Bull. Aust. Math. Soc. **112** (2025), 550–560
doi:10.1017/S0004972725000012

OPTIMAL INSDEL CODES FROM ALMOST MDS CODES

JIAHUI WANG and YANG DING

(Received 9 December 2024; accepted 18 December 2024; first published online 17 February 2025)

Abstract

Determining the insdel distance of linear codes is a very challenging problem. Recently, Ji *et al.* ['Strict half-Singleton bound, strict direct upper bound for linear insertion-deletion codes and optimal codes', *IEEE Trans. Inform. Theory* **69**(5) (2023), 2900–2910] proposed their strict half-Singleton bound for linear codes that do not include the all-one vector. A natural question asks for linear codes with both large Hamming distance and insdel distance. Almost maximum distance separable (MDS) codes are a special kind of linear code with good Hamming error-correcting capability. We present a sufficient condition for the insdel distance of almost MDS codes to be bounded by the strict half-Singleton bound. In addition, we construct a class of two-dimensional near MDS codes that achieve the strict half-Singleton bound using twisted Reed–Solomon codes. Finally, we present a construction of optimal almost MDS insdel codes from Reed–Solomon codes for large dimensions.

2020 *Mathematics subject classification*: primary 94B05; secondary 94B60.

Keywords and phrases: insdel codes, almost MDS code, near MDS codes, strict half-Singleton bound, twisted Reed–Solomon codes.

1. Introduction

Communication systems suffer not only from substitution errors, but also from synchronisation errors such as insertion and deletion errors. Insertion-deletion codes (insdel codes) were first proposed by Levenshtein [20] to correct insertion and deletion errors and are extensively employed in DNA-based data storage, racetrack memories and language processing [2, 3, 7, 14, 18, 22, 25]. To correct insdel errors, the concept of insdel distance was introduced in [20]. Many studies are concerned with the tradeoff between the minimum insdel distance d_I and code size M [4–6, 12, 15, 21, 26]. Most recently, Ji *et al.* [19] proposed a strict half-Singleton bound, indicating that $d_I \leq 2(n - 2k + 1)$ if the linear code with dimension k does not contain the all-ones codeword.

Codes that achieve these bounds are called optimal codes. Real communication systems not only have to resist synchronous errors but also substitution errors.

This work was supported by the National Natural Science Foundation of China under grant numbers 12171309 and 12271084.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Australian Mathematical Publishing Association Inc.

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

Therefore, many constructions are based on utilising Reed–Solomon codes, which are maximum distance separable (MDS) codes [4, 9, 10, 12, 21]. Almost MDS (AMDS) codes and near MDS (NMDS) codes are special classes of linear code that can be regarded as relaxed versions of MDS codes and also have good capabilities for correcting adversary errors. We consider the insdel correcting capabilities of AMDS codes in this paper.

The rest of this paper is organised as follows. Section 2 is a brief introduction of AMDS codes and insdel codes. In Section 3, we provide a bound for the insdel distance of AMDS codes. In Section 4, we first give a class of optimal two-dimensional NMDS insdel codes and then we proceed to construct optimal AMDS insdel codes for arbitrary dimensions. We demonstrate that the bound in Theorem 3.1 is tight. A problem worth considering is to determine the upper bound on the insdel distance for AMDS codes that do not satisfy the conditions of Theorem 3.1.

2. Preliminaries

Let $\mathbb{F}_q$ denote the finite field with q elements and let $\mathbb{F}_q^n$ be the set of all vectors of length n over $\mathbb{F}_q$. For a nonnegative integer t with $t \le q$, a t-subset $A \subseteq \mathbb{F}_q$ is called a *zero-sum t-subset* if $\sum_{a\in A} a = 0$. For a vector $\mathbf{a} \in \mathbb{F}_q^n$, its *support* is defined by $\mathrm{supp}(\mathbf{a}) = \{i : a_i \neq 0\}$. The *Hamming weight* $w_H(\mathbf{a})$ of $\mathbf{a}$ is the cardinality of its support. A linear code over $\mathbb{F}_q$ of length n, dimension k and minimum Hamming distance d_H is called an $[n, k, d_H]_q$ linear code. Two linear codes over $\mathbb{F}_q$ are considered *equivalent* if one can be derived from the other by a combination of coordinate permutations and scalar multiplication, where each coordinate is multiplied by a nonzero element in $\mathbb{F}_q$. It is evident that equivalent codes have the same parameters.

Now we introduce AMDS and NMDS codes.

DEFINITION 2.1. An $[n, k, d_H]_q$ linear code is AMDS if $d_H = n - k$. An $[n, k, d_H]_q$ linear code is NMDS if both C and the dual of C are AMDS codes.

The next result gives properties of the generator matrix of AMDS and NMDS codes.

LEMMA 2.2 [11]. *Let G be a $k \times n$ matrix over $\mathbb{F}_q$. Assume that:*

(N1) *every $k-1$ columns of G are linearly independent;*
(N2) *there exist k linearly dependent columns in G; and*
(N3) *any $k+1$ columns of G have full rank.*

Let C be a linear code with generator matrix G. Then:

(1) *C is an AMDS code if and only if G satisfies (N2) and (N3); and*
(2) *C is an NMDS code if and only if G satisfies (N1), (N2) and (N3).*

There are many known constructions of NMDS codes. Roth–Lempel in [24] give a class of NMDS codes as follows. Let n and k be two integers such that $3 \le k \le n \le q$. Let $A = \{a_1, \ldots, a_n\}$ be an n-subset of $\mathbb{F}_q$. Then an $[n+2, k]$ *Roth–Lempel* code,

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

$RL(A, k, n+2)$, over $\mathbb{F}_q$ is defined by the generator matrix

$$G_n = \begin{pmatrix} 1 & 1 & \cdots & 1 & 0 & 0 \\ a_1 & a_2 & \cdots & a_n & 0 & 0 \\ a_1^2 & a_2^2 & \cdots & a_n^2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_1^{k-2} & a_2^{k-2} & \cdots & a_n^{k-2} & 0 & 1 \\ a_1^{k-1} & a_2^{k-1} & \cdots & a_n^{k-1} & 1 & 0 \end{pmatrix}.$$

It has been proved that the linear code $RL(A, k, n+2)$ is an NMDS code if and only if A contains a zero-sum $(k-1)$-subset [16].

Now we introduce the insdel distance and insdel codes.

DEFINITION 2.3. Let $\mathbf{a} = (a_1, a_2, \ldots, a_n), \mathbf{b} = (b_1, b_2, \ldots, b_n)$ be two vectors in $\mathbb{F}_q^n$.

(1) The insdel distance $d_I(\mathbf{a}, \mathbf{b})$ between $\mathbf{a}$ and $\mathbf{b}$ is the minimal number of insertions and deletions that are needed to transform $\mathbf{a}$ into $\mathbf{b}$.

(2) A common subsequence of $\mathbf{a}$ and $\mathbf{b}$ is a sequence $(c_1, c_2, \ldots, c_m)$ such that $c_s = a_{i_s} = b_{j_s}$, where $1 \le s \le m, 1 \le i_1 < i_2 < \cdots < i_m \le n$ and $1 \le j_1 < j_2 < \cdots < j_m \le n$. We use $l(\mathbf{a}, \mathbf{b})$ to denote the length of a longest common subsequence of $\mathbf{a}$ and $\mathbf{b}$.

It can be confirmed that $d_I(\mathbf{a}, \mathbf{b})$ is a metric on $\mathbb{F}_q^n$. Moreover, the insdel distance between any two vectors can be determined by their longest common subsequences.

LEMMA 2.4 [12]. *Let $\mathbf{a}, \boldsymbol{b} \in \mathbb{F}_q^n$. Then $d_I(\mathbf{a}, \boldsymbol{b}) = 2n - 2l(\mathbf{a}, \boldsymbol{b})$.*

The insdel distance of a linear insdel code C over $\mathbb{F}_q$ is defined as the minimum insdel distance between any two different codewords in C, that is,

$$d_I(C) = \min_{\mathbf{a}, \mathbf{b} \in C, \mathbf{a} \neq \mathbf{b}} \{d_I(\mathbf{a}, \mathbf{b})\}.$$

It is easy to show that $d_I(C) \le 2d_H(C)$ for arbitrary insdel codes. Haeupler–Shahrasbi [15] gave the direct Singleton bound $d_I(C) \le 2(n - 2k + 1)$. After that, Chen–Zhang [4] improved the direct Singleton bound of linear insdel codes. Further, Cheng *et al.* [6] and Chen [5] introduced the half-Singleton bound for linear insdel codes.

LEMMA 2.5 [5, 6]. *Let C be an $[n, k]$ linear code over $\mathbb{F}_q$ with $n > k \ge 1$. Then $d_I(C) \le \max\{2(n - 2k + 2), 2\}$.*

According to Lemma 2.5, the linear codes can correct insdel errors only for $n \ge 2k$. This conclusion was also proposed in [1]. Recently, Ji *et al.* [19] proposed a strict half-Singleton bound on the insdel distance for linear insdel codes without the codeword with all ones.

LEMMA 2.6 (Strict half-Singleton bound, [19]). *Let C be an $[n, k]$ linear code over $\mathbb{F}_q$ with $n > k \ge 1$. If $\mathbf{1} = (1, \ldots, 1) \notin C$, then the insdel distance of C satisfies*

$$d_I(C) \le \max\{2(n - 2k + 1), 2\}.$$

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

3. Upper bound of the insdel distance for AMDS code

In this section, we provide a sufficient condition for the insdel distance of an AMDS code to be bounded by the strict half-Singleton bound. For a positive integer n, we denote $[n] = \{1, 2, \ldots, n\}$. For a linear $[n, k]_q$ code $C \subset \mathbb{F}_q^n$, the subset $S = \{i_1, i_2, \ldots, i_h\} \subset [n]$ of h coordinate positions is called an *information free coordinate subset* if the natural projection $\phi_S : C \to \mathbb{F}_q^h$, defined by $\phi_S((c_1, \ldots, c_n)) = (c_{i_1}, \ldots, c_{i_h})$, is a surjection. Clearly, $h \le k$. When $h = k$, S is called an *information set*.

For positive integers n and s with $s \le n$, we define $[n]^s$ as the set of all vectors of length s whose coordinates are from $[n]$. We say that a vector $I = (I_1, I_2, \ldots, I_s) \in [n]^s$ is an *increasing vector* if $1 \le I_1 < I_2 < \cdots < I_s \le n$. Let $I \cap J$ be an increasing vector formed by the corresponding equal components of I and J: that is, $I \cap J = (r_1, r_2, \ldots, r_t), t \le s$. Here, $r_i = I_{e_i} = J_{e_i}$ represents the e_ith component of I and J for $1 \le i \le t$. Let $|I \cap J|$ denote the length of the vector $I \cap J$.

THEOREM 3.1. *Let C be a linear $[n, k]_q$ AMDS code with $n \ge \max\{3k - 1, 2k + 1\}$. Suppose that there exists a minimum Hamming weight codeword $\boldsymbol{c} \in C$ with k consecutive indices in* supp($\boldsymbol{c}$). *Then*

$$d_I(C) \le 2(n - 2k + 1).$$

PROOF. Let G be a generator matrix of C. Let $J = [n] \setminus \mathrm{supp}(\mathbf{c})$ and let I be the set of k consecutive indices in supp($\mathbf{c}$). We have $|I| = k$. We discuss the insdel distance of C in two cases.

Case 1. If $\mathrm{rank}(G_I) = k$, then I is an information set. We can assume that $I = \{i, i + 1, \ldots, i + k - 1\}$, where $1 \le i < i + 1 < \cdots < i + k - 1 \le n$. Since I is an information set, we can find a codeword $\mathbf{b} \in C$ satisfying that $\mathbf{b}_I = (b_i, b_{i+1}, \ldots, b_{i+k-1})$ with the coordinate values

$$\begin{aligned} b_{i+1} &= b_i + c_i, \\ b_{i+2} &= b_{i+1} + c_{i+1}, \\ &\cdots \\ b_{i+k-1} &= b_{i+k-2} + c_{i+k-2}. \end{aligned}$$

Here, b_i is an arbitrary element in $\mathbb{F}_q$. Then

$$\begin{gathered} (\mathbf{b} + \mathbf{c})_I = (b_i + c_i, \ldots, b_{i+k-1} + c_{i+k-1}) = (b_{i+1}, \ldots, b_{i+k-1}, b_{i+k-1} + c_{i+k-1}), \\ (\mathbf{b} + \mathbf{c})_J = \mathbf{b}_J + \mathbf{c}_J = \mathbf{b}_J. \end{gathered}$$

We have $l(\mathbf{b}, \mathbf{b} + \mathbf{c}) \ge l(\mathbf{b}_I, (\mathbf{b} + \mathbf{c})_I) + k = k - 1 + k = 2k - 1$. Therefore,

$$d_I(C) \le d_I(\mathbf{b}, \mathbf{b} + \mathbf{c}) = 2n - 2l(\mathbf{b}, \mathbf{b} + \mathbf{c}) \le 2(n - 2k + 1).$$

Case 2. If $\mathrm{rank}(G_I) < k$, then there exists $\mathbf{0} \ne \mathbf{m} \in \mathbb{F}_q^k$ such that $\mathbf{m}G_I = \mathbf{0}$. Let $\mathbf{d} = \mathbf{m}G$. Then $w_H(\mathbf{d}) = n - k$. Since $n \ge 3k - 1$, by the pigeonhole principle, we can guarantee that there is $S \subset [n] \setminus I$ such that S is a consecutive index set with size k. We analyse this case by considering two subcases.

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

Case 2.1. If $\operatorname{rank}(G_S) = k$, then S is an information set and the conclusion follows from Case 1.

Case 2.2. If $\operatorname{rank}(G_S) < k$, since $n \geq \max\{3k-1, 2k+1\}$, we can find a consecutive index set $D \subset [n] \setminus I$ or $[n] \setminus S$ with size $k+1$. Assume that $D \subset [n] \setminus S$. Then we must have a consecutive information set in D because any $k+1$ columns of G have full rank. Then we can complete the proof in a similar way to the proof in Case 1. □

If the first or the last k columns of a generator matrix of an AMDS code are linearly dependent, then the bound is also true.

COROLLARY 3.2. *Let C be a linear $[n,k]_q$ AMDS code with $n \geq 2k+1$. If the first or the last k columns of G are linearly dependent, then $d_I(C) \leq 2(n-2k+1)$.*

PROOF. Without loss of generality, we assume that the first k columns of G are linearly dependent. Let $J = \{1, 2, \ldots, k\}$ and $S = [n] \setminus J$. Consequently, there exists $\mathbf{0} \neq \mathbf{m} \in \mathbb{F}_q^k$ such that $\mathbf{m}G = \mathbf{0}$. If $\mathbf{c} = \mathbf{m}G$, then $w_H(\mathbf{c}) = n-k$. Since $n \geq 2k+1$, we can find an information set $I \subset S$. The conclusion follows by a similar proof to that in Case 1 of Theorem 3.1. □

Now we can give two examples of optimal insdel codes with respect to the strict half-Singleton bound from Roth–Lempel codes.

EXAMPLE 3.3. Let $A = \{0, 1, 4, 2, 3\}$ and let $RL(A, 3, 7)$ be a linear code over $\mathbb{F}_5$ with generator matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 4 & 2 & 3 & 0 & 1 \\ 0 & 1 & 1 & 4 & 4 & 1 & 0 \end{pmatrix}.$$

With the help of MAGMA [23], we can calculate that $RL(A, 3, 7)$ is an optimal $[7, 3]_5$ NMDS insdel code with $d_I = 4$. All the MAGMA codes mentioned in this and the following examples are available by contacting the authors.

EXAMPLE 3.4. Let $q = 25$ and let ω be a primitive element of $\mathbb{F}_q$. Let C be a linear code over $\mathbb{F}_q$ with generator matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ \omega & 1 & \omega^2 & 2 & 0 & 3 & 0 & \omega^{11} & 0 \\ \omega^2 & 1 & \omega^4 & 4 & 0 & 4 & 0 & \omega^{22} & 1 \\ \omega^3 & 1 & \omega^6 & 3 & 1 & 2 & 0 & \omega^9 & 0 \end{pmatrix}.$$

With the help of MAGMA, we can calculate that C is an optimal $[9, 4]_{25}$ NMDS insdel code with $d_I = 4$.

REMARK 3.5. Let $A = \{\omega, \omega^2, \omega^{11}, 0, 1, 2, 3\}$. It is clear that C in Example 3.4 is equivalent to $RL(A, 4, 9)$. However, we can calculate $d_I(RL(A, 4, 9)) = 2$, which shows that the coordinate orderings strongly affect the insdel distance of linear codes.

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

In fact, the insdel distance of a linear $[n, k]$ AMDS code is bounded by the strict half-Singleton bound when n is large enough.

COROLLARY 3.6. *If C be a linear $[n, k]_q$ AMDS code with $n \geq k^2 + k$, then we must have $d_I(C) \leq 2(n - 2k + 1)$.*

PROOF. Suppose that $\mathbf{c} \in C$ with $w_H(\mathbf{c}) = n - k$. Given that $n \geq k^2 + k$, by the pigeonhole principle, we can guarantee that there exists $S \subset \mathrm{supp}(\mathbf{c})$ such that S is a consecutive index set with size k. According to Theorem 3.1, $d_I(C) \leq 2(n - 2k + 1)$. □

REMARK 3.7. If C is a linear $[n, k]_q$ AMDS code, then C can achieve the half-Singleton bound only for $n \leq k^2 + k - 1$.

From Remark 3.7, we know that we can find optimal two-dimensional NMDS insdel codes with respect to the half-Singleton bound only for $n \leq 5$. We close this section with an example.

EXAMPLE 3.8. Let $q = 3^{25}$ and let ω be a primitive element of $\mathbb{F}_q$. Let

$$C = \{(f(\omega), f(\omega^2), f(\omega^4), f(-\omega^2), f(\omega^8)) \mid f(x) = a_0 + a_1 x^2, a_0, a_1 \in \mathbb{F}_q\}.$$

With the help of MAGMA, we can show that C is an optimal $[5, 2, d_I = 6]_q$ NMDS insdel code with respect to the half-Singleton bound.

4. Constructions of optimal AMDS insdel codes

This section is composed of two subsections. In Section 4.1, we present a class of two-dimensional optimal NMDS insdel codes that achieve the strict half-Singleton bound by employing twisted Reed–Solomon codes. In Section 4.2, we construct optimal AMDS insdel codes with respect to the strict half-Singleton bound for arbitrary dimensions. These results show that the strict half-Singleton bound is tight for AMDS codes that satisfy the conditions of Theorem 3.1.

4.1. Optimal two-dimensional NMDS insdel codes. In this subsection, we focus on two-dimensional NMDS insdel codes. We utilise twisted Reed–Solomon (TRS) codes to construct a class of optimal NMDS insdel codes with respect to the strict half-Singleton bound. Now we introduce the definition of TRS codes.

Let $\mathbf{a} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are distinct elements in $\mathbb{F}_q$ and $\eta \in \mathbb{F}_q^*$. Let $\mathbb{F}_q[x]$ be the polynomial ring over $\mathbb{F}_q$. Consider the TRS code $C_k(\mathbf{a}, \eta)$ with length n and dimension k, where

$$C_k(\mathbf{a}, \eta) = \left\{ \mathbf{c}_f = (f(\alpha_1), \ldots, f(\alpha_n)) \,\middle|\, f(x) = \sum_{i=0}^{k-1} a_i x^i + \eta a_{k-1} x^k \in \mathbb{F}_q[x] \right\}.$$

Let $S_2 = \{\alpha_i + \alpha_j \mid 1 \leq i < j \leq n\}$. Huang *et al.* [17] studied the NMDS property of $C_2(\mathbf{a}, \eta)$.

LEMMA 4.1 [17]. *The code $C_2(\mathbf{a}, \eta)$ is NMDS if and only if $-\eta^{-1} \in S_2$.*

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

To construct our codes, we also need the concept of Sidon sets.

DEFINITION 4.2. A set D of integers is called a Sidon set if all pairwise sums of its elements are distinct. Equivalently, the equation $a + b = c + d$ has only the trivial solution $\{a, b\} = \{c, d\}$ in D.

We use the TRS code to construct a family of two-dimensional insdel NMDS codes. Let $q = l^2$, where $l = p^e$ is a prime power and $e > 1$ is a positive integer. Let θ be the primitive element of $\mathbb{F}_l$, that is, $\mathbb{F}_l = \mathbb{F}_p(\theta)$. Let $n \geq 5$ be an integer and let $i_1, i_3, i_4, \ldots, i_n$ be $n-1$ integers such that $0 \leq i_1 < i_3 < \cdots < i_n \leq l-2$. Choose $\eta \in \mathbb{F}_l^*$ such that $1 + 4\eta\theta^{i_1} \neq 0$. Let α_1, α_2 be the two roots of the polynomial $x + \eta x^2 = \theta^{i_1}$ and let $\alpha_3, \alpha_4, \ldots, \alpha_n$ be the roots of the polynomials $x + \eta x^2 = \theta^{i_t} (t = 3, 4, \ldots, n)$, respectively. Obviously, $\alpha_1, \alpha_2 \ldots, \alpha_n$ are distinct elements in $\mathbb{F}_q^*$. Then

$$C_2(\mathbf{a}, \eta) = \{\mathbf{c}_f = (f(\alpha_1), \ldots, f(\alpha_n)) \mid f(x) = \lambda + \mu(x + \eta x^2), \lambda, \mu \in \mathbb{F}_q\}$$

is a two-dimensional TRS code of length n.

The following theorem provides a sufficient condition for the insdel distance of $C_2(\mathbf{a}, \eta)$ to achieve the strict half-Singleton bound.

THEOREM 4.3. *If $D = \{i_1, i_3, \ldots, i_n\} \subseteq \{0, 1, \ldots, l-2\}$ satisfies:*

(1) *D is a Sidon set; and*
(2) *$i_{n-1} + i_n < e$,*

then the minimum insdel distance of $C_2(\mathbf{a}, \eta)$ is equal to $2n-6$: that is, $C_2(\mathbf{a}, \eta)$ is an optimal NMDS insdel code with respect to the strict half-Singleton bound.

PROOF. According to Lemma 4.1, $C_2(\mathbf{a}, \eta)$ is a two-dimensional NMDS code since

$$-\eta^{-1} = \alpha_1 + \alpha_2 \in S_2.$$

Moreover, the first two columns of the generator matrix of $C_2(\mathbf{a}, \eta)$ are

$$\begin{pmatrix} 1 \\ \theta^{i_1} \end{pmatrix}, \begin{pmatrix} 1 \\ \theta^{i_1} \end{pmatrix}.$$

By Corollary 3.2, $d_I(C_2(\mathbf{a}, \eta)) \leq 2n-6$. It remains to show that $d_I(C_2(\mathbf{a}, \eta)) \geq 2n-6$.

Let

$$f(x) = \lambda_0 + \mu_0(x + \eta x^2), g(x) = \lambda_1 + \mu_1(x + \eta x^2),$$

where $\lambda_0, \mu_0, \lambda_1, \mu_1 \in \mathbb{F}_q$. From Lemma 2.4, it is enough to prove the following claim.

Claim. $l(\mathbf{c}_f, \mathbf{c}_g) \leq 3$ for any $f(x) \neq g(x)$.

Suppose that $f(x) \neq g(x)$ with $l(\mathbf{c}_f, \mathbf{c}_g) \geq 4$. Then there exist eight integers $k_1, k_2, k_3, k_4, r_1, r_2, r_3, r_4$ with $1 \leq k_1 < k_2 < k_3 < k_4 \leq n, 1 \leq r_1 < r_2 < r_3 < r_4 \leq n$

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

such that

$$\begin{cases} \lambda_0 + \mu_0(\alpha_{k_1} + \eta\alpha_{k_1}^2) = \lambda_1 + \mu_1(\alpha_{r_1} + \eta\alpha_{r_1}^2), \\ \lambda_0 + \mu_0(\alpha_{k_2} + \eta\alpha_{k_2}^2) = \lambda_1 + \mu_1(\alpha_{r_2} + \eta\alpha_{r_2}^2), \\ \lambda_0 + \mu_0(\alpha_{k_3} + \eta\alpha_{k_3}^2) = \lambda_1 + \mu_1(\alpha_{r_3} + \eta\alpha_{r_3}^2), \\ \lambda_0 + \mu_0(\alpha_{k_4} + \eta\alpha_{k_4}^2) = \lambda_1 + \mu_1(\alpha_{r_4} + \eta\alpha_{r_4}^2). \end{cases}$$

Then

$$\begin{cases} \lambda_0 + \mu_0\theta^{i_{k_2}} = \lambda_1 + \mu_1\theta^{i_{r_2}}, \\ \lambda_0 + \mu_0\theta^{i_{k_3}} = \lambda_1 + \mu_1\theta^{i_{r_3}}, \\ \lambda_0 + \mu_0\theta^{i_{k_4}} = \lambda_1 + \mu_1\theta^{i_{r_4}}, \end{cases}$$

with $k_2 < k_3 < k_4$ and $r_2 < r_3 < r_4$, which is equivalent to saying that

$$\begin{pmatrix} \theta^{i_{k_2}} & \theta^{i_{r_2}} & 1 \\ \theta^{i_{k_3}} & \theta^{i_{r_3}} & 1 \\ \theta^{i_{k_4}} & \theta^{i_{r_4}} & 1 \end{pmatrix} \begin{pmatrix} \mu_0 \\ -\mu_1 \\ \lambda_0 - \lambda_1 \end{pmatrix} = \mathbf{0}.$$

Since $f(x) \neq g(x)$, we know that $\mu_0, -\mu_1$ and $\lambda_0 - \lambda_1$ are not all equal to 0. Thus, we must have

$$\begin{vmatrix} \theta^{i_{k_2}} & \theta^{i_{r_2}} & 1 \\ \theta^{i_{k_3}} & \theta^{i_{r_3}} & 1 \\ \theta^{i_{k_4}} & \theta^{i_{r_4}} & 1 \end{vmatrix} = 0.$$

Consider the polynomial

$$M(x) = \begin{vmatrix} x^{i_{k_2}} & x^{i_{r_2}} & 1 \\ x^{i_{k_3}} & x^{i_{r_3}} & 1 \\ x^{i_{k_4}} & x^{i_{r_4}} & 1 \end{vmatrix} = x^{i_{k_2}+i_{r_3}} + x^{i_{k_3}+i_{r_4}} + x^{i_{k_4}+i_{r_2}} - x^{i_{k_2}+i_{r_4}} - x^{i_{k_3}+i_{r_2}} - x^{i_{k_4}+i_{r_3}}.$$

It is clear that $M(x) \in \mathbb{F}_p[x]$ and $M(\theta) = 0$. We want to prove that $M(x) \neq 0$. Observing that $k_2 < k_3 < k_4$ and $r_2 < r_3 < r_4$, we have

$$\begin{aligned} i_{k_3} + i_{r_4} > \max\{i_{k_2} + i_{r_3}, i_{k_2} + i_{r_4}\}, \quad i_{k_4} + i_{r_3} > \max\{i_{k_3} + i_{r_2}, i_{k_4} + i_{r_3}\}, \\ i_{k_2} + i_{r_3} < \min\{i_{k_3} + i_{r_4}, i_{k_2} + i_{r_4}\}, \quad i_{k_3} + i_{r_2} < \min\{i_{k_4} + i_{r_2}, i_{k_4} + i_{r_3}\}. \end{aligned}$$

Thus, the term with the maximum degree in the polynomial $M(x)$ is $x^{i_{k_3}+i_{r_4}}$ or $-x^{i_{k_4}+i_{r_3}}$ and the term with the minimum degree in the polynomial $M(x)$ is $x^{i_{k_2}+i_{r_3}}$ or $-x^{i_{k_3}+i_{r_2}}$. If $M(x) = 0$, then we must have

$$i_{k_3} + i_{r_4} = i_{k_4} + i_{r_3}, \quad i_{k_2} + i_{r_3} = i_{k_3} + i_{r_2}.$$

Since D is a Sidon set, this implies that

$$\{i_{k_3}, i_{r_4}\} = \{i_{k_4}, i_{r_3}\}, \quad \{i_{k_2}, i_{r_3}\} = \{i_{k_3}, i_{r_2}\}.$$

Since $i_{k_2} < i_{k_3} < i_{k_4}$ and $i_{r_2} < i_{r_3} < i_{r_4}$, it follows that $i_{k_t} = i_{r_t}$ for $t = 2, 3, 4$. This means that $f(\alpha_{i_{k_t}}) = g(\alpha_{i_{k_t}})$ for $t = 2, 3, 4$, forcing $f(x) = g(x)$, which is a contradiction. Therefore, we conclude that $M(x) \neq 0$.

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

Thus, we have $0 \neq M(x) \in \mathbb{F}_p[x]$ with $M(\theta) = 0$. However, since θ is a primitive element of $\mathbb{F}_l = \mathbb{F}_{p^e}$ and $\deg M(x) \leq i_{n-1} + i_n < e$, we arrive at a contradiction. This completes the proof of the claim $l(\mathbf{c}_f, \mathbf{c}_g) \leq 3$ for any distinct codewords $\mathbf{c}_f, \mathbf{c}_g \in C_2(\mathbf{a}, \eta)$.

Consequently, $d_I(C_2(\mathbf{a}, \eta)) \geq 2n - 6$. This completes the proof. □

REMARK 4.4. (1) Let $S(n)$ denote the maximum size of a Sidon set D when $D \subset [n]$. Erdős and Turán [13] proved that $S(n) \leq \sqrt{n} + O(n^{1/4})$. Thus, we require a field size at least $\Omega(\exp(n^2))$ in Theorem 4.3.

(2) It is easy to check that $\mathbf{1} = (1, \ldots, 1) \in C_2(\mathbf{a}, \eta)$. This observation indicates that $C_2(\mathbf{a}, \eta)$ does not satisfy the condition presented in Lemma 2.6.

4.2. Optimal AMDS insdel codes. In this subsection, we aim to generalise the idea for optimal two-dimensional NMDS insdel codes and construct optimal AMDS insdel codes for arbitrary dimensions.

Recently, Con *et al.* [8] showed that random Reed–Solomon codes approach the half-Singleton bound with high probability. Let C be an $[n, k]_q$ Reed–Solomon code with generator matrix

$$G = (\mathbf{g_1}, \mathbf{g_2}, \ldots, \mathbf{g_n})$$

that achieves the half-Singleton bound, where $n \geq 2k$. Let C_1 be the linear code spanned by the row vectors of

$$G_1 = (\mathbf{g_1}, G) = (\mathbf{g_1}, \mathbf{g_1}, \mathbf{g_2}, \ldots, \mathbf{g_n}).$$

LEMMA 4.5. *C_1 is an $[n + 1, k]_q$ AMDS code.*

PROOF. Consider the map

$$\varphi : C \to C_1$$
$$(c_1, c_2, \ldots, c_n) \mapsto (c_1, c_1, c_2, \ldots, c_n).$$

It is clear that φ is a bijection, and thus $\dim(C_1) = k$. Since G_1 is the generator matrix of C_1 and any k columns of G have full rank, it follows that any $k + 1$ columns of G_1 have full rank and the first k columns of G_1 are linearly dependent. This implies that C_1 is an $[n + 1, k]_q$ AMDS code according to Lemma 2.2. □

REMARK 4.6. It is easy to verify that $\mathbf{c} = (1, -1, 0, \ldots, 0)$ is a codeword of the dual of C_1. This indicates that C_1 cannot be an NMDS code when $k \geq 3$.

According to the bijection φ between C and C_1, it is easy to see the following lemma.

LEMMA 4.7. *For any distinct codewords $\mathbf{a}, \boldsymbol{b} \in C_1$,*

$$l(\varphi^{-1}(\mathbf{a}), \varphi^{-1}(\boldsymbol{b})) \leq l(\mathbf{a}, \boldsymbol{b}) \leq l(\varphi^{-1}(\mathbf{a}), \varphi^{-1}(\boldsymbol{b})) + 1.$$

Next, we show that C_1 is an optimal AMDS insdel code.

THEOREM 4.8. *C_1 is an $[n + 1, k, d_I = 2(n + 1 - 2k + 1)]_q$ AMDS insdel code. Consequently, C_1 is an optimal AMDS code with respect to the strict half-Singleton bound.*

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

PROOF. It remains to show that $d_I(C_1) = 2(n - 2k + 2)$. According to Corollary 3.2, we have $d_I(C_1) \le 2(n - 2k + 2)$ since the first k columns of G_1 are linearly dependent.

Suppose that $d_I(C_1) \le 2(n - 2k + 1)$. Then there exist two distinct codewords $\mathbf{a}, \mathbf{b}$ in C_1 such that

$$l(\mathbf{a}, \mathbf{b}) = \tfrac{1}{2}(2(n + 1) - d_I(\mathbf{a}, \mathbf{b})) \ge 2(n + 1) - 2(n - 2k + 1) = 2k.$$

From this, $l(\varphi^{-1}(\mathbf{a}), \varphi^{-1}(\mathbf{b})) \ge l(\mathbf{a}, \mathbf{b}) - 1 = 2k - 1$, which indicates that

$$d_I(C) \le d_I(\varphi^{-1}(\mathbf{a}), \varphi^{-1}(\mathbf{b})) = 2n - 2l(\varphi^{-1}(\mathbf{a}), \varphi^{-1}(\mathbf{b})) \le 2(n - 2k + 1).$$

This contradicts $d_I(C) = 2(n - 2k + 2)$. Thus, we have $d_I(C_1) = 2(n - 2k + 2)$. □

References

[1] K. A. S. Abdel-Ghaffar, H. C. Ferreira and L. Cheng, 'Correcting deletions using linear and cyclic codes', *IEEE Trans. Inform. Theory* **56**(10) (2010), 5223–5234.

[2] E. Brill and R. C. Moore, 'An improved error model for noisy channel spelling correction', in: *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics* (ed. H. Iida) (Association for Computational Linguistics, Kerrville, TX, 2000), 286–293.

[3] Y. M. Chee, H. M. Kiah, A. Vardy, E. Yaakobi and V. K. Vu, 'Codes correcting position errors in racetrack memories', in: *2017 IEEE Information Theory Workshop*, Kaohsiung, Taiwan, 2017 (eds. P.-N. Chen, G. Kramer and C.-P. Li) (IEEE, New York, 2018), 161–165.

[4] B. Chen and G. Zhang, 'Improved Singleton bound on insertion-deletion codes and optimal constructions', *IEEE Trans. Inform. Theory* **68**(5) (2022), 3028–3033.

[5] H. Chen, 'Coordinate-ordering-free upper bounds for linear insertion-deletion codes', *IEEE Trans. Inform. Theory* **68**(8) (2022), 5126–5132.

[6] K. Cheng, V. Guruswami, B. Haeupler and X. Li, 'Efficient linear and affine codes for correcting insertions/deletions', *SIAM J. Discrete Math.* **37**(2) (2023), 748–778.

[7] K. Cheng, Z. Jin, X. Li and K. Wu, 'Deterministic document exchange protocols and almost optimal binary codes for edit errors', in: *IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, Paris France (ed. M. Thorup) (IEEE, New York, 2018), 200–211.

[8] R. Con, Z. Guo, R. Li and Z. Zhang, 'Random Reed–Solomon codes achieve the half-singleton bound for insertions and deletions over linear-sized alphabets', Preprint, 2024, arXiv:2407.07299.

[9] R. Con, A. Shpilka and I. Tamo, 'Reed–Solomon codes against adversarial insertions and deletions', in: *2022 IEEE International Symposium on Information Theory (ISIT)*, Espoo, Finland (eds. T. Charalambous, C. Hollanti and D. Tirkkonen) (IEEE, New York, 2022), 2940–2945.

[10] R. Con, A. Shpilka and I. Tamo, 'Optimal two-dimensional Reed–Solomon codes correcting insertions and deletions', *IEEE Trans. Inform. Theory* **70**(7) (2024), 5012–5016.

[11] S. M. Dodunekov and I. N. Landgev, 'Near MDS codes over some small fields', *Discrete Math.* **213**(1) (2000), 55–65.

[12] T. D. Duc, S. Liu, I. Tjuawinata and C. Xing, 'Explicit constructions of two-dimensional Reed–Solomon codes in high insertion and deletion noise regime', *IEEE Trans. Inform. Theory* **67**(5) (2021), 2808–2820.

[13] P. Erdős and P. Turán, 'On a problem of Sidon in additive number theory, and on some related problems', *J. Lond. Math. Soc. (2)* **16**(4) (1941), 212–215.

[14] B. Haeupler, 'Optimal document exchange and new codes for insertions and deletions', in: *IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, Baltimore, MD (ed D. Zuckerman) (IEEE, New York, 2019), 334–347.

[15] B. Haeupler and A. Shahrasbi, 'Synchronization strings: codes for insertions and deletions approaching the Singleton bound', *J. ACM* **68**(5) (2021), Article no. 36, 39 pages.

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press

[16] D. Han and C. Fan, 'Roth–Lempel NMDS codes of non-elliptic-curve type', *IEEE Trans. Inform. Theory* **69**(9) (2023), 5670–5675.
[17] D. Huang, Q. Yue, Y. Niu and X. Li, 'MDS or NMDS self-dual codes from twisted generalized Reed–Solomon codes', *Des. Codes Cryptogr.* **89**(9) (2021), 2195–2209.
[18] S. Jain, F. Hassanzadeh, M. Schwartz and J. Bruck, 'Duplication-correcting codes for data storage in the DNA of living organisms', *IEEE Trans. Inform. Theory* **63**(8) (2017), 4996–5010.
[19] Q. Ji, D. Zheng, H. Chen and X. Wang, 'Strict half-Singleton bound, strict direct upper bound for linear insertion-deletion codes and optimal codes', *IEEE Trans. Inform. Theory* **69**(5) (2023), 2900–2910.
[20] V. I. Levenshtein, 'Binary codes capable of correcting deletions, insertions, and reversals', *Soviet Phys. Dokl.* **10** (1965), 707–710.
[21] S. Liu and C. Xing, 'Bounds and constructions for insertion and deletion codes', *IEEE Trans. Inform. Theory* **69**(2) (2023), 928–940.
[22] S. Liu, C. Xing and Y. Zhang, 'Codes with biochemical constraints and single error correction for DNA-based data storage', Preprint, 2023, arXiv:2307.00221.
[23] Magma Computer Algebra System. http://magma.maths.usyd.edu.au/.
[24] R. Roth and A. Lempel, 'A construction of non-Reed–Solomon type MDS codes', *IEEE Trans. Inform. Theory* **35**(3) (1989), 655–657.
[25] C. Schoeny, A. Wachter-Zeh, R. Gabrys and E. Yaakobi, 'Codes correcting a burst of deletions or insertions', *IEEE Trans. Inform. Theory* **63**(4) (2017), 1971–1985.
[26] C. Xie, H. Chen, L. Qu and L. Liu, 'New dimension-independent upper bounds on linear insdel codes', *Adv. Math. Commun.* **18**(6) (2024), 1575–1589.

JIAHUI WANG, Department of Mathematics,
Shanghai University, Shanghai, China
e-mail: 15056330992@shu.edu.cn

YANG DING, Department of Mathematics,
Shanghai University, Shanghai, China
e-mail: dingyang@t.shu.edu.cn

https://doi.org/10.1017/S0004972725000012 Published online by Cambridge University Press