described and their encoding and decoding methods are also explained. Based on the theory given here, some efficient t-Sy0EC systematic codes are given in [42].

Please note that based on the theory developed in this paper and [6], [8], [35], [36], [37], [38], [39], [40], [41], [43], whenever it is possible to define an isometry from the metric space which characterizes a given coding problem to the L_1 metric (as the mapping V in (15)), any information on codes for the L_1 metric reflects in the analogous information for that coding problem. In particular, as we mentioned in Section II, the sticky channel error control problem [11], [25], [37] can be reduced to the L_1 metric error control problem through the isometry given by the composition of the Gray mapping and the V mapping. Also, using Theorem 4.6 and L_1 error control codes over $\mathbb{Z}_m$, with $m \in \mathbb{N} \cup \{\infty\}$, lower bounds, upper bounds, code designs and decoding algorithms can be given for the t-Sy0EC codes which satisfy the $RLL(d,k)$ constraint [23], [26]. This is because the set of all $RLL(d,k)$ binary words of length n and weight w with the $d_{0\text{-}D/I}$ metric can be put in bijection with $\left(\mathbb{Z}_{k-d+1}^{w+1}, d_{L_1}^{sy}\right)$ through the following isometry

$$0^{v_1}\mathbf{1}0^{v_2}\mathbf{1}\ldots 0^{v_w}\mathbf{1}0^{v_{w+1}} \;\leftrightarrow\; (v_1-d, v_2-d, \ldots, v_w-d).$$

Likewise, the bit-shift coding problem described in [17] and [23] can be solved with the following isometry from the appropriate metric space $(\mathcal{S}(\mathbb{Z}_2, n, w), d_{bit-shift})$ into the metric space $\left(\mathbb{Z}_n^w, d_{L_1}^{sy}\right)$,

$$0^{v_1}\mathbf{1}0^{v_2}\mathbf{1}\ldots 0^{v_w}\mathbf{1}0^{v_{w+1}} \;\leftrightarrow\; (v_1, v_1+v_2+1, \ldots, v_1+v_2+\ldots+v_w+w-1)$$

which associates any binary word with its support. Also, the generalization to the q-ary case, $q \in \mathbb{N}$, of this 0-error problem is possible. In this case, assuming for notational convenience that x-errors are possible if, and only if $x \in \mathbb{Z}_q - \{0\} \stackrel{\text{def}}{=} \{1, 2, \ldots, (q-1)\}$, then this, say "0-reliable symbol problem", can be solved with the isometry

$$\begin{aligned} &1^{v_1}2^{v_2}\ldots(q-1)^{v_{(q-1)}}\mathbf{0} \\ &\quad 1^{v_q}2^{v_{q+1}}\ldots(q-1)^{v_{2(q-1)}}\mathbf{0} \\ &\quad \vdots \\ &\quad 1^{v_{w(q-1)+1}}2^{v_{w(q-1)+2}}\ldots(q-1)^{v_{(w+1)(q-1)}} \;\leftrightarrow \\ &\qquad (v_1, v_2, \ldots, v_{(w+1)(q-1)}); \end{aligned}$$

where w indicates the number of 0 in any q-ary word. Instead, note that the q-ary repetition error problem can be solved with the simpler isometry mentioned in [37]. Also the above mentioned binary problems and many others could be generalized to the q-ary case and addressed analogously.

Appendix

Theorem 5.1 (Sphere Packing Upper Bound on the Cardinality of Asymmetric Error Control Codes Over $\mathbb{Z}_m$):

For all $m \in \mathbb{N} \cup \{\infty\}$ and $n, w, t \in \mathbb{N}$ the following relation holds.

$$CW(\mathbb{Z}_m, n+1, w, t) = LW(\mathbb{Z}_m, n, w, t) \leq \left\lfloor \sum_{v=w-(m-1)}^{w+t} \binom{n}{v}_{m+\mu} \Big/ \sum_{\tau=0}^{t} \binom{n}{\tau}_m \right\rfloor;$$

where

$$\mu \stackrel{\text{def}}{=} \mu(m,t) \stackrel{\text{def}}{=} \min\{m-1, t\}.$$

In particular, if $m = \infty$ then

$$LW(\mathbb{N}, n, w, t) \leq \left\lfloor \binom{n+w+t}{n} \Big/ \binom{n+t}{n} \right\rfloor,$$

Proof: The above upper bound for $LW(\mathbb{Z}_m, n, w, t)$ can be derived by using a sphere packing argument as follows. For any $X \in \mathbb{Z}_m^n$ consider the "(positive) m-ary asymmetric ball centered at X with radius t" defined (in Frobenius notation) as

$$\mathcal{B}(X,t) \stackrel{\text{def}}{=} \mathcal{B}_{L_1}^{as}(n, \mathbb{Z}_m, X, t) = X + \mathcal{B}(\emptyset, t) = \{Z \in \mathbb{N}^n :\; Z = X + E,\; E \in \mathbb{Z}_m^n \text{ and } |E| \leq t\}; \tag{74}$$

where, from (7),

$$\mathcal{B}(\emptyset,t) = \{E \in \mathbb{Z}_m^n :\; |E| \leq t\} = \bigcup_{\tau=0}^{t} \{E \in \mathbb{Z}_m^n :\; |E| = \tau\} = \bigcup_{\tau=0}^{t} \mathcal{S}(\mathbb{Z}_m, n, \tau).$$

Since the above union is a disjoint union of spheres, from (74) it follows

$$|\mathcal{B}(X,t)| = |\mathcal{B}(\emptyset,t)| = \sum_{\tau=0}^{t} |\mathcal{S}(\mathbb{Z}_m, n, \tau)| = \sum_{\tau=0}^{t} \binom{n}{\tau}_m. \tag{75}$$

In particular, if $m = \infty$ then

$$|\mathcal{B}(X,t)| = |\mathcal{B}(\emptyset,t)| = \sum_{\tau=0}^{t} |\mathcal{S}(\mathbb{N}, n, \tau)| = \sum_{\tau=0}^{t} \binom{n+\tau-1}{n-1} = \binom{n+t}{n} \tag{76}$$

because of (12). Now, let $\hat{\mathcal{A}} \subseteq \mathbb{Z}_m^n$ be any m-ary code of length n and

$$\mathcal{B}(\hat{\mathcal{A}}, t) \stackrel{\text{def}}{=} \bigcup_{X \in \hat{\mathcal{A}}} \mathcal{B}(X,t). \tag{77}$$

If the minimum asymmetric L_1 distance of $\hat{\mathcal{A}}$ is $d_{L_1}^{as}(\hat{\mathcal{A}}) > t$ then the code $\hat{\mathcal{A}}$ is a $(0,t)$-EC code, and so,

$$\text{for all } X, Y \in \hat{\mathcal{A}},\;\; X \neq Y \;\Rightarrow\; \mathcal{B}(X,t) \cap \mathcal{B}(Y,t) = \emptyset. \tag{78}$$

In fact, if $Z \in \mathcal{B}(X,t) \cap \mathcal{B}(Y,t) \neq \emptyset$ then $|X \dot{-} Y| \leq |X \dot{-} Z| + |Z \dot{-} Y| \leq 0 + t = t$ and $|Y \dot{-} X| \leq |Y \dot{-} Z| + |Z \dot{-} X| \leq 0 + t = t$; and so, $d_{L_1}^{as}(X,Y) = \min\{|X \dot{-} Y|, |Y \dot{-} X|\} \leq t$. That is, $d_{L_1}^{as}(\hat{\mathcal{A}}) \leq t$. From (78), the union in (77) is a disjoint union of m-ary asymmetric balls. From (75),

these m-ary asymmetric balls have the same cardinality. So, if $d_{L_1}^{as}(\hat{\mathcal{A}}) > t$ then

$$|\mathcal{B}(\hat{\mathcal{A}},t)| = \left|\bigcup_{X\in\hat{\mathcal{A}}} \mathcal{B}(X,t)\right| = \sum_{X\in\hat{\mathcal{A}}} |\mathcal{B}(X,t)| = |\mathcal{B}(\emptyset,t)| \cdot |\hat{\mathcal{A}}|;$$

that is,

$$|\hat{\mathcal{A}}| = \frac{|\mathcal{B}(\hat{\mathcal{A}},t)|}{|\mathcal{B}(\emptyset,t)|}. \tag{79}$$

Note that, if $A = \mathbb{Z}_m$ then $\max_{a\in A} a = (m-1)$. So, if $\hat{\mathcal{A}}$ is a constrained weight

$$w \in \left[w - \max_{a\in A} a, w\right] = [w-(m-1), w]$$

code then, from (77),

$$\mathcal{B}(\hat{\mathcal{A}},t) \stackrel{\text{def}}{=} \bigcup_{X\in\hat{\mathcal{A}}} \mathcal{B}(X,t) \subseteq \bigcup_{v=w-(m-1)}^{w+t} \mathcal{S}(\mathbb{Z}_{m+\mu}^n, n, v);$$

where $\mu \stackrel{\text{def}}{=} \mu(m,t) \stackrel{\text{def}}{=} \min\{m-1, t\}$. Since, the rightmost union above is a disjoint union, it follows

$$\begin{aligned} |\mathcal{B}(\hat{\mathcal{A}},t)| &\le \left|\bigcup_{v=w-(m-1)}^{w+t} \mathcal{S}(\mathbb{Z}_{m+\mu}, n, v)\right| = \\ &\sum_{v=w-(m-1)}^{w+t} |\mathcal{S}(\mathbb{Z}_{m+\mu}, n, v)| = \\ &\sum_{v=w-(m-1)}^{w+t} \binom{n}{v}_{m+\mu}. \end{aligned} \tag{80}$$

In particular, if $m = \infty$ then

$$\begin{aligned} |\mathcal{B}(\hat{\mathcal{A}},t)| &\le \left|\bigcup_{v=0}^{w+t} \mathcal{S}(\mathbb{N}, n, v)\right| = \sum_{v=0}^{w+t} |\mathcal{S}(\mathbb{N}, n, v)| = \\ &\sum_{v=0}^{w+t} \binom{n+v-1}{n-1} = \binom{n+w+t}{n}. \end{aligned} \tag{81}$$

because of (12). So, if $\hat{\mathcal{A}}$ is a constrained weight $w \in [w-(m-1), w]$ code and $d_{L_1}^{as}(\hat{\mathcal{A}}) > t$ then, from (79), (75) and (80),

$$|\hat{\mathcal{A}}| = \frac{|\mathcal{B}(\hat{\mathcal{A}},t)|}{|\mathcal{B}(\emptyset,t)|} \le \sum_{v=w-(m-1)}^{w+t} \binom{n}{v}_{m+\mu} \Bigg/ \sum_{\tau=0}^{t} \binom{n}{\tau}_{m}.$$

In particular, if $m = \infty$ then

$$|\hat{\mathcal{A}}| = \frac{|\mathcal{B}(\hat{\mathcal{A}},t)|}{|\mathcal{B}(\emptyset,t)|} \le \binom{n+w+t}{n} \Bigg/ \binom{n+t}{n} = \frac{|\mathcal{S}(\mathbb{N}, n+1, w+t)|}{|\mathcal{S}(\mathbb{N}, n+1, t)|}.$$

because of (79), (76) and (81). At this point, the upper bounds on $LW(\mathbb{Z}_m, n, w, t)$, for all $m \in \mathbb{N} \cup \{\infty\}$, follow because the last two inequalities are valid for any constrained weight $w \in [w-(m-1), w]$ code, $\hat{\mathcal{A}}$, such that $d_{L_1}^{as}(\hat{\mathcal{A}}) > t$. ■

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their helpful inputs which greatly improve the quality of the paper.

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 Trans. Inf. Theory*, vol. 58, no. 3, pp. 1804–1808, Mar. 2012.

[2] M. Abroshan, R. Venkataramanan, and A. G. I. Fàbregas, "Coding for segmented edit channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3086–3098, Apr. 2018.

[3] *Bertrand's Postulate*. Accessed: Sep. 4, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Bertrand's_postulate

[4] M. Blaum, *Codes for Detecting Correcting Unidirectional Errors*, IEEE Comput. Soc. Press, Washington, DC, USA, 1993.

[5] R. C. Bose and S. Chowla, "Theorems in the additive theory of numbers," *Commentarii Mathematici Helvetici*, vol. 37, no. 1, pp. 141–147, Dec. 1962.

[6] B. Bose and T. R. N. Rao, "Theory of undirectional error correcting/detecting codes," *IEEE Trans. Comput.*, vol. C-31, no. 6, pp. 521–530, Jun. 1982.

[7] S. D. Constantin and T. R. N. Rao, "On the theory of binary asymmetric error correcting codes," *Inf. Control*, vol. 40, no. 1, pp. 20–36, Jan. 1979.

[8] B. Bose, N. Elarief, and L. G. Tallini, "On codes achieving zero error capacities in limited magnitude error channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 257–273, Jan. 2018.

[9] *Cramér's Conjecture*. Accessed: Sep. 4, 2022. [Online]. Available: https://en.wikipedia.org/wiki/Cramér's_conjecture

[10] H. Derksen, "Error-correcting codes and B_h-sequences," *IEEE Trans. Inform. Theory*, vol. 50, no. 3, pp. 476–485, Mar. 2004.

[11] L. Dolecek and V. Anantharam, "Repetition error correcting sets: Explicit constructions and prefixing methods," *SIAM J. Discrete Math.*, vol. 23, no. 4, pp. 2120–2146, 2010.

[12] A. Fazeli, A. Vardy, and E. Yaakobi, "Generalized sphere packing bound," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2313–2334, May 2015.

[13] H. C. Ferreira, W. A. Clarke, A. S. J. Helberg, K. A. S. Abdel-Ghaffar, and A. J. H. Vinck, "Insertion/deletion correction with spectral nulls," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 722–732, Mar. 1997.

[14] V. Guruswami and J. Håstad, "Explicit two-deletion codes with redundancy matching the existential bound," in *Proc. 32nd Annu. ACM-SIAM Symp. Discrete Algorithms*, Jan. 2021, pp. 21–32.

[15] A. S. J. Helberg and H. C. Ferreira, "On multiple insertion/deletion correcting codes," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 305–308, Jan. 2002.

[16] M. Kovačević and V. Y. F. Tan, "Asymptotically optimal codes correcting fixed-length duplication errors in DNA storage systems," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2194–2197, Nov. 2018.

[17] M. Kovačević, "Runlength-limited sequences and shift-correcting codes: Asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 65, no. 8, pp. 4804–4814, Aug. 2019.

[18] T. Kløve, "Error correction codes for the asymmetric channel," Rep., Dept. Informat., Univ. Bergen, Bergen, Norway, Tech. Rep., 1981.

[19] A. A. Kulkarni, "Insertion and deletion errors with a forbidden symbol," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2014, pp. 596–600.

[20] A. A. Kulkarni and N. Kiyavash, "Nonasymptotic upper bounds for deletion correcting codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 5115–5130, Aug. 2013.

[21] V. I. Levenshtein, "Binary codes with correction for deletions and insertions of the symbol 1," (in Russian), *Problems Peredachi Inf.*, vol. 1, no. 1, pp. 12–25, 1965.

[22] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions and reversals," *Sov. Phys. Dokl.*, vol. 10, pp. 707–710, Feb. 1966.

[23] V. I. Levenshtein and A. J. H. Vinck, "Perfect (d,k)-codes capable of correcting single peak-shifts," *IEEE Trans. Inform. Theory*, vol. 39, no. 2, pp. 656–662, Mar. 1993.

[24] Y. Liron and M. Langberg, "A characterization of the number of subsequences obtained via the deletion channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2300–2312, May 2015.

[25] H. Mahdavifar and A. Vardy, "Asymptotically optimal sticky-insertion-correcting codes with efficient encoding and decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2683–2687.
[26] 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 Trans. Inf. Theory*, vol. 58, no. 3, pp. 1809–1824, Mar. 2012.
[27] D. Pelusi, S. Elmougy, L. G. Tallini, and B. Bose, "m-ary balanced codes with parallel decoding," *IEEE Trans. Inform. Theory*, vol. 61, no. 6, pp. 3251–3264, Jun. 2015.
[28] L. Pezza, L. G. Tallini, and B. Bose, "Variable length unordered codes," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 548–569, Feb. 2012.
[29] *Prime Gap*. Accessed: Sep. 4, 2022. [Online]. Available: https://en.wikipedia.org/wiki/prime_gap
[30] F. Sellers, "Bit loss and gain correction code," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 35–38, Jan. 1962.
[31] J. Sima and J. Bruck, "Optimal k-deletion correcting codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 847–851.
[32] N. J. A. Sloane, "On single-deletion-correcting codes," in *Codes Designs*, R.-C. Festschrift, Ed. Ohio State Univ., 2000, pp. 273–291. [Online]. Available: https://arxiv.org/abs/math/0207197
[33] L. G. Tallini and B. Bose, "Design of balanced and constant weight codes for VLSI systems," *IEEE Trans. Comput.*, vol. CI-47, no. 5, pp. 556–572, May 1998.
[34] L. G. Tallini and U. Vaccaro, "Efficient m-ary balanced codes," *Discrete Appl. Math.*, vol. 92, no. 1, pp. 17–56, Mar. 1999.
[35] L. G. Tallini and B. Bose, "On a new class of error control codes and symmetric functions," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 980–984.
[36] L. G. Tallini and B. Bose, "On decoding some error control codes using the elementary symmetric functions," in *Trends Incidence Galois Geometries: A Tribute to Giuseppe Tallini—Quaderni di Matematica*, vol. 19, F. Mazzocca, N. Melone and D. Olanda, Eds. Caserta, Italy: Dipartimento di Matematica, Seconda Università di Napoli, 2010, pp. 265–297.
[37] L. G. Tallini, N. Elarief, and B. Bose, "On efficient repetition error correcting codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2010, pp. 1012–1016.
[38] L. G. Tallini and B. Bose, "On L_1-distance error control codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul./Aug. 2011, pp. 1026–1030.
[39] L. G. Tallini and B. Bose, "On symmetric L_1 distance error control codes and elementary symmetric functions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2012, pp. 741–745.
[40] L. G. Tallini and B. Bose, "On L_1 metric asymmetric/unidirectional error control codes, constrained weight codes and σ-codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2013, pp. 694–698.
[41] L. G. Tallini and B. Bose, "On some new Z_m linear codes based on elementary symmetric functions," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2018, pp. 1665–1669.
[42] L. G. Tallini, N. Alqwaifly, and B. Bose, "Efficient systematic deletions/insertions of 0's error control codes," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2022.
[43] J. H. Weber, C. de Vroedt, and D. E. Boekee, "Necessary and sufficient conditions on block codes correcting/detecting errors of various types," *IEEE Trans. Comput.*, vol. 41, no. 9, pp. 1189–1193, Sep. 1992.

Luca G. Tallini was born in Frascati, Rome, Italy. He received the Laurea degree *(magna cum laude)* in mathematics from the University of Rome "La Sapienza" in 1991, and the M.S. and Ph.D. degrees in computer science from Oregon State University, Corvallis OR, USA, in 1994 and 1996, respectively. From June 1997 to June 1998, he had been with the Dipartimento di Informatica ed Applicazioni, University of Salerno, Italy. From June 1998 to December 2002, he had been an Assistant Professor at the Politecnico di Milano, Italy. Since 2003, he has been with the University of Teramo, Italy, where he is currently a Professor at the Faculty of Communication Sciences. His research interests include coding theory, information theory, combinatorics, combinatorial algorithms, combinatorial geometry, fault-tolerant computing, parallel computing, neural networks, and VLSI.

Nawaf Alqwaifly received the B.S. degree in electrical engineering from Qassim University, Saudi Arabia, in 2008, the M.S. degree in electrical engineering from the University of Southern California in 2015, and the Ph.D. degree in electrical engineering from Oregon State University in 2021. He is currently an Assistant Professor with the Department of Electrical Engineering, Qassim University. His research interests include error control codes and information theory.

Bella Bose (Life Fellow, IEEE) received the B.E. degree in electrical engineering from Madras University, India, in 1973, the M.E. degree in electrical engineering from the Indian Institute of Science, Bengaluru, in 1975, and the M.S. and Ph.D. degrees in computer science and engineering from Southern Methodist University, Dallas, TX, USA, in 1979 and 1980, respectively. Since 1980, he has been with Oregon State University, where he is currently a Professor with the School of Electrical Engineering and Computer Science. His current research interests include error control codes, fault-tolerant computing, parallel processing, and computer networks. He is a fellow of both the ACM and the IEEE.