Location-Correcting Codes*

Ron M. Roth† Gadiel Seroussi‡

May 10, 2016

Abstract

We study codes over $GF(q)$ that can correct t channel errors assuming the error *values* are known. This is a counterpart to the well-known problem of erasure correction, where error values are found assuming the locations are known. The correction capabilities of these so called t-location correcting codes (t-LCCs) are characterized by a new metric, the *decomposability distance*, which plays a role analogous to that of the Hamming metric in conventional error-correcting codes (ECCs). Based on the new metric, we present bounds on the parameters of t-LCCs that are counterparts to the classical Singleton, sphere packing and Gilbert-Varshamov bounds for ECCs. In particular, we show examples of perfect LCCs, and we study optimal (MDS-like) LCCs that attain the Singleton-type bound on the redundancy. We show that these optimal codes are generally much shorter than their erasure (or conventional ECC) analogues: The length n of any t-LCC that attains the Singleton-type bound for $t > 1$ is bounded from above by $t + O(\sqrt{q})$, compared to length $q+1$ which is attainable in the conventional ECC case. We show constructions of optimal t-LCCs for $t \in \{1, 2, n-2, n-1, n\}$ that attain the asymptotic length upper bounds, and constructions for other values of t that are optimal, yet their lengths fall short of the upper bounds. The resulting asymptotic gap remains an open research problem. All the constructions presented can be efficiently decoded.

Index terms: location-correcting codes, error locations, error values, decomposability distance, bounds on code parameters, Sidon sets.

---

*A preliminary version of this work was presented in part at the *IEEE International Symposium on Information Theory,* Trondheim, Norway, June 1994.

†Computer Science Department, Technion, and Hewlett-Packard Israel Science Center, Technion City, Haifa 32000, Israel.

‡Hewlett-Packard Laboratories, 1501 Page Mill Road, Palo Alto, CA 94304, USA.

1 Introduction

Consider a code $\mathcal{C}$ of length n over $F = GF(q)$. We define the *dimension* of $\mathcal{C}$ as $k = \log_q |\mathcal{C}|$ and the *redundancy* of $\mathcal{C}$ as $r = n-k$. For linear codes, these correspond to the usual algebraic definitions. We will use the notation (n, k) to denote general nonlinear codes, and $[n, k]$ to denote linear codes. The task of a conventional decoder for a t-error-correcting code $\mathcal{C}$ can be described as follows: given a received word $\mathbf{y} \in F^n$, find $\tau \leq t$ *error locations* $1 \leq i_1, i_2, \ldots, i_\tau \leq n$ and τ nonzero *error values* $E_1, E_2, \ldots, E_\tau \in F$ such that $\mathbf{y} - \mathbf{e} \in \mathcal{C}$, where $\mathbf{e} = [\, e_1 \; e_2 \; \ldots \; e_n \,]$ is a vector whose nonzero entries are given by $e_{i_\ell} = E_\ell$, $\ell = 1, 2, \ldots, \tau$. In the conventional model, it is assumed that the encoder has no a priori information about the locations or values of the errors and that decoding is carried out by using no input other than the received word $\mathbf{y}$. Recently, a model of *localized errors* has been studied by several authors: Ahlswede, Bassalygo, and Pinsker [1], [2], Bassalygo, Gelfand, and Pinsker [5], [6], [7], and Kabatyansky [16]. In this model, the encoder (but not the decoder) does have knowledge of the locations i_ℓ of the errors that might occur, thus allowing the encoder to adapt the codeword list accordingly. On the other hand, in the pure *erasure correction* problem, the locations i_ℓ are assumed to be known only at the decoding side, and the task of the decoder is to determine the values $e_{i_\ell} = E_\ell$. (The remaining case where both the encoder and the decoder know the error locations is by far easier.)

In this work, we consider a converse of the erasure correction problem in the following sense: we assume the decoder knows the error *values*, while trying to determine the unknown error locations. More specifically, the input to the decoder is a received word $\mathbf{y}$ and a multiset consisting of τ nonzero values $E_1, E_2, \ldots, E_\tau \in F$. The task of the decoder is to find τ distinct locations $1 \leq i_1, i_2, \ldots, i_\tau \leq n$ such that, as before, $\mathbf{y} - \mathbf{e} \in \mathcal{C}$, and the nonzero entries of $\mathbf{e}$ are given by $e_{i_\ell} = E_\ell$.

A code $\mathcal{C}$ capable of correcting all patterns of t errors or less given their values is called a *t-location-correcting code* (in short, t-LCC). In the sequel, we always assume that $0 < t \leq n$.

A model where the error values are known may arise in applications where the decoder has side information about the possible values of the errors through, say, monitoring of the channel. This information, on the other hand, is not available to the encoder. The study of such an error model has a theoretical motivation as well: A conventional t-error-correcting code (t-ECC) must have enough redundancy r to allow reconstruction of both the error locations and the error values. Hence, we might think of the information coded in the r redundancy symbols as carrying 'error-location information' and 'error-value information.' When correcting erasures, only the error-value information is required, whereas in the error model investigated here we need only the error-location information.

The question is: how does the smallest possible r for a t-ECC relate to the *sum* of the redundancies of a t-LCC and a t-erasure-correcting code — all of the same length?

We point out that, in the binary case, the problem of location correction is equivalent to that of error correction provided that the number τ of errors that have actually occurred is known to the decoder. Assuming that τ is at most a prescribed number t, it is easy to verify that a necessary and sufficient condition for a binary code $\mathcal{C}$ of length n to be a t-LCC is that the *even* Hamming distances between distinct codewords in $\mathcal{C}$ all be greater than $2t$. It follows that the codewords of even Hamming weight in $\mathcal{C}$ (and, respectively, the codewords of odd weight) must form a code of minimum Hamming distance $\geq 2t+2$. Hence, a binary t-LCC of length n with a maximal number of codewords can always be obtained from a binary t-ECC of length $n-1$ as follows: for each codeword $\mathbf{c}$ of the latter, we introduce two codewords, $[\,0\ \mathbf{c}\,]$ and $[\,1\ \mathbf{c}\,]$, in the former. Therefore, the problem of location correction is of interest mostly when the alphabet size is greater than 2. In fact, most of the results in this paper refer to values of q that increase with n.

We also mention an early work by Wolf and Elspas [23], who considered the binary case, but with a somewhat different model: The codewords there are assumed to be divided into non-overlapping sub-blocks of size m, and the decoder needs to identify the erroneous sub-blocks — rather than find the exact error locations — given that the number of binary errors is bounded from above by a prescribed number t. For related work see, for instance, also [13] and [22].

The rest of the paper is organized as follows: In Section 2, we define a metric that plays, for LCCs, a role analogous to that of the Hamming metric for ECCs. In Section 3 we present bounds on the redundancy and length of LCCs, which apply to the general class of nonlinear codes. In particular, we present counterparts to the classical Singleton, sphere packing, and Gilbert-Varshamov bounds of conventional ECCs. Using the sphere packing bound, we show that the length of any t-LCC that attains the Singleton-type bound for $t > 1$ is bounded from above by $t + O(\sqrt{q})$. Therefore, when $1 < t \ll q$, the length of LCCs that are optimal in the Singleton bound sense is much smaller than $q+1$ — a length which is attainable by conventional MDS codes. Section 3 also contains examples of nontrivial LCCs that are both optimal (in the sense of attaining the Singleton-type bound) and perfect (in the sense of attaining the sphere packing bound). In Section 4, we present constructions of t-LCCs for various values of t. The emphasis here is on constructing the longest possible codes that are optimal in the Singleton bound sense, as a function of increasing values of q. All the constructed codes can be efficiently decoded. A table summarizing the various bounds and constructions is given in Section 5.

2 Decomposability distance

In subsequent results we will make use of the following definitions.

A vector $\mathbf{c} \in F^n$ is called *τ-decomposable* if there exist two vectors $\mathbf{x}, \mathbf{y} \in F^n$, both containing the same multiset of at most τ nonzero entries, such that $\mathbf{c} = \mathbf{x} - \mathbf{y}$.

For $i = 1, 2, \ldots, n$, let $\mathbf{u}_i \in F^n$ be the unit vector having 1 in its ith coordinate. For every $i, j \in \{1, 2, \ldots, n\}$, $i \neq j$, we define the vector $\mathbf{u}_{i,j}$ to be the difference $\mathbf{u}_i - \mathbf{u}_j$. The set of all vectors $\mathbf{u}_{i,j} \in F^n$ will be denoted by $\mathcal{W}_F(n)$.

The following lemma establishes an equivalent definition of τ-decomposability.

Lemma 1. *Let $\mathbf{c} \in F^n$. Then, $\mathbf{c}$ is τ-decomposable if and only if it can be written as a linear combination of τ elements of $\mathcal{W}_F(n)$; namely, there exist τ elements $E_\ell \in F$ and respective vectors $\mathbf{u}_{i_\ell,j_\ell} \in \mathcal{W}_F(n)$ such that*

$$\mathbf{c} \;=\; \sum_{\ell=1}^{\tau} E_\ell \, \mathbf{u}_{i_\ell,j_\ell} \tag{1}$$

(an empty sum is defined as zero).

Proof. The "only if" part follows from the definition of τ-decomposability. For the "if" part we show how to transform the $\mathcal{W}_F(n)$-expansion (1) of $\mathbf{c}$ into one where the indexes i_ℓ are all distinct, and so are the indexes j_ℓ. This is accomplished by performing a Gaussian elimination process for $m = 1, 2, \ldots, \tau$, as follows: Suppose that $i_\ell = i_m$ for some index $\ell > m$. Noting that

$$E_m \mathbf{u}_{i_m,j_m} \;+\; E_\ell \mathbf{u}_{i_\ell,j_\ell} \;=\; (E_m + E_\ell)\mathbf{u}_{i_m,j_m} \;+\; E_\ell \mathbf{u}_{j_m,j_\ell} \;,$$

we can use $\mathbf{u}_{i_m,j_m}$ as a pivot element and eliminate all vectors $\mathbf{u}_{i_\ell,j_\ell}$ with $\ell > m$ and $i_\ell = i_m$ from the expansion (1) of $\mathbf{c}$ without increasing the number of vectors $\mathbf{u}_{i,j}$ in the expansion. Similarly, we can also eliminate all vectors $\mathbf{u}_{i_\ell,j_\ell} = -\mathbf{u}_{j_\ell,i_\ell}$ with $\ell > m$ and $j_\ell = i_m$. Having done so, we check for $\ell > m$ whether any of the indexes i_ℓ or j_ℓ is equal to j_m. If there exists such an index, we re-order the vectors $\mathbf{u}_{i_\ell,j_\ell}$, $\ell > m$, possibly with a sign change $\mathbf{u}_{i_\ell,j_\ell} = -\mathbf{u}_{j_\ell,i_\ell}$, so that $i_{m+1} = j_m$. We now continue with the pivot element $\mathbf{u}_{i_{m+1},j_{m+1}}$ and iterate this procedure until we reach the last pivot element $\mathbf{u}_{i_\tau,j_\tau}$. Once we have an expansion (1) with distinct indexes i_ℓ and distinct indexes j_ℓ, we construct $\mathbf{x} = \sum_{\ell=1}^{\tau} E_\ell \mathbf{u}_{i_\ell}$ and $\mathbf{y} = \sum_{\ell=1}^{\tau} E_\ell \mathbf{u}_{j_\ell}$, which satisfy the condition of the lemma. Notice that in the process of Gaussian elimination, some of the coefficients E_ℓ might vanish,

so the actual number of nonzero entries in each of the vectors $\mathbf{x}$ and $\mathbf{y}$ might be less than τ. □

Clearly, the entries of every τ-decomposable vector in F^n must sum to zero. On the other hand, by Lemma 1 it can be readily verified that if the entries of a vector sum to zero, then the vector is $(n-1)$-decomposable. Therefore, we can assume that $\tau \leq n-1$.

We define the *decomposability weight* of $\mathbf{c} \in F^n$ as the smallest nonnegative integer τ, if any, such that $\mathbf{c}$ is τ-decomposable. If no such τ exists, we define the decomposability weight of $\mathbf{c}$ as infinity. The *decomposability distance* between two vectors $\mathbf{c}$ and $\mathbf{c}'$ in F^n is defined as the decomposability weight of $\mathbf{c}-\mathbf{c}'$. Note that this distance is finite (and, thus, less than n) if and only if the sum of entries of $\mathbf{c}$ equals the sum of entries of $\mathbf{c}'$. It can be easily verified that decomposability distance is a metric. We denote by $\mathcal{B}_F(n; 2w)$ the set of all vectors in F^n of decomposability weight equaling at most w. Thus, in the decomposability metric, $\mathcal{B}_F(n; 2w)$ is a sphere (including the interior) of radius w centered at $\mathbf{0}$. We also define $\mathcal{B}_F(n; 2w+1)$ as the set of all vectors in F^n of the form $\mathbf{c} + \mathbf{u}_i$, where $\mathbf{c} \in \mathcal{B}_F(n; 2w)$ and $\mathbf{u}_i$ is a unit vector in F^n. The set $\mathcal{B}_F(n; 2w+1)$ is a union of n spheres of radius w, each centered at a distinct unit vector $\mathbf{u}_i$ of F^n.

The *minimum decomposability distance* (in short, MDD) of a code $\mathcal{C} \subseteq F^n$ is the smallest among all decomposability distances between any two distinct elements of $\mathcal{C}$. If $\mathcal{C}$ is linear, then its MDD is equal to the minimum decomposability weight of any nonzero codeword of $\mathcal{C}$.

Theorem 1. *A code over F is a t-LCC if and only if its MDD is at least $t+1$.*

Proof. This follows from the fact that two distinct vectors $\mathbf{c}, \mathbf{c}' \in F^n$ are at decomposability distance at most t if and only if $\mathbf{c} + \mathbf{e} = \mathbf{c}' + \mathbf{e}'$, where $\mathbf{e}$ and $\mathbf{e}'$ have the same multiset of t nonzero entries. □

In particular, a code is an $(n-1)$-LCC if and only if its MDD is at least n, namely, infinity.

Using the result of Lemma 1, one can observe that a t-LCC can also correct t errors under a more general model where multiple errors in the same location are allowed, i.e., we do not require that the error locations i_ℓ be distinct, and the nonzero entries of the error vector $\mathbf{e}$ are given by $e_i = \sum_{\ell \,:\, i_\ell = i} E_\ell$. Hence, these two error models are equivalent from a code construction point of view. In the sequel, we consider only the conceptually simpler model with a one-to-one correspondence between the given error values and the error locations.

The next lemma presents another equivalent definition of τ-decomposability, which will be used in the appendix to prove that computing the decomposability weight of a vector is an intractable (NP-complete) problem [12]. Still, this is not necessarily an impediment to the construction of codes with efficient decoding algorithms, as we show in the sequel.

Lemma 2. *Let $\mathbf{c} = [\, c_1 \; c_2 \; \ldots \; c_n \,] \in F^n$ and let $\tau \leq n-1$. Then, $\mathbf{c}$ is τ-decomposable if and only if there is a partition of $\{1, 2, \ldots, n\}$ into $n-\tau$ nonempty disjoint subsets S_k, $k = 1, 2, \ldots, n-\tau$, such that*

$$\sum_{i \in S_k} c_i \;=\; 0 \quad \textit{for} \quad k = 1, 2, \ldots, n-\tau \; .$$

Proof. Suppose that such a partition exists and for $k = 1, 2, \ldots, n-\tau$, let $\mathbf{c}_k = [\, c_{k,1} \; c_{k,2} \; \ldots \; c_{k,n} \,]$ be the vector defined by

$$c_{k,i} \;=\; \begin{cases} c_i & \text{if } i \in S_k \\ 0 & \text{otherwise} \end{cases} \quad .$$

There are at most $|S_k|$ nonzero entries in $\mathbf{c}_k$ and the sum of those entries is zero. Hence, the vector $\mathbf{c}_k$ is $(|S_k|-1)$-decomposable (this applies also to the case where $|S_k| = 1$, in which case $\mathbf{c}_k = \mathbf{0}$). Noting that $\mathbf{c} = \sum_{k=1}^{n-\tau} \mathbf{c}_k$, it thus follows that $\mathbf{c}$ is η-decomposable, where $\eta = \sum_{k=1}^{n-\tau} (|S_k| - 1) = n - (n - \tau) = \tau$. This concludes the "if" part.

As for the other direction, let $\mathbf{c} = [\, c_i \,]_{i=1}^n$ be a τ-decomposable vector and let

$$\mathbf{c} \;=\; \sum_{\ell=1}^{\tau} E_\ell \, \mathbf{u}_{i_\ell, j_\ell} \tag{2}$$

be a decomposition of $\mathbf{c}$ with distinct values i_ℓ and distinct values j_ℓ, as produced by the Gaussian elimination process described in the proof of Lemma 1. We prove the result by induction on τ. For $\tau = 0$ we have $\mathbf{c} = \mathbf{0}$ and the singletons $S_k = \{k\}$, $k = 1, 2, \ldots, n$, yield the desired partition.

Assume now that the result is valid for any $\tau' < \tau$. Let I and J denote the sets $\{\, i_\ell \,\}_{\ell=1}^{\tau}$ and $\{\, j_\ell \,\}_{\ell=1}^{\tau}$, respectively, where i_ℓ and j_ℓ are as in (2). If $I = J$, then the τ vectors $\mathbf{u}_{i_\ell, j_\ell}$ are linearly dependent (their sum is zero), which implies that $\mathbf{c}$ is in effect $(\tau-1)$-decomposable, and the desired result follows from the induction hypothesis (a partition into $n-\tau+1$ subsets is trivially converted into one with $n-\tau$ subsets by joining two subsets). Hence, we assume $I \neq J$ and, in particular, we can assume that $I - J$ contains the element i_1. If $j_1 \in I$, then, by renaming of indexes we can assume that

$j_1 = i_2$. We repeat this process with j_2 and continue in this fashion until we reach an index $m \geq 1$ such that $j_\ell = i_{\ell+1}$ for $\ell = 1, 2, \ldots m{-}1$ and $j_m \notin I$. We then define

$$S_1 \;=\; \{\, i_\ell \,\}_{\ell=1}^{m} \cup \{\, j_m \,\} \;,$$

satisfying

$$\sum_{i \in S_1} c_i \;=\; E_1 \;+\; \Big(\sum_{\ell=2}^{m} (E_\ell - E_{\ell-1})\Big) \;-\; E_m \;=\; 0 \;,$$

as desired. Let $\overline{S}_1$ denote the set $\{1, 2, \ldots, n\} - S_1$ and consider the vector $\mathbf{c}' = [\, c'_i \,]_{i=1}^{n}$ which is given by

$$c'_i \;=\; \begin{cases} c_i & \text{if } i \in \overline{S}_1 \\ 0 & \text{otherwise} \end{cases} \;.$$

We have $\mathbf{c}' = \sum_{\ell=m+1}^{\tau} E_\ell \, \mathbf{u}_{i_\ell, j_\ell}$, implying that $\mathbf{c}'$ is $(\tau{-}m)$-decomposable. Furthermore, since $i_\ell, j_\ell \notin S_1$ for any $\ell > m$, it follows that the subvector $\mathbf{c}'' \in F^{n-m-1}$ of $\mathbf{c}'$ which is indexed by $\overline{S}_1$ is also $(\tau{-}m)$-decomposable. Applying the induction hypothesis on $\mathbf{c}''$, it follows that there exists a partition of $\overline{S}_1$ into $n{-}\tau{-}1$ subsets $S_2, S_3, \ldots, S_{n-\tau}$ which, together with S_1, form the desired partition of $\{1, 2, \ldots, n\}$. □

For a code $\mathcal{C} \subseteq F^n$ and an element $a \in F$, denote by $\mathcal{C}(a)$ the set of all codewords $\mathbf{c} = [\, c_i \,]_{i=1}^{n} \in \mathcal{C}$ such that $\sum_{i=1}^{n} c_i = a$. Clearly, the sets $\mathcal{C}(a)$ (referred to as *classes*) form a partition of $\mathcal{C}$ and, so,

$$|\mathcal{C}| \;=\; \sum_{a \in F} |\mathcal{C}(a)| \;.$$

Furthermore, the decomposability distance between two codewords that belong to distinct classes $\mathcal{C}(a)$ is infinity. Hence, the MDD of $\mathcal{C}$ equals the smallest among the MDDs of any of the classes $\mathcal{C}(a)$.

For any vector $\mathbf{v} \in F^n$ and subset $X \subseteq F^n$, denote by $\mathbf{v} + X$ the set $\{\mathbf{v} + \mathbf{x} \mid \mathbf{x} \in X\}$. Now, let $\mathcal{C}$ be a t-LCC over $F = GF(q)$ and let $\mathcal{C}(b)$ denote the largest class in $\mathcal{C}$. Consider the code $\mathcal{C}'$ given by

$$\mathcal{C}' \;=\; \bigcup_{a \in F} \left((a - b)\mathbf{u}_1 + \mathcal{C}(b)\right) \;.$$

We have $\mathcal{C}'(a) = (a - b)\mathbf{u}_1 + \mathcal{C}(b)$; namely, all classes of $\mathcal{C}'$ have the same size and the same MDD as $\mathcal{C}(b)$. Hence,

$$|\mathcal{C}'| \;=\; q \cdot |\mathcal{C}(b)| \;\geq\; |\mathcal{C}| \;,$$

and the MDD of $\mathcal{C}'$ is at least as large as that of $\mathcal{C}$. As we are interested in codes $\mathcal{C}$ which are as large as possible for any given length and MDD, we can assume, when appropriate, that $|\mathcal{C}| = q \cdot |\mathcal{C}(0)|$ and that

$$\mathcal{C} \;=\; \bigcup_{a \in F} \left(a\mathbf{u}_1 + \mathcal{C}(0)\right) \;.$$

Notice that the minimum Hamming distance of a code of this form is 1. In the special case of linear codes, the classes $\mathcal{C}(a)$ become cosets of the linear subcode $\mathcal{C}(0)$ in $\mathcal{C}$. We can assume that the generator matrix of $\mathcal{C}$ contains the row $\mathbf{u}_1$ and that the parity-check matrix has a zero column.

Example 1. Consider the $[4, 2]$ code $\mathcal{C}_5$ over $GF(5)$ defined by the parity-check matrix

$$H_5 = \begin{bmatrix} 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 2 \end{bmatrix}.$$

If we take the six unordered pairs of columns of H_5 and compute the difference of the elements in each pair, we obtain the six columns of a parity-check matrix of a $[6, 4]$ 1-error correcting Hamming code over $GF(5)$. Hence, no such difference is a scalar multiple of the other and, so, $\mathcal{C}_5$ cannot contain any 2-decomposable codewords. Therefore, the MDD of $\mathcal{C}_5$ is at least 3 and, as such, it is a 2-LCC. The subcode $\mathcal{C}_5(0)$ is a $[4, 1]$ code which is spanned by the codeword $[\,1\ 1\ 1\ 2\,]$. It is not difficult to verify that this codeword is 3-decomposable. Hence, the MDD of $\mathcal{C}_5$ is exactly 3.

We end this section by stating the following simple relationship between MDD and minimum Hamming distance.

Lemma 3. *Let $\mathcal{C}$ be a code over F and let $d(a)$ denote the minimum Hamming distance of $\mathcal{C}(a)$. Then the MDD of $\mathcal{C}$ is bounded from below by $\frac{1}{2}\min_{a\in F} d(a)$. In particular, the MDD of $\mathcal{C}$ is at least half the minimum Hamming distance of $\mathcal{C}$.*

Proof. This follows from the fact that the Hamming weight of any τ-decomposable vector is at most 2τ. □

The bound of Lemma 3 is not tight, as exhibited by the code $\mathcal{C}_5$ of Example 1: the minimum Hamming distance of every class $\mathcal{C}_5(a)$ is 4, while the MDD is 3. The following weaker converse does hold for fields $F = GF(q)$ with $q > 2$: for every code $\mathcal{C}$ whose minimum Hamming distance is $d > 1$, there exists an *equivalent* code (in the general sense of [17, Ch. 2]) whose MDD is $\lceil d/2 \rceil$. Such a code can be obtained through scaling each coordinate of $\mathcal{C}$ by an appropriate constant from F. For $q = 2$, the result applies to the minimum *even* Hamming distance, as noted in Section 1.

3 Bounds on the parameters of LCCs

3.1 Singleton-type bound

The well-known Singleton bound [17] for a conventional t-ECC states that the redundancy r of such a code must satisfy

$$r \geq 2t \,. \tag{3}$$

The respective bound for pure erasure correction is given by

$$r \geq t \,. \tag{4}$$

Codes that attain these bounds are known as *maximum-distance separable* (MDS) codes. Primary examples of MDS codes are Reed-Solomon codes and their extensions and generalizations, yielding constructions of length $q+1$ for the dimension range between 2 and q [17, Ch. 11].

Theorem 2. (Singleton-type bound for LCCs.) *Let $\mathcal{C}$ be a t-LCC of length n over $F = GF(q)$, with $1 \leq t \leq n-1$. Then the redundancy of $\mathcal{C}$ satisfies*

$$r \geq t \,.$$

Proof. We prove by contradiction. Assume $r < t$. Then, we have $|\mathcal{C}| > q^{n-t}$, and, for some $a \in F$, we must have $|\mathcal{C}(a)| > q^{n-t-1}$. Hence, there exist two codewords in $\mathcal{C}(a)$ whose suffixes of length $n-t-1$ (≥ 0) are identical. The decomposability distance between these two codewords is finite, and it is equal to the decomposability distance between their prefixes of length $t+1$. This distance is at most t, which, by Theorem 1, contradicts our assumption on $\mathcal{C}$. □

Codes that attain the bound of Theorem 2 will be referred to as *optimal LCCs*. They are analogous to conventional MDS codes. The code of Example 1 is an optimal $[4, 2]$ 2-LCC over $GF(5)$.

Example 2. The code $\mathcal{C}_{27}$ over $GF(27)$ is defined by the parity-check matrix

$$H_{27} = \begin{bmatrix} 0 & 1 & 0 & \alpha & \alpha^2 & \alpha^6 & \alpha^{11} & \alpha^{19} \\ 0 & 0 & 1 & \alpha & \alpha^{24} & \alpha^3 & \alpha^{21} & \alpha^{18} \end{bmatrix} ,$$

where α is a root of the ternary polynomial x^3+2x+1. Again, if we take the 28 unordered pairs of columns of H_{27} and compute the difference of the elements in each pair, we obtain the 28 columns of a parity-check matrix of a $[28, 26]$ 1-error correcting Hamming code over $GF(27)$. Therefore, $\mathcal{C}_{27}$ is an optimal 2-LCC.

We recall that by permuting the columns of a parity-check matrix of a given $[n, n-r]$ code we obtain an equivalent code with the same MDD and dimension. Note also that when we translate the columns of a parity-check matrix — namely, when we add the same vector to each column of the matrix — the MDD is preserved. Applying elementary operations on the rows of the matrix, we can assume, without loss of generality, that a parity-check matrix has a systematic form, containing a zero column followed by an $r \times r$ identity matrix. Such systematic matrices were presented in Examples 1 and 2.

We remark that the bound of Theorem 2 does not hold when $t = n$. In this case, a bound $r \geq t-1 = n-1$ is obtained by observing that, obviously, a t-LCC is also a $(t-1)$-LCC. Indeed, there exists an $[n, 1]$ n-LCC over $GF(q)$ for any n and q, as we show in Section 4.

A close analysis of the proof of Theorem 2 reveals another link between optimal LCCs and MDS codes, as shown in the following corollary.

Corollary 1. *Let $\mathcal{C}$ be an optimal t-LCC of length n. Then, each class $\mathcal{C}(a)$ of $\mathcal{C}$ is a conventional $(n, n-t-1)$ MDS code.*

Proof. It follows from the argument in the proof of Theorem 2 that for all $a \in F$, the suffixes of length $\ell = n-t-1$ of codewords in $\mathcal{C}(a)$ must all be distinct. This implies that $|\mathcal{C}(a)|$ must equal q^ℓ and that the suffixes exhaust all q^ℓ different ℓ-tuples over $GF(q)$. Now, the same must apply to the projection of the codewords into any subset of ℓ coordinates. This property characterizes an (n, ℓ) MDS code [17, Ch. 11]. □

We point out that the converse of Corollary 1 is not true: an $(n, n-t-1)$ MDS code has minimum Hamming distance $t+2$ and, as such, it can have an MDD which is as small as $\lceil t/2 \rceil + 1$ (see Lemma 3 and its ensuing discussion). Such a code cannot be a class $\mathcal{C}(a)$ of an optimal LCC when $t \geq 2$.

3.2 Sphere packing bound

Our next result is a sphere packing bound for t-LCCs (with spheres defined by the decomposability metric).

Theorem 3. (Sphere packing bound for LCCs.) *Let $\mathcal{C}$ be a t-LCC of length n over $F = GF(q)$. Then the redundancy of $\mathcal{C}$ satisfies*

$$r \;\geq\; \log_q |\mathcal{B}_F(n; t)| \; .$$

Proof. We first note that when t is even, the spheres $\mathbf{c}+\mathcal{B}_F(n;t)$ must be disjoint for distinct codewords $\mathbf{c} \in \mathcal{C}$, or else we would have two codewords in $\mathcal{C}$ at decomposability distance t or less. We observe that this applies also when t is odd, in which case $\mathcal{B}_F(n;t)$ is a union of n spheres, and $(\mathbf{c}+\mathcal{B}_F(n;t)) \cap (\mathbf{c}'+\mathcal{B}_F(n;t)) \neq \emptyset$ implies the existence of $\mathbf{e}, \mathbf{e}' \in \mathcal{B}_F(n;t-1)$ and $\mathbf{u}, \mathbf{u}' \in \mathcal{W}_F(n)$ such that

$$\mathbf{c} + \mathbf{e} + \mathbf{u} \;=\; \mathbf{c}' + \mathbf{e}' + \mathbf{u}' \;,$$

namely $\mathbf{c} - \mathbf{c}' \in \mathcal{B}_F(n;2t)$. Adding the volumes of the disjoint sets $\mathbf{c} + \mathcal{B}_F(n;t)$, we must have

$$|\mathcal{C}| \cdot |\mathcal{B}_F(n;t)| \;\leq\; q^n \;,$$

which implies the claim of the theorem. □

LCCs that attain the bound of Theorem 3 will be called *perfect LCCs*.

We examine now some specific values of t. For $t = 1$, the size of $|\mathcal{B}_F(n,1)|$ equals n, yielding by Theorem 3 an upper bound

$$n \;\leq\; q^r \tag{5}$$

on the length n of any $(n, n-r)$ 1-LCC over $GF(q)$. As we show in Section 4 (Construction 1), this bound can be attained for every q and r.

The exact value of $|\mathcal{B}_F(n,2)|$ is $\binom{n}{2}(q-1)+1$, yielding the upper bound

$$n(n-1) \;\leq\; 2 \cdot \frac{q^r - 1}{q - 1} \tag{6}$$

on the length n of any $(n, n-r)$ 2-LCC over $GF(q)$. Indeed, let $H = [\,\mathbf{h}_1\ \mathbf{h}_2\ \ldots\ \mathbf{h}_n\,]$ be an $r \times n$ parity-check matrix of a linear code $\mathcal{C}$ over $GF(q)$. Then $\mathcal{C}$ is a 2-LCC if and only if the differences $\mathbf{h}_i - \mathbf{h}_j$ of all $n(n-1)/2$ pairs of columns in H with indexes $i < j$ form a matrix H' in which no column is a scalar multiple of the other, thus yielding (6). In particular, the code $\mathcal{C}$ is a perfect 2-LCC if and only if H' is a parity-check matrix of a Hamming code of length $(q^r-1)/(q-1)$ over $GF(q)$ (see Examples 1 and 2).

For the special case of optimal 2-LCCs (namely, when $r = t = 2$), the bound (6) becomes

$$n(n-1) \;\leq\; 2(q+1) \;, \tag{7}$$

which is equivalent to

$$(n+1)(n-2) \;\leq\; 2q \;. \tag{8}$$

However, since q is a power of a prime and $\gcd(n+1, n-2) = \gcd(n+1, 3)$, equality in (8) can be attained only in the following cases:

(i) $n = 3$ and $q = 2$, in which case there exists a $[3, 1]$ 2-LCC, as we show in Section 4 (Construction 2).

(ii) $n = 4$ and $q = 5$, in which case we have the perfect 2-LCC $\mathcal{C}_5$ presented in Example 1.

(iii) $n = 5$ and $q = 9$; a simple exhaustive search has revealed that there is no *linear* $[5, 3]$ 2-LCC over $GF(9)$. Yet, there exists a nonlinear perfect $(5, 3)$ 2-LCC over $GF(9)$ which is described in Example 3 below.

(iv) $n = 8$ and $q = 27$, in which case we have the linear perfect 2-LCC $\mathcal{C}_{27}$ of Example 2.

In all other cases, we can tighten (7) to

$$n(n-1) \;\leq\; 2q\,.$$

Again, since $\gcd(n, n-1) = 1$, equality can be attained only when $n = q = 3$, in which case Construction 2 in Section 4 yields an optimal code. We can therefore summarize as follows.

Corollary 2. *The length n of every optimal 2-LCC over $GF(q)$ satisfies the inequality*

$$n(n-1) \;\leq\; \begin{cases} 2(q+1) & \text{for } q \in \{\,2,\,5,\,9,\,27\,\}\,, \\ 2q & \text{for } q = 3\,, \\ 2(q-1) & \text{otherwise.} \end{cases}$$

An $(n, n-2)$ 2-LCC that satisfies the equality $n(n-1) = 2(q-1)$ will be called *semi-perfect*. Note that the stricter inequality $n(n-1) \leq 2(q-1)$ holds also for linear codes over $GF(9)$.

Example 3. We define a perfect $(5, 3)$ 2-LCC $\mathcal{C}_9$ over $GF(9)$ as follows. Let α be an element of $GF(9)-GF(3)$, and consider the matrix

$$G \;=\; \begin{bmatrix} \alpha & 0 & \alpha{+}1 & 1 & \alpha{+}1 \\ 1 & 0 & 1 & 2\alpha{+}2 & \alpha{+}2 \\ 0 & \alpha & \alpha & 2\alpha{+}2 & 2\alpha{+}1 \\ 0 & 1 & 2\alpha{+}1 & 2\alpha{+}1 & 2\alpha \end{bmatrix}\,.$$

The class $\mathcal{C}_9(0)$ of $\mathcal{C}_9$ is the set of all linear combinations, *with coefficients in* $GF(3)$, of rows of G. The other classes of $\mathcal{C}_9$ are given by $\mathcal{C}_9(a) = \mathcal{C}_9(0) + a\mathbf{u}_1$, $a \in GF(9)$. Although $\mathcal{C}_9$ is nonlinear over $GF(9)$, it is a vector space over $GF(3)$. Hence, in order to show that

$\mathcal{C}_9$ is a 2-LCC, it suffices to check that none of the 80 nonzero codewords in $\mathcal{C}_9(0)$ is 2-decomposable. A simple enumeration reveals that this is indeed the case. Expanding each entry $f\alpha+g$ of G into its components $f, g \in GF(3)$, so that each column of G is split into two columns over $GF(3)$, we obtain a generator matrix $\hat{G}$ for a $[10, 4]$ ternary code. This code turns out to be a shortened extended ternary Golay code. Indeed, we observe that the columns of $\hat{G}$ correspond, up to permutation and scaling, to the nonzero columns of the first four rows of the generator matrix of the extended ternary Golay code $\mathcal{G}_{12}$, as described in [17, p. 510]. A computer search, based on the strict constraints imposed by Corollary 1, revealed that a code with the parameters of $\mathcal{C}_9$ is essentially unique.

Next we provide an estimate on the size of $\mathcal{B}_F(n;t)$ for general t, which will allow us to analyze the bound of Theorem 3.

Lemma 4. *For $F = GF(q)$,*

$$\sum_{\tau=0}^{w} \binom{n}{2\tau} (q-1)^\tau \;\le\; |\mathcal{B}_F(n;2w)| \;\le\; \sum_{\tau=0}^{w} \binom{n(n-1)/2}{\tau} (q-1)^\tau \;.$$

and

$$\sum_{\tau=0}^{w} \binom{n}{2\tau+1} (q-1)^\tau \;\le\; |\mathcal{B}_F(n;2w+1)| \;\le\; n \cdot \sum_{\tau=0}^{w} \binom{n(n-1)/2}{\tau} (q-1)^\tau \;.$$

Proof. The upper bounds follow from the fact that, by definition, each vector in $\mathcal{B}_F(n;2w)$ can be written as a linear combination of w elements of $\mathcal{W}_F(n)$.

As for the lower bounds, consider a vector $\mathbf{x}$ of Hamming weight $2\tau \le 2w$ whose ith nonzero entry is the additive inverse of its $(i+\tau)$th nonzero entry for $i = 1, 2, \ldots, \tau$. Clearly, the vector $\mathbf{x}$ is an element of $\mathcal{B}_F(n;2w)$, and there are $\binom{n}{2\tau}(q-1)^\tau$ such vectors $\mathbf{x}$ in F^n. This yields the lower bound on $\mathcal{B}_F(n;2w)$. The lower bound on $\mathcal{B}_F(n;2w+1)$ is obtained in a similar manner by considering vectors of Hamming weight $2\tau+1$ whose first 2τ nonzero entries satisfy the same property as before and whose last nonzero entry is 1. □

Corollary 3. *For every $(n, n-r)$ t-LCC over $F = GF(q)$,*

$$r \;\ge\; t \log_q n \;+\; \lfloor t/2 \rfloor \log_q(q-1) \;-\; t \log_q t \;.$$

Proof. This follows from Theorem 3, Lemma 4 and the inequality $\binom{n}{t} \geq (n/t)^t$. $\square$

We remark that it is possible to improve the lower bounds on $|\mathcal{B}_F(n;t)|$ of Lemma 4 by taking into account permutations of the 2τ nonzero entries of the vectors $\mathbf{x}$ considered in the proof of the lemma. This, in turn, will improve the bound of Corollary 3 when $t \ll q$ by an additive term of approximately $(t/2)\log_q t$, which is the most we can expect due to the upper bound in Lemma 4. However, for the application that follows, the bound of Corollary 3 is sufficient.

Consider the case $n = q{+}1$. We recall that the bounds (3) and (4) for conventional ECCs can be attained with equality for this length for any t in the appropriate range ($t \leq n/2$ for (3), $t \leq n$ for (4)). On the other hand, substituting $n = q{+}1$ in Corollary 3 yields

$$r \;\geq\; \left\lfloor \tfrac{3}{2}t \right\rfloor \;-\; t\log_q t\;.$$

It thus follows that when $2 \leq t \ll q$ and $n = q{+}1$, the sum of the redundancies of a t-LCC and a t-erasure-correcting code of length n must be strictly greater than the smallest attainable redundancy of a t-ECC of length n.

The following theorem will allow us to get a stronger bound than that of Corollary 3 in the special case of optimal t-LCCs.

Let $N(t,q)$ denote the maximal length, if any, of any optimal t-LCC over $GF(q)$.

Theorem 4. *For any $t \geq 2$,*

$$N(t,q) \;\leq\; N(t{-}1,q) \;+\; 1\;.$$

Proof. Let $\mathcal{C}$ be an optimal $(n, n{-}t)$ t-LCC over $F = GF(q)$ with $n = N(t,q)$. Define $\mathcal{C}'$ as the set of all vectors in F^{n-1} obtained by substituting the first two coordinates in each codeword of $\mathcal{C}$ with their sum. Each vector in $\mathcal{C}'$ corresponds to exactly one codeword in $\mathcal{C}$, since otherwise $\mathcal{C}$ would contain codewords at decomposability distance 1 from each other. Thus, we have $|\mathcal{C}'| = |\mathcal{C}| = q^{n-t}$. We claim that the MDD of $\mathcal{C}'$ is at least t. Suppose to the contrary that $\mathbf{c} = [\,c_1\;\; c_2\;\; \mathbf{d}\,]$ and $\mathbf{c}' = [\,c'_1\;\; c'_2\;\; \mathbf{d}'\,]$ are distinct codewords in $\mathcal{C}$ whose respective corresponding vectors $[\,(c_1{+}c_2)\;\; \mathbf{d}\,]$ and $[\,(c'_1{+}c'_2)\;\; \mathbf{d}'\,]$ in $\mathcal{C}'$ are at decomposability distance less than t. Then, we have

$$\begin{aligned} \mathbf{c} - \mathbf{c}' &= [\,(c_1{-}c'_1)\;\; (c_2{-}c'_2)\;\; (\mathbf{d}{-}\mathbf{d}')\,] \\ &= (c_1{-}c'_1)\,\mathbf{u}_{1,2} \;+\; [\,0\;\; (c_1{+}c_2{-}c'_1{-}c'_2)\;\; (\mathbf{d}{-}\mathbf{d}')\,]\;, \end{aligned}$$

which implies that $\mathbf{c}$ and $\mathbf{c}'$ are at decomposability distance at most t, contradicting our assumption on $\mathcal{C}$. Therefore, $\mathcal{C}'$ is an optimal $(n{-}1, n{-}t)$ $(t{-}1)$-LCC over F and, thus, $n{-}1 \leq N(t{-}1,q)$. $\square$

Corollary 4. *The following bounds hold:*

$$N(t,q) \;\leq\; \begin{cases} q & \text{for } t=1 \\ t-2+\left\lceil \sqrt{2(q+1)} \right\rceil & \text{for } t \geq 2 \end{cases} .$$

Proof. The cases $t=1$ and $t=2$ follow, respectively, from (5) and Corollary 2. The bounds for larger values of t follow from Theorem 4. □

In particular, when $2 \leq t \ll q$, Corollary 4 implies that we cannot find optimal t-LCCs that are counterparts to extended Reed-Solomon codes of length $q+1$.

The following result is an analogue of Theorem 4 with respect to the maximal length $M(k,q)$ of any optimal LCC of dimension k over $GF(q)$.

Theorem 5. *For any $k \geq 2$,*

$$M(k,q) \;\leq\; M(k-1,q) \;+\; 1 .$$

Proof. Let $\mathcal{C}$ be an optimal (n,k) t-LCC over $F=GF(q)$ with $n=M(k,q)$ and $t=n-k$. Define $\mathcal{C}'$ as the set of all vectors in F^{n-1} obtained by deleting the first coordinate from the codewords of $\mathcal{C}(0)$. We claim that the MDD of $\mathcal{C}'$ is at least $t+1$. Suppose to the contrary that $\mathbf{c}$ and $\mathbf{c}'$ are two vectors in $\mathcal{C}'$ at decomposability distance t or less. Then both vectors belong to the same class $\mathcal{C}'(a)$ in $\mathcal{C}'$, and $[\,-a \;\; \mathbf{c}\,]$ and $[\,-a \;\; \mathbf{c}'\,]$ must be codewords in $\mathcal{C}$. But these codewords are at decomposability distance at most t of each other, which is a contradiction. Therefore, $\mathcal{C}'$ is an $(n-1,k')$ t-LCC with $t=n-k$. By Corollary 1 we have $k'=k-1$, namely, $\mathcal{C}'$ is optimal. Hence, $n-1 \leq M(k-1,q)$. □

We end this section by mentioning that, using an argument similar to the proof of the conventional Gilbert-Varshamov bound for ECCs [17, Ch. 1], one can show that there exists an $[n,n-r]$ t-LCC over $GF(q)$ whenever

$$q^r \;>\; |\mathcal{B}_F(n-1;2t-1)| . \tag{9}$$

By Lemma 4 and the inequality $\sum_{\tau=0}^{w} \binom{n(n-1)/2}{\tau} \leq (n^2/2)^w$ (which holds for $n>2$), it follows from (9) that for every $1 \leq t < n$ there exists an $[n,n-r]$ t-LCC over $GF(q)$ with

$$r \;\leq\; \left\lceil\; (2t-1)\log_q n \;\;+\;\; (t-1)\log_q\left(\frac{q-1}{2}\right) \;\right\rceil .$$

4 Constructions of optimal LCCs

We will mainly be interested in constructing the longest possible optimal t-LCCs for various values of t.

4.1 The cases $t = 1$ and $t \geq n-1$

Construction 1. (The case $t = 1$.) *Let $\mathcal{C}$ be an $[n, n-r]$ code over $F = GF(q)$. Then $\mathcal{C}$ is 1-LCC if and only if the columns of its parity-check matrix H are distinct.*

Proof. A nonzero vector in F^n is 1-decomposable if and only if it is a scalar multiple of a vector in $\mathcal{W}_F(n)$. Such a vector is a codeword in $\mathcal{C}$ if and only if there are two identical columns in H. □

In particular, we can build an optimal linear 1-LCC of length q over $GF(q)$ by letting the entries of the single-row parity-check matrix H range over all elements of $GF(q)$. By Corollary 4, this is the longest possible optimal 1-LCC.

We next turn to the other extreme values of t. For $t \in \{n-1, n\}$, Theorem 2 allows for a code of dimension at most 1. The next construction shows that dimension 1 is indeed attainable.

Construction 2. (The case $t \in \{n-1, n\}$.) *Let $\mathcal{C}$ be an $[n, 1]$ code over F with a generator matrix*

$$G \;=\; [\,\alpha_1\ \alpha_2\ \alpha_3\ \ldots\ \alpha_n]\,,$$

where the α_i are elements of F (not necessarily distinct) whose sum is nonzero. Then $\mathcal{C}$ is an n-LCC.

Proof. Each class $\mathcal{C}(a)$ contains one codeword, so the MDD of $\mathcal{C}$ is infinity. The following procedure corrects up to n errors: Let $\mathbf{c} = uG$ be the transmitted codeword and $\mathbf{e} = [\,e_i\,]_{i=1}^n$ be the error vector. Knowing the error values, the decoder also knows the sum $s = \sum_{\ell=1}^t E_\ell = \sum_{i=1}^n e_i$. Hence, by summing up the entries of the received word, the decoder obtains the value $s + u\sum_{i=1}^n \alpha_i$, from which it can recover u and, hence, $\mathbf{c}$.□

A code as in Construction 2 is optimal for $t \in \{n-1, n\}$ and can be obtained for any arbitrary length, by taking, say, a sufficiently long unit vector as G. Observe that unlike conventional MDS codes, equality in the bound of Theorem 2 is not preserved under duality: the length of an optimal 1-LCC is bounded from above by q (Corollary 4), while we can have arbitrarily long optimal $(n-1)$-LCCs.

4.2 The case $t = 2$

Our constructions for $t = 2$ make use of the following definition. A subset S of an Abelian group A is called a *weak Sidon set* if for any four distinct elements $x, y, z, w \in S$ we have $x + y \neq z + w$ (this is also referred to as an S_2-*set* in [8]; see also [14]). A *(strong) Sidon set* is defined similarly, except that the inequality $x + y \neq z + w$ is required for any four elements of which at least three are different [4]. It is easy to check that the two definitions coincide when the elements of A other than unity all have order 2 (as is the case when A is the additive group of $GF(2^m)$).

Construction 3. (The case $t = 2$.) *Let $\mathcal{C}$ be an $[n, n-2]$ code over $F = GF(q)$ with a parity-check matrix of the form*

$$H \;=\; \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \\ \alpha_1^{-1} & \alpha_2^{-1} & \dots & \alpha_n^{-1} \end{bmatrix} , \tag{10}$$

where $S = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ is a subset of distinct nonzero elements of F. Then, $\mathcal{C}$ is a 2-LCC if and only if S is a weak Sidon set of the multiplicative group F^* of F.

Proof. Let $\mathbf{e} = [\, e_i \,]_{i=1}^n$ be an error vector whose nonzero entries are given by $e_{i_1} = E_1$ and $e_{i_2} = E_2$, and let $\mathbf{e}' = [\, e'_j \,]_{j=1}^n$ be an error vector whose nonzero entries are $e'_{j_1} = E_1$ and $e'_{j_2} = E_2$. Suppose that $\mathbf{e}H^T = \mathbf{e}'H^T$. We write $\alpha = \alpha_{i_1}$, $\beta = \alpha_{i_2}$, $\gamma = \alpha_{j_1}$, and $\delta = \alpha_{j_2}$. Also, since $\mathcal{C}$ is a 1-LCC by Construction 1, we assume that $\alpha \neq \gamma$ and $\beta \neq \delta$. We have

$$E_1\alpha + E_2\beta \;=\; E_1\gamma + E_2\delta \tag{11}$$

and

$$E_1\alpha^{-1} + E_2\beta^{-1} \;=\; E_1\gamma^{-1} + E_2\delta^{-1} \; . \tag{12}$$

These two equations, in turn, are equivalent to

$$E_1(\alpha - \gamma) \;=\; E_2(\delta - \beta) \tag{13}$$

and

$$E_1\frac{\alpha - \gamma}{\alpha \cdot \gamma} \;=\; E_2\frac{\delta - \beta}{\beta \cdot \delta} \; . \tag{14}$$

Dividing each side of (14) by the respective side of (13), we obtain

$$\alpha \cdot \gamma \;=\; \beta \cdot \delta \; . \tag{15}$$

Now, if S is a weak Sidon set of the multiplicative group F^*, then (15) cannot hold unless at most three of the values $\alpha, \beta, \gamma, \delta$ are different. This may happen only in the following two cases:

(i) $\alpha = \beta$ (implying by (15) also the equality $\gamma = \delta$), in which case both $\mathbf{e}$ and $\mathbf{e}'$ contain at most one nonzero location with value $E_1 + E_2$. Since $\mathcal{C}$ is 1-LCC, we must have $\mathbf{e} = \mathbf{e}'$.

(ii) $\alpha = \delta$ (and so $\beta = \gamma$), in which case we must have by (11) the equality $(E_1 - E_2)(\alpha - \beta) = 0$. So, if $\alpha \neq \beta$, then $E_1 = E_2$, implying $\mathbf{e} = \mathbf{e}'$.

On the other hand, if S is not a weak Sidon set, then there exist distinct α, β, γ and δ in S such that (15) holds. Now, select E_1 and E_2 so that (13) is satisfied (and, thus, so is (14)). The respective vectors $\mathbf{e}$ and $\mathbf{e}'$ have distinct supports and therefore are different. However, since both (11) and (12) hold, these vectors have the same syndrome and, so, the code $\mathcal{C}$ contains a 2-decomposable codeword $\mathbf{e} - \mathbf{e}'$. □

When s is a power of a prime, we can get strong Sidon sets of size $s+1$ of the additive group of the integers modulo s^2+s+1 by taking Singer difference sets in projective planes [15, Ch. 11]. This yields strong Sidon sets of size $\Omega(\sqrt{\ell})$ for any cyclic group of size ℓ and, in particular, for the cyclic multiplicative group F^*. Thus, Construction 3 yields codes of length $n = \Omega(\sqrt{q})$, attaining the asymptotic upper bound of Corollary 2 (up to a constant multiplier of the length).

Next, we present another construction of 2-LCCs, based on weak Sidon sets of the *additive* group of the field.

Construction 4. *Let $\mathcal{C}$ be an $[n, n-2]$ code over F with a parity-check matrix of the form*

$$H \;=\; \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_n^2 \end{bmatrix} ,$$

where $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a subset of distinct elements of F. Then, $\mathcal{C}$ is a 2-LCC if and only if S is a weak Sidon set of the additive group of F.

Proof. The proof is similar to that of Construction 3. Equation (11) remains as is, whereas Equation (12) becomes

$$E_1\alpha^2 + E_2\beta^2 \;=\; E_1\gamma^2 + E_2\delta^2 \; .$$

Rearranging terms and dividing respective sides, we obtain

$$\alpha \;+\; \gamma \;=\; \beta \;+\; \delta \; .$$

We continue now as in the proof of Construction 3. □

Bounds on sizes of strong Sidon sets for general finite Abelian groups can be found in [3], [4], and [14]. In particular, it is easy to verify that in a finite Abelian group A

of odd order, any strong Sidon set S must satisfy the inequality $|S|(|S|-1)+1 \leq |A|$. Singer difference sets satisfy this bound with equality.

As for weak Sidon sets for a finite Abelian group A, it is known [8] that

$$|S|(|S|-1) \ \leq \ 2|A| \ .$$

Table V in [8] (see also Table IV in [14]) lists the smallest cardinality $v(k)$ of any Abelian group A containing a weak Sidon set of size k, for $2 \leq k \leq 28$ (this is, in a sense, the inverse of the function we are interested in). When A is the additive group of a finite field, Construction 4 in conjunction with Corollary 2 yield the tighter bound

$$|S|(|S|-1) \ \leq \ 2(|A|-1) \tag{16}$$

for all but a finite set of values of $q = |A|$. Inequality (16) is proven in [14] for the case where A is the ring of integers modulo $|A|$. This is also the best known upper bound on the size of *strong* Sidon sets of the additive group of $GF(2^m)$ [4]. For weak Sidon sets of Abelian groups A of odd order we have the following.

Lemma 5. *Let A be a finite Abelian group of odd size. Then, any weak Sidon set S satisfies the inequality*

$$|S|(|S|-3)+1 \ \leq \ |A| \ .$$

Proof. Let T denote the set of all $|S|(|S|-1)$ ordered pairs $\{(x,y) \mid x,y \in S, x \neq y\}$, and let (a,b) and (z,w) be distinct pairs in T such that $a-b = z-w$. Since S is a weak Sidon set, we must have either $a = w$ or $b = z$. In the first case we have $2a = b+z$ and the set $\{b,z\} \subseteq S$ is uniquely defined by a, and in the second case we have $2b = a+w$ and the set $\{a,w\} \subseteq S$ is uniquely defined by b. It thus follows that for any given $a \in S$, there are at most two pairs in T of the form (a,u) such that the difference $a-u$ equals the difference $z-w$ associated with some other pair $(z,w) \in T$. Since such a difference is never zero, we have $|S|(|S|-1) = |T| \leq |A| - 1 + 2|S|$. □

Example 4. The set $S = \{0,1,2,4,7\}$ is a weak Sidon set for the additive group of $GF(11)$ [14]. The size of this set attains the bound of Lemma 5. Let $\mathcal{C}_{11}$ be the linear $[5,3]$ code over $GF(11)$ with the parity-check matrix

$$H_{11} \ = \ \begin{bmatrix} 0 & 1 & 2 & 4 & 7 \\ 0 & 1 & 4 & 5 & 5 \end{bmatrix} \ .$$

By Construction 4, the code $\mathcal{C}_{11}$ is a 2-LCC. Furthermore, the code $\mathcal{C}_{11}$ attains the bound of Corollary 2 and, as such, it is a semi-perfect 2-LCC.

Example 5. The set $S = \{0, 1, 2, 4, 7, 12\}$ is a weak Sidon set for the additive group of $GF(19)$ which attains the bound of Lemma 5 [14]. By Construction 4, S leads to a $[6, 4]$ code $\mathcal{C}_{19}$ over $GF(19)$ which is a 2-LCC. By Corollary 2 this is the longest optimal 2-LCC possible, even though it is not a semi-perfect code.

We remark that Constructions 3 and 4 do not always yield the longest optimal 2-LCCs. For example, the largest weak Sidon set in the multiplicative group of $GF(5)$ is of size 3, which is also the size of the largest weak Sidon set in the additive group of $GF(5)$ (Lemma 5 is not tight in this case). Nevertheless, we showed in Example 1 an optimal 2-LCC of length 4 over $GF(5)$. It can be verified that the construction of Example 1 can be described by way of appending a column to the parity-check matrix defined by Constructions 3 or 4.

Example 6. The set $S = \{0, 1, 2, 5, 9, 18, 24\}$ is a weak Sidon set of size 7 for the additive group of $GF(29)$ and, as such, it attains the bound of Lemma 5. (In fact, a weak Sidon set of this size is unique up to scaling and translation, namely, up to a transformation $x \mapsto ax + b$ applied to each element $x \in S$ for some constants $a \neq 0$ and b.) Now, Corollary 2 allows for a semi-perfect optimal 2-LCC of length 8 over $GF(29)$. However, no column can be added to any parity-check matrix obtained from S (and its translates) by using Construction 4 to form an optimal 2-LCC of length 8. Nevertheless, there is a semi-perfect optimal 2-LCC over $GF(29)$ with the parity-check matrix

$$H_{29} = \begin{bmatrix} 0 & 1 & 0 & 2 & 3 & 4 & 24 & 27 \\ 0 & 0 & 1 & 2 & 5 & 18 & 7 & 22 \end{bmatrix}.$$

We turn now to fields of size $q = 2^m$. We can form a Sidon set of the additive group of $GF(2^m)$ by taking the columns of a parity-check matrix of an $[n, n-m]$ 2-ECC binary code, together with the zero column. For examples of the best possible such codes, see [17, p. 675]. As mentioned before, the best upper bound on the size of Sidon sets S in the additive group of fields $GF(2^m)$ is $|S|(|S| - 1) \leq 2(2^m - 1)$. This bound is attained for $m = 4$ by $S = \{0, 1, \alpha, \alpha^2, \alpha^3, 1+\alpha+\alpha^2+\alpha^3\}$, where α is any element of $GF(16) - GF(4)$, yielding by Construction 4 a semi-perfect 2-LCC of length 6 over $GF(16)$. For $q = 32$, the largest Sidon set in the additive group of $GF(32)$ is of size 7, from which, by extension, we can obtain an $[8, 6]$ 2-LCC with the parity-check matrix

$$H_{32} = \begin{bmatrix} 0 & 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 & \alpha^{26} & 0 \\ 0 & 1 & \alpha^2 & \alpha^4 & \alpha^6 & \alpha^8 & \alpha^{21} & 1 \end{bmatrix},$$

where α is a root of the binary polynomial $x^5 + x^2 + 1$.

For even q, Construction 4 will turn out to be a special case of the more general Construction 7 described below.

We end this section by presenting a simple construction of $[q, q-3]$ 2-LCCs over $GF(q)$. These codes are not optimal, but they get rather close to the sphere packing bound (6).

Construction 5. *Let $\mathcal{C}$ be a $[q, q-3]$ code over $GF(q)$ with a parity-check matrix of the form*

$$H \;=\; \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_q \\ \alpha_1^2 & \alpha_2^2 & \ldots & \alpha_q^2 \\ \alpha_1^3 & \alpha_2^3 & \ldots & \alpha_q^3 \end{bmatrix}, \tag{17}$$

where the α_j range over all distinct elements of $GF(q)$. Then, $\mathcal{C}$ is a 2-LCC.

Proof. It follows from Lemma 3 and the linearity of $\mathcal{C}$ that the MDD of $\mathcal{C}$ is at least half the minimum Hamming distance of $\mathcal{C}(0)$. Now, the subcode $\mathcal{C}(0)$ is a $[q, q-4]$ MDS code with minimum Hamming distance 5. Therefore, the MDD of $\mathcal{C}$ is at least 3. □

The codes of Construction 5 are close to the bound (6) in two respects: For the given length $n = q$, they attain the minimum integer redundancy $r = 3$ allowed by (6) for codes over $GF(q)$ with $q \geq 4$; and, for the given redundancy $r = 3$, their length is asymptotically within a factor $\sqrt{2}$ of the length $q\sqrt{2}+o(q)$ allowed by (6). Construction 5 can be extended by taking the α_j to be elements of $GF(q^m)$. We obtain a construction of an $[n, n-r]$ 2-LCC over $GF(q)$ with $r \leq 3m$ and $n = q^m \geq q^{r/3}$ (compared to the upper bound $O(q^{(r-1)/2})$ of (6)).

The 2-LCCs of Constructions 3 and 4 can be efficiently decoded by solving quadratic equations over $GF(q)$. For example, in the case of Construction 3, having the syndrome vector $[\, s_1 \; s_2 \,] = \mathbf{y}H^T$ for the received word $\mathbf{y}$ and the parity-check matrix H in (10), we obtain the two equations

$$s_1 \;=\; E_1\alpha_{j_1} + E_2\alpha_{j_2} \quad \text{and} \quad s_2 \;=\; E_1\alpha_{j_1}^{-1} + E_2\alpha_{j_2}^{-1}$$

in the unknown error locators α_{j_1} and α_{j_2}, where E_1 and E_2 are the given error values. The proofs of the constructions guarantee that the solution for the pair of error locators $(\alpha_{j_1}, \alpha_{j_2})$ is unique, provided that they are taken from the set S. The code of Construction 5 can be decoded by using any of the decoding algorithms for a double-error-correcting Reed-Solomon code whose parity-check matrix is obtained by adding an all-one row to the matrix H in (17). The syndrome value that corresponds to this row is the sum of the error values, which is known to the decoder.

4.3 The case $t = n-2$

Construction 6. (The case $t = n-2$.) *Let $\mathcal{C}$ be an $[n, 2]$ code over F with a generator matrix of the form*

$$G = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_n \\ \beta_1 & \beta_2 & \beta_3 & \dots & \beta_n \end{bmatrix},$$

where the α_i are elements of F (not necessarily distinct) such that $\sum_{i=1}^{n} \alpha_i \neq 0$, and the β_i are elements of F (not necessarily distinct) such that $\sum_{i=1}^{n} \beta_i = 0$ but no sum of less than n of these elements is zero. Then, $\mathcal{C}$ is an optimal LCC.

Proof. The class $\mathcal{C}(0)$ consists of all scalar multiples of the second row of G. By Lemma 2, the MDD of $\mathcal{C}(0)$ (and therefore of any other class $\mathcal{C}(a)$) equals $n-1$. Hence, this is also the MDD of $\mathcal{C}$. □

Let $q = p^m$, where p is the characteristic of $F = GF(q)$, and let S be a multiset of elements of F constructed by taking $p-1$ copies of each element of a basis of F over $GF(p)$. Then the multiset $S \cup \{-\sum_{a \in S} a\}$ can be used for $\{\beta_i\}_{i=1}^{n}$ in Construction 6. The resulting code is of length $n = m(p-1)+1$. It will follow from Corollary 5 below that this is the maximum attainable length for an optimal LCC of dimension 2 over $GF(p^m)$.

The code of Construction 6 can be decoded as follows: Let $\mathbf{c} = u_1\mathbf{g}_1 + u_2\mathbf{g}_2$ be the transmitted codeword, where $\mathbf{g}_1$ and $\mathbf{g}_2$ denote the rows of G. Also, let $\mathbf{y} = [\, y_i \,]_{i=1}^{n}$ be the received vector and let $E_1, E_2, \dots, E_{n-2}$ be the given error values. We recover the coefficient u_1 similarly to the decoding of Construction 2. As for the coefficient u_2, we proceed as follows. For $i = 1, 2, \dots, n-1$, we make the assumption that the ith component of $\mathbf{y} - u_1\mathbf{g}_1$ is error-free and compute the value of u_2 under this hypothesis. We then verify our hypothesis by checking whether the multiset of entries of the implied error vector $\mathbf{y} - u_1\mathbf{g}_1 - u_2\mathbf{g}_2$ is consistent with the multiset of the given values E_ℓ. Such a comparison of multisets can be done in $O(n \log n)$ operations on elements of $GF(q)$ by simple sorting methods. The total complexity of the procedure is therefore $O(n^2 \log n)$.

4.4 Other values of t

The existence of a multiset $\{\beta_i\}_{i=1}^{n}$ satisfying the conditions of Construction 6 was shown to be a sufficient condition for the existence of an optimal t-LCC with $t = n-2$. We now show that it is also a necessary condition for $t \leq n-2$.

Theorem 6. *Let $\mathcal{C}$ be an optimal t-LCC of dimension ≥ 2 over F. Then, there exists a multiset S of $t+1$ elements of F, such that no nonempty subset of S sums to zero.*

Proof. Let $\mathcal{C}$ be an $(n, n-t)$ t-LCC with $t \leq n-2$ and consider its partition into classes $\mathcal{C}(a)$. Then, for at least one element $a \in F$, we must have $|\mathcal{C}(a)| \geq |\mathcal{C}|/q = q^{n-t-1}$. Thus, we can find two distinct codewords $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}(a)$ that agree in their last $n-t-2$ coordinates. Define $\mathbf{d} = [\, d_i \,]_{i=1}^{n} = \mathbf{c}_1 - \mathbf{c}_2$. Then, we have $d_i = 0$ for $i > t+2$ and, by the definition of $\mathcal{C}(a)$, we also have $\sum_{i=1}^{t+2} d_i = \sum_{i=1}^{n} d_i = 0$. It follows that the vector $\mathbf{d}' = [\, d_1 \; d_2 \; \ldots \; d_{t+2} \,]$ is $(t+1)$-decomposable. On the other hand, the vector $\mathbf{d}'$ cannot be t-decomposable. Hence, by Lemma 2, the multiset $S = \{d_i\}_{i=1}^{t+1}$ must be such that no nonempty subset of S sums to zero. □

Olson [18], [19] has shown that, for $F = GF(p^m)$, the maximal size of a multiset S satisfying the conditions of Theorem 6 is $m(p-1)$ (see also [10]). This size is attained by the multiset S described in the discussion following Construction 6. Olson's results, together with Theorem 6, lead to the following.

Corollary 5. *There exists an optimal t-LCC of dimension ≥ 2 over $GF(p^m)$ only if $t \leq (p-1)m - 1$.*

It follows from Corollary 5 that the length range of optimal t-LCCs for large values of t depends crucially on the structure of the field $GF(q)$. When q is a prime p, we can apply Construction 6 to obtain an optimal $[p, 2]$ LCC over $GF(p)$. On the other hand, when $q = 2^m$, the largest multiset in $GF(q)$ that satisfies the conditions of Theorem 6 is of size m. Hence, the maximum length of an optimal LCC of dimension 2 over $GF(q)$ varies from being logarithmic in q for $q = 2^m$ to being linear in q when q is prime.

Next, we describe a construction of optimal t-LCCs for arbitrary t over extension fields $F = GF(p^m)$ for $m \geq t$, with a corresponding decoding algorithm. The proof of the construction makes use of the following lemma.

Lemma 6. [17, p. 119]. *Let $a_1, a_2, \ldots, a_\rho$ be elements of $GF(p^m)$ that are linearly independent over $GF(p)$. Then the $\rho \times \rho$ matrix $[\, a_j^{p^{i-1}} \,]_{i,j=1}^{\rho}$ is nonsingular.*

Lemma 6 does not require p to be prime. The same applies to the following construction, although it can attain larger values of t when p is the characteristic of F.

Construction 7. *Let $t \leq \min\{n, m\}$ and let $\mathcal{C}$ be an $[n, n-t]$ code over $F = GF(p^m)$ with a parity-check matrix of the form*

$$H = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1^p & \alpha_2^p & \cdots & \alpha_n^p \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1^{p^{t-1}} & \alpha_2^{p^{t-1}} & \cdots & \alpha_n^{p^{t-1}} \end{bmatrix}, \tag{18}$$

where $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ is a set of n distinct elements of F. Then, $\mathcal{C}$ is a t-LCC if every subset of $2t$ or less nonzero elements of S are linearly independent over $GF(p)$.

Proof. We present a decoding algorithm that uniquely recovers the error vector $\mathbf{e}$ provided that the number τ of errors is t or less. The algorithm is based upon the one described in [21] for decoding maximum-rank array codes (see also [11]).

Let $\mathbf{y} \in F^n$ be the received word and let $E_1, E_2, \ldots, E_\tau$ be the given error values, which we regard as vectors in the linear space $F = GF(p^m)$ over $GF(p)$. Also, let $\boldsymbol{\delta} = [\, \delta_1 \; \delta_2 \; \ldots \; \delta_\rho \,]$ be a vector over F whose entries form a basis of size $\rho \leq \tau \leq t$ of the linear span of the set of error values E_ℓ. Consider the syndrome vector $\mathbf{s} = [\, s_1 \; s_2 \; \ldots \; s_t \,]$ of the received word with respect to the matrix H of Equation (18), i.e., $\mathbf{s} = \mathbf{y}H^T$. In order to prove the construction, it suffices to show that there is a unique $\rho \times n$ matrix U over $GF(p)$ such that

$$\mathbf{s} = \boldsymbol{\delta} U H^T \,. \tag{19}$$

The unique error vector $\mathbf{e}$ is then given by $\mathbf{e} = \boldsymbol{\delta} U$.

Now, write

$$\beta_j \;=\; \sum_{i=1}^{n} U_{j,i}\alpha_i \;, \quad j = 1, 2, \ldots, \rho \;, \tag{20}$$

where $U_{j,i}$ stands for the (j,i)th entry of U. Recalling that each such entry is in $GF(p)$, we obtain by (19) the following set of linear equations over F:

$$s_\ell \;=\; \sum_{j=1}^{\rho} \delta_j \beta_j^{p^{\ell-1}} \;, \quad \ell = 1, 2, \ldots, t \;.$$

For $\ell = 1, 2, \ldots, \rho$, we raise each side of the ℓth equation to the power $p^{m-\ell+1}$, thus obtaining

$$s_\ell^{p^{m-\ell+1}} \;=\; \sum_{j=1}^{\rho} \delta_j^{p^{m-\ell+1}} \beta_j \;, \quad \ell = 1, 2, \ldots, t \;. \tag{21}$$

Since the elements δ_j are linearly independent over $GF(p)$, so are the elements $\delta_j^{p^{m-\rho+1}}$. Hence, by Lemma 6, the matrix $[\, \delta_j^{p^{m-\ell+1}} \,]_{\ell,j=1}^{\rho}$ is nonsingular. Therefore, we can solve (21) uniquely for the elements β_j, $j = 1, 2, \ldots, \rho$.

Finally, in order to recover the matrix U, we make use of the fact that the nonzero elements of S can be regarded as columns of a parity-check matrix of a t-ECC with redundancy m over $GF(p)$. Therefore, knowing β_j, we can solve (20) uniquely for the $U_{j,i}$, except when S contains the zero element, in which case (20) does not give information on entries $U_{j,i}$ corresponding to $\alpha_i = 0$. This missing column of U, which corresponds to an entry of $\mathbf{e}$, can be recovered from the knowledge of the sum $\sum_{i=1}^{n} e_i = \sum_{\ell=1}^{\tau} E_\ell$. $\square$

We pointed out that the nonzero elements of S correspond to the columns of a parity-check matrix of an $[n, n-m]$ t-ECC over $GF(p)$ (or an $[n-1, n-1-m]$ t-ECC when $0 \in S$). Such a code can be obtained by constructing a BCH code of designed distance $2t+1$ and redundancy at most m. The redundancy constraint is satisfied by taking the roots of the code from $GF(p^h)$, where $\lceil 2t(p-1)/p \rceil h \leq m$. This leads to a code length $n > (1/p)q^{1/\lceil 2t(p-1)/p \rceil} = \Omega((1/p)\sqrt[2t]{q})$. For the special case $p = 2$, the constraint on h becomes $t\,h \leq m$, yielding codes of length $\Omega(\sqrt[t]{q})$.

When BCH codes are used to construct S, the proof of Construction 7 provides a polynomial-time decoding algorithm: The computation of the elements β_j amounts to Gaussian elimination and solving linear equations over F, whereas the reconstruction of the matrix U out of the elements β_j can be carried out by any of the known decoding algorithms for BCH codes over $GF(p)$.

Note that Construction 7 becomes vacuous for $p = 2$ and $t \log_2 t \geq m = \log_2 q$ (compare with Corollary 5), since in this case we have $n \leq 2^h \leq t$. For odd values of q and $t = 2$, Construction 3 yields considerably longer codes than those of Construction 7.

5 Summary

Table 1 summarizes the bounds on the length of optimal LCCs, and the main constructions presented in the paper.

Case	Upper Bound on Length		Construction Length	
$t = 1$	q	(Corollary 4)	q	(Construction 1)
$t = 2$	$O(\sqrt{q})$	(Corollary 2)	$\Omega(\sqrt{q})$	(Constructions 3,4)
$t = n-1$	none		any length	(Construction 2)
$q = p^m$, $t = n-2$	$m(p-1)+1$	(Corollary 5)	$m(p-1)+1$	(Construction 6)
$q = 2^m$, $t \leq m$	$O(\sqrt{q})$	(Corollary 4)	$\Omega(\sqrt[t]{q})$	(Construction 7)
$q = p^m$, $t \leq m$	$O(\sqrt{q})$	(Corollary 4)	$\Omega(\frac{1}{p}\sqrt[2t]{q})$	(Construction 7)

Table 1: Summary of bounds and constructions for optimal LCCs.

Closing the asymptotic gap in code length between Construction 7 and the bounds of Corollaries 4 and 5 remains an open research problem.

Acknowledgment

We thank Noga Alon and Tuvi Etzion for introducing us to the literature on Sidon sets.

Appendix

We prove here that computing the decomposability weight of a vector of length n over $GF(p)$ is NP-complete in the strong sense [12, Section 4.2]; namely, the existence of an algorithm for computing the decomposability weight in time complexity which is polynomial in n and p implies P = NP ("strong sense" refers to the fact that we allow complexity which is polynomial in the value of p rather than the size, $\log p$, of its representation).

Theorem 7. *The following* DECOMPOSABILITY *decision problem is NP-complete in the strong sense:*

Instance: A prime p and a vector $\mathbf{c}$ of length $4m$ over $GF(p)$.

Question: Is $\mathbf{c}$ $(3m)$-decomposable?

Proof. The problem is clearly in NP. As for completeness, we prove by reduction from the 3-PARTITION problem which is defined as follows (see [12, Section 4.2.2]):

Instance: A positive integer b and a set A of $3m$ positive integers such that $b/4 < a < b/2$ for every $a \in A$ and $\sum_{a \in A} a = m\, b$.

Question: Can we partition A into m disjoint sets $A_1, A_2, \ldots, A_m$ such that $\sum_{a \in A_k} a = b$ for every $k = 1, 2, \ldots, m$?

3-PARTITION is known to be NP-complete in the strong sense. Note that if a qualifying partition of A exists, then each partition class A_k must contain exactly three elements of A.

Given an instance $(b\,;\, A = \{a_1, a_2, \ldots, a_{3m}\})$ of 3-PARTITION, we construct an instance of DECOMPOSABILITY as follows. Let s be an odd positive integer less than $b/2$ such that $p = 2b + s$ is prime. By the Prime Number Theorem [9],[20], there always exists such an integer s for $b > 4$, and s can be found by a brute-force search in time complexity which is polynomial in the value of b. Consider the integer vector $\mathbf{c} = [\, c_1 \; c_2 \; \ldots \; c_{4m} \,]$ defined by

$$c_i \;=\; \begin{cases} 2a_i & \text{for } i = 1, 2, \ldots, 3m \\ s & \text{for } i = 3m{+}1, 3m{+}2, \ldots, 4m \end{cases} .$$

We show that the set $\{1, 2, \ldots, 4m\}$ can be partitioned into m disjoint nonempty subsets $S_1, S_2, \ldots, S_m$ such that $\sum_{i \in S_k} c_i \equiv 0 \pmod{p}$ if and only if the set $A = \{a_i\}_{i=1}^{3m}$ can be partitioned into m disjoint sets $A_1, A_2, \ldots, A_m$, each consisting of three elements summing up to b. The result will then follow from Lemma 2.

The "if" part is straightforward: take

$$S_k \;=\; \{\, i \mid a_i \in A_k \,\} \;\cup\; \{\, 3m{+}k \,\}\,, \quad k = 1, 2, \ldots, m\,.$$

As for the "only if" part, we first show that $|S_k| \geq 4$ and characterize the case where equality holds. Write $\sum_{i \in S_k} c_i = \ell \cdot p$ for some positive integer ℓ. If $\ell > 1$, then we must have $|S_k| > 4$. If $\ell = 1$, then the number of elements in S_k which are greater than $3m$ must be odd. If this odd number is 3 or more, then, again, $|S_k| > 4$. Otherwise, the subset S_k can be written as $\{i_1, i_2, i_3, 3m{+}j\}$, where $1 \leq i_1 < i_2 < i_3 \leq 3m$, $a_{i_1} + a_{i_2} + a_{i_3} = b$, and $j \in \{1, 2, \ldots, m\}$.

We conclude that if there is a partition of $\{1, 2, \ldots 4m\}$ into m subsets S_k such that $\sum_{i \in S_k} c_i \equiv 0 \pmod{p}$, then each S_k must have size 4 and, therefore, it must be of the form $\{\, i \mid a_i \in A_k \,\} \cup \{\, 3m{+}k \,\}$, where $A_1, A_2, \ldots, A_m$ forms a qualifying partition of 3-PARTITION. □

References

[1] R. Ahlswede, L.A. Bassalygo, M.S. Pinsker, *Asymptotically dense nonbinary codes correcting a constant number of localized errors, Proc. Third Int'l Workshop on Algebraic and Combinat. Coding Theory,* Tyrnovo, Bulgaria (1992).

[2] R. Ahlswede, L.A. Bassalygo, M.S. Pinsker, *Nonbinary codes correcting localized errors, IEEE Trans. Inform. Theory,* 39 (1993), 1413–1416.

[3] N. Alon, H. Lefmann, V. Rödl, *On an anti-Ramsey type result, Colloq. Math. Soc. János Bolyai 60: Sets, Graphs, and Numbers,* Budapest (1991), 9–22.

[4] L. Babai, V.T. Sós, *Sidon sets in groups and induced subgraphs of Cayley graphs, Europ. J. Comb.,* 6 (1985), 101–114.

[5] L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, *Coding for channels with localized errors, Proc. Fourth Soviet–Swedish Workshop in Inform. Theory,* Gotland, Sweden (1989), 95–99.

[6] L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, *Coding for channels with partially localized errors, IEEE Trans. Inform. Theory,* 37 (1991), 880–884.

[7] L.A. Bassalygo, S.I. Gelfand, M.S. Pinsker, *Simple methods of deduction of lower bounds in coding Theory, Probl. Peredac. Inform.,* 27, No. 4 (1991), 3–8.

[8] A.E. Brouwer, J.B. Shearer, N.J.A. Sloane, W.D. Smith, *A new table of constant weight codes, IEEE Trans. Inform. Theory,* 36 (1990), 1334–1380.

[9] H. Davenport, *Multiplicative Number Theory,* Second Edition, revised by H.L. Montgomery, Springer, Berlin, 1980.

[10] G.T. Diderrich, H.B. Mann, *Combinatorial problems in finite Abelian groups, A Survey of Combinatorial Theory,* J.N. Srivastava et al., eds., North-Holland, Amsterdam (1973).

[11] E.M. Gabidulin, *Theory of codes with maximum rank distance, Probl. Peredach. Inform.,* 21 (1985), 3–16 (in Russian; pp. 1–12 in the English translation).

[12] M.R. Garey, D.S. Johnson, *Computers and Intractability — A Guide to the Theory of NP-Completeness,* Freeman, New York, 1979.

[13] J.M. Goethals, *Cyclic error-locating codes, Inform. Control,* 10 (1967), 378–385.

[14] R.L. Graham, N.J.A. Sloane, *On additive bases and harmonious graphs, SIAM J. Alg. Disc. Meth.,* 1 (1980), 382–404.

[15] M. Hall, Jr., *Combinatorial Theory,* Second edition, John Wiley, New York, 1986.

[16] G.A. Kabatyansky, *The construction of code correcting single localized error, Proc. Third Int'l Workshop on Algebraic and Combinat. Coding Theory,* Tyrnovo, Bulgaria (1992).

[17] F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes,* North-Holland, Amsterdam, 1977.

[18] J.E. Olson, *A combinatorial problem on finite Abelian groups, I, J. Number Theory,* 1 (1969), 8–10.

[19] J.E. Olson, *A combinatorial problem on finite Abelian groups, II, J. Number Theory,* 1 (1969), 195–199.

[20] J.B Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers, Illinois J. of Math,* 6 (1962), 64–94.

[21] R.M. Roth, *Maximum-rank array codes and their application to crisscross error correction, IEEE Trans. Inform. Theory,* 37 (1991), 328–336.

[22] J.K. Wolf, *On an extended class of error-locating codes, Inform. Control,* 8 (1963), 163–169.

[23] J.K. Wolf, B. Elspas, *Error locating codes — a new concept in error control, IEEE Trans. Inform. Theory,* 9 (1963), 113–117.