Optimal Two-Dimensional Reed–Solomon Codes Correcting Insertions and Deletions

Roni Con[*]   Amir Shpilka [†]   Itzhak Tamo [‡]

Abstract

Constructing Reed–Solomon (RS) codes that can correct insertions and deletions (insdel errors) has been considered in numerous recent works. For the special case of two-dimensional RS-codes, it is known [CST23] that an $[n, 2]_q$ RS-code that can correct from $n - 3$ insdel errors satisfies that $q = \Omega(n^3)$. On the other hand, there are several known constructions of $[n, 2]_q$ RS-codes that can correct from $n-3$ insdel errors, where the smallest field size is $q = O(n^4)$. In this short paper, we construct $[n, 2]_q$ Reed–Solomon codes that can correct $n - 3$ insdel errors with $q = O(n^3)$, thereby resolving the minimum field size needed for such codes.

[*]Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel. Email: roni.con93@gmail.com
[‡]Department of EE-Systems, Tel Aviv University, Tel Aviv, Israel. Email: zactamo@gmail.com.

1   Introduction

Constructing error-correcting codes against *synchronization* errors has received a lot of attention recently. The most common model for studying synchronization errors is the insertion-deletion model (insdel for short): an insertion error is when a new symbol is inserted between two symbols of the transmitted word. A deletion is when a symbol is removed from the transmitted word. These are errors that affect the length of the received word. For example, over the binary alphabet, when 100110 is transmitted, we may receive the word 11011000, which is obtained from three insertions (1 at the beginning and two 0s at the end) and one deletion (one of the 0's at the beginning of the transmitted word). Generally, insdel errors cause the sending and receiving parties to become "out of sync" which makes them inherently more difficult to deal with.

Insdel errors appear in diverse settings such as optical recording, semiconductor devices, integrated circuits, and synchronous digital communication networks. Due to the importance of the model and our lack of understanding of some basic problems concerning it, constructing efficient codes that can handle insdel errors is the topic of many recent works [HS17, BGZ17, GW17, SWZGY17, CJLW18, Hae19, CGHL21, GHS20, GH21, LTX23] (see also the excellent survey [HS21]). Further, codes for correcting insdel errors attract a lot of attention due to their possible application in correcting errors in DNA-based storage systems [LSWZY19, HMG19, SH⁺22]. This recent increased interest was paved by substantial progress in synthesis and sequencing technologies. The main advantages of DNA-based storage over classical storage technologies are very high data densities and long-term reliability without an electrical supply.

Reed-Solomon codes are the most widely used family of codes in theory and practice. Their extensive use can be credited to their simplicity, as well as their efficient encoding and decoding algorithms. Some of their notable applications include QR codes [Soo08], secret sharing schemes [MS81], and distributed storage systems [TB14, GW16]. As such, it is an important problem to understand whether they can also decode from insdel errors. This problem received a lot of attention recently [SNW02, WMSN04, TSN07, DLTX19, LT21, CZ21, LX22, CST23]. An open problem that was raised in these line of works is what is the minimal field size for which there exist an $[n, k]_q$ RS-code that can correct from $n - 2k + 1$ insdel errors, the maximum possible insdel errors that can be corrected. In this paper, we give an answer to this question for RS-codes of dimension $k = 2$, and show that the minimal field size is $q = \Theta(n^3)$.

1.1   Basic definitions and notation

For an integer k, we denote $[k] = \{1, 2, \ldots, k\}$. Throughout this paper, $\log(x)$ refers to the base-2 logarithm. For a prime power q, we denote with $\mathbb{F}_q$ the field of size q.

A linear code over a field $\mathbb{F}$ is a linear subspace $\mathcal{C} \subseteq \mathbb{F}^n$. The rate of a linear code $\mathcal{C}$ of block length n is $\mathcal{R} = \dim(\mathcal{C})/n$. Every linear code of dimension k can be described as the image of a linear map, which, abusing notation, we also denote with $\mathcal{C}$, i.e., $\mathcal{C} : \mathbb{F}^k \to \mathbb{F}^n$. When $\mathcal{C} \subseteq \mathbb{F}_q^n$ has dimension k we say that it is an $[n, k]_q$ code (or an $[n, k]$ code defined over $\mathbb{F}_q$). The minimal distance of $\mathcal{C}$ with respect to a metric $d(\cdot, \cdot)$ is defined as $\text{dist}_{\mathcal{C}} := \min_{v \neq u \in \mathcal{C}} d(v, u)$. Naturally, we would like the rate to be as large as possible, but there is an inherent tension between the rate of the code and the minimal

distance (or the number of errors that a code can decode from). In this work, we focus on codes against insertions and deletions.

Definition 1.1. *Let s be a string over the alphabet Σ. The operation in which we remove a symbol from s is called a* deletion *and the operation in which we place a new symbol from Σ between two consecutive symbols in s, in the beginning, or at the end of s, is called an* insertion.

We next define Reed-Solomon codes (RS-codes from now on).

Definition 1.2 (Reed-Solomon codes)**.** *Let $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{F}_q$ be distinct points in a finite field $\mathbb{F}_q$ of order $q \geq n$. For $k \leq n$ the $[n, k]_q$ RS-code defined by the evaluation vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is the set of codewords*

$$\{c_f = (f(\alpha_1), \ldots, f(\alpha_n)) \mid f \in \mathbb{F}_q[x], \deg f < k\} \ .$$

Namely, a codeword of an $[n, k]_q$ RS-code is the evaluation vector of some polynomial of degree less than k at n predetermined distinct points. It is well known (and easy to see) that the rate of $[n, k]_q$ RS-code is k/n and the minimal distance, with respect to the Hamming metric, is $n - k + 1$.

1.2 Related work

Linear codes against insdel errors. The basic question of whether there exist good *linear* codes for the insdel model was first addressed in the work of Cheng, Guruswami, Haeupler, and Li [CGHL21]. Specifically, they showed that there are linear codes of rate $R = (1 - \delta)/2 - h(\delta)/\log_2(q)$. that can correct from δ fraction of insdel errors. They also showed an almost matching upper bound which they called the *half-Singleton bound*, given next.

Theorem 1.3 (Half-Singleton bound: Corollary 5.1 in [CGHL21])**.** *Every linear insdel code which is capable of correcting a δ fraction of deletions has rate at most $(1-\delta)/2+o(1)$.*

Remark 1.4. *The following non-asymptotic version of the Half-Singleton bound can be derived from the proof of Corollary 5.1 in [CGHL21]: An $[n, k]$ linear code can correct at most $n - 2k + 1$ insdel errors.*

Cheng et al. [CGHL21] also constructed the first asymptotically good binary linear codes for insdel errors. Their codes have rate $R < 2^{-80}$ and can correct efficiently from $\delta < 1/400$ insdel errors. Then, Con et al. [CST22] constructed linear codes with better rate-distance tradeoffs, and Cheng et al. [CJL+23] constructed asymptotically good codes in the high-noise and high-rate regimes, covering regimes of rate-distance that are not achievable by the codes of [CST22]. We note that the open question of constructing codes over *small alphabets* that achieve the half-Singleton bound is still open.

RS-codes against insdel errors. To the best of our knowledge, Safavi-Naini and Wang [SNW02] were the first to study the performance of RS-codes against insdel errors. They gave an algebraic condition that is sufficient for an RS-code to correct from insdel errors, yet they did not provide any construction.

Work	Field size
[DLTX19]	$\exp(n)$
[LX22]	$O(n^5)$
[CST23]	$O(n^4)$
This work	$O(n^3)$

Table 1: Previous $[n, 2]_q$ RS-code cosntructions that can correct from $n - 3$ insdel errors and their field size.

Wang, McAven, and Safavi-Naini [WMSN04] constructed a $[5, 2]$ RS-code capable of correcting a single deletion. Then, in [TSN07], Tonien and Safavi-Naini constructed an $[n, k]$ generalized-RS-codes capable of correcting from $\log_{k+1} n - 1$ insdel errors. Con, Shpilka, and Tamo [CST23] showed the existence of $[n, k]_q$ RS-codes that can correct from $n - 2k + 1$ where $q = n^{O(k)}$. These codes are the first linear codes that achieve the half-Singleton bound. They also provided deterministic construction of such code over a field of size $n^{k^{O(k)}}$.

Much attention was given to the specific case of 2-dimensional RS-codes. Specifically, the goal is to construct $[n, 2]_q$ RS-codes correcting $n - 3$ insdel errors with the smallest q possible. By Remark 1.4, it should be noted that these codes are optimal for correcting insdel errors, as they correct the maximum possible number of such errors. Numerous constructions were published [DLTX19, CZ21, CST23, LX22] (see table Table 1). The current best construction is due to [CST23] that presented an $[n, 2]_q$ RS-code correcting $n - 3$ insdel errors where $q = O(n^4)$. They also proved that the field size of an $[n, 2]_q$ RS-code correcting $n - 3$ deletions must be $q = \Omega(n^3)$. This work closes the gap and shows that $q = \Theta(n^3)$.

1.3 Our results

Our main result is the following theorem.

Theorem 1.5. *For any $n \geq 3$, there exists an explicit $[n, 2]_q$ RS-code that can correct from $n - 3$ insdel errors, where $q = O(n^3)$.*

As described earlier, an explicit construction of an RS-code amounts to specifying an evaluation vector. In our constructions, we define sets of size n and prove that any ordering of these n points into an evaluation vector defines a two-dimensional RS-code that can correct from $n - 3$.

1.4 An algebraic condition

In this section, we recall the algebraic condition presented in [CST23]. We first make the following definitions: We say that a vector of indices $I \in [n]^s$ is an *increasing* vector if its coordinates are monotonically increasing, i.e., for any $1 \leq i < j \leq s$, $I_i < I_j$, where I_i is the ith coordinate of I. For two vectors $I, J \in [n]^{2k-1}$ with distinct coordinates we define the following (variant of a) vandermonde matrix of order $(2k - 1) \times (2k - 1)$ in the formal variables $\mathbf{X} = (X_1, \ldots, X_n)$:

$$V_{I,J}(\mathbf{X}) = \begin{pmatrix} 1 & X_{I_1} & \dots & X_{I_1}^{k-1} & X_{J_1} & \dots & X_{J_1}^{k-1} \\ 1 & X_{I_2} & \dots & X_{I_2}^{k-1} & X_{J_2} & \dots & X_{J_2}^{k-1} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & X_{I_{2k-1}} & \dots & X_{I_{2k-1}}^{k-1} & X_{J_{2k-1}} & \dots & X_{J_{2k-1}}^{k-1} \end{pmatrix}. \tag{1}$$

Proposition 1.6. *[CST23, Proposition 2.1] Consider the $[n,k]_q$ RS-code defined by an evaluation vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$. If for every two increasing vectors $I, J \in [n]^{2k-1}$ that agree on at most $k-1$ coordinates, it holds that $\det(V_{I,J}(\boldsymbol{\alpha})) \neq 0$, then the code can correct any $n - 2k + 1$ insdel errors. Moreover, if the code can correct any $n - 2k + 1$ insdel errors, then the only possible vectors in $Kernel(V_{I,J}(\boldsymbol{\alpha}))$ are of the form $(0, f_1, \dots, f_{k-1}, -f_1, \dots, -f_{k-1})$.*

2 Explicit construction for $k = 2$ with qubic field size

In this section, we prove Theorem 1.5, which is restated for convenience.

Theorem 1.5. *For any $n \geq 3$, there exists an explicit $[n,2]_q$ RS-code that can correct from $n - 3$ insdel errors, where $q = O(n^3)$.*

The proof of Theorem 1.5 will follow from two code constructions of an $[n,2]$ RS-code that can correct from $n - 3$ insdel errors, over a field of size $O(n^3)$. The first code construction works for any characteristic, whereas the second construction works only for fields of characteristic not equal to 2. However, for these characteristics, given the same field size, the latter construction provides a slightly longer code compared to the first construction.

2.1 First Construction

In this section, we shall present the first construction that works for any finite field. Both the construction and its proof are given in the following proposition.

Proposition 2.1. *Let $\mathcal{A} \subseteq \mathbb{F}_q^*$ be a set of size n such that for any two distinct elements $\delta, \delta' \in \mathcal{A}$, $\delta \neq -\delta'$, and let γ be a root of a degree 3 irreducible polynomial over $\mathbb{F}_q$. Let the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$ be some ordering of the n elements $\delta + \delta^{-1} \cdot \gamma, \delta \in \mathcal{A}$. Then, the $[n,2]$ RS-code defined over $\mathbb{F}_{q^3}$ with the evaluation vector $\boldsymbol{\alpha}$ can correct any $n - 3$ insdel errors.*

Remark 2.2. *Note that if the characteristic of the field $\mathbb{F}_q$ does not equal to 2, then, necessarily, the length of the code is at most $(q-1)/2$. As for each $\delta \in \mathcal{A}$, $-\delta \notin \mathcal{A}$. On the other hand, if the characteristic equals to 2, then it can be as large as $q - 1$. However, in both cases $q = O(n^3)$.*

Proof. Assume towards a contradiction that the claim is false, then by Proposition 1.6 there exist two vectors of distinct evaluation points $(\beta_1, \beta_2, \beta_3), (\beta_4, \beta_5, \beta_6)$, that agree on at most one coordinate, such that,

$$\left| \begin{pmatrix} 1 & \beta_1 & \beta_4 \\ 1 & \beta_2 & \beta_5 \\ 1 & \beta_3 & \beta_6 \end{pmatrix} \right| = 0 \,.$$

Equivalently,
$$(\beta_1 - \beta_2)(\beta_5 - \beta_6) - (\beta_4 - \beta_5)(\beta_2 - \beta_3) = 0 \ .$$
Write $\beta_i := \delta_i + \delta_i^{-1} \cdot \gamma$ for $i \in [6]$ and observe that the LHS is a polynomial in γ of degree less than 3 over $\mathbb{F}_q$. Namely,
$$p_0(\boldsymbol{\delta}) + p_1(\boldsymbol{\delta}) \cdot \gamma + p_2(\boldsymbol{\delta}) \cdot \gamma^2 = 0$$
where,
$$\begin{aligned}
p_0(\boldsymbol{\delta}) &= (\delta_1 - \delta_2)(\delta_5 - \delta_6) - (\delta_2 - \delta_3)(\delta_4 - \delta_5) \\
p_1(\boldsymbol{\delta}) &= (\delta_1 - \delta_2)(\delta_5^{-1} - \delta_6^{-1}) + (\delta_1^{-1} - \delta_2^{-1})(\delta_5 - \delta_6) \\
&\quad - (\delta_2^{-1} - \delta_3^{-1})(\delta_4 - \delta_5) - (\delta_2 - \delta_3)(\delta_4^{-1} - \delta_5^{-1}) \\
p_2(\boldsymbol{\delta}) &= (\delta_1^{-1} - \delta_2^{-1})(\delta_5^{-1} - \delta_6^{-1}) - (\delta_2^{-1} - \delta_3^{-1})(\delta_4^{-1} - \delta_5^{-1}) \ .
\end{aligned}$$
$p_1(\boldsymbol{\delta})$ and $p_2(\boldsymbol{\delta})$ can be simplified further,
$$\begin{aligned}
p_1(\boldsymbol{\delta}) &= (\delta_1 - \delta_2)(\delta_5 - \delta_6)\left(-(\delta_1\delta_2)^{-1} - (\delta_5\delta_6)^{-1}\right) \\
&\quad - (\delta_2 - \delta_3)(\delta_4 - \delta_5)\left(-(\delta_2\delta_3)^{-1} - (\delta_4\delta_5)^{-1}\right) \\
p_2(\boldsymbol{\delta}) &= (\delta_1\delta_2\delta_5\delta_6)^{-1}(\delta_1 - \delta_2)(\delta_5 - \delta_6) \\
&\quad - (\delta_2\delta_3\delta_4\delta_5)^{-1}(\delta_2 - \delta_3)(\delta_4 - \delta_5) \ .
\end{aligned}$$

Next, since the minimal polynomial of γ over $\mathbb{F}_q$ is of degree 3, $p_i(\boldsymbol{\delta}) = 0$ for $i = 0, 1, 2$. $p_0(\boldsymbol{\delta}) = 0$ implies that,
$$(\delta_1 - \delta_2)(\delta_5 - \delta_6) = (\delta_2 - \delta_3)(\delta_4 - \delta_5) \neq 0 \ , \tag{2}$$

where the inequality follows since the coordinates of each of the vectors $(\delta_1, \delta_2, \delta_3)$ and $(\delta_4, \delta_5, \delta_6)$ are distinct. Substituting (2) in the equation $p_2(\boldsymbol{\delta}) = 0$ gives,
$$\delta_1\delta_6 = \delta_3\delta_4, \tag{3}$$
and by $p_1(\boldsymbol{\delta}) = 0$ and (3),
$$\delta_1\delta_2 + \delta_5\delta_6 = \delta_2\delta_3 + \delta_4\delta_5 \ . \tag{4}$$
(2) and (3) imply that,
$$\delta_1\delta_5 + \delta_2\delta_6 = \delta_2\delta_4 + \delta_3\delta_5 \ .$$
Subtracting the last equation from (4) we get,
$$(\delta_1 - \delta_6)(\delta_2 - \delta_5) = (\delta_3 - \delta_4)(\delta_2 - \delta_5) \ . \tag{5}$$

If $\delta_2 = \delta_5$ then by (4) $\delta_1 + \delta_6 = \delta_3 + \delta_4$. Together with (3), we get that both $\{\delta_1, \delta_6\}$ and $\{\delta_3, \delta_4\}$ are solutions to the quadratic equation,
$$x^2 - (\delta_3 + \delta_4)x + \delta_3\delta_4 = 0 \ .$$
Therefore, $\{\delta_1, \delta_6\} = \{\delta_3, \delta_4\}$, and we arrive at a contradiction to the assumption that two evaluation vectors have distinct coordinates and they agree on at most one coordinate (indeed, if $\delta_1 = \delta_4$ and $\delta_3 = \delta_6$ then $\beta_1 = \beta_4$ and $\beta_3 = \beta_6$). Otherwise, $\delta_2 \neq \delta_5$ and (5) becomes $\delta_1 - \delta_6 = \delta_3 - \delta_4$. Combining it with (3), we get that, as before, both $\{\delta_1, -\delta_6\}$ and $\{\delta_3, -\delta_4\}$ are solutions to the quadratic equation $x^2 - (\delta_3 - \delta_4)x - \delta_3\delta_4 = 0$, and therefore $\{\delta_1, -\delta_6\} = \{\delta_3, -\delta_4\}$. Recall that for $\delta \in \mathcal{A}, -\delta \notin \mathcal{A}$, thus, the only possibility is that $\delta_1 = \delta_3$ and $\delta_4 = \delta_6$ and again we arrived at a contradiction. The result follows. $\square$

2.2 Second Construction

In this section, we give the second code construction that improves the length of the code when the characteristic of the finite field $\mathbb{F}_q$ does not equal to 2 (See Remark 2.2). Specifically, the code's length can be as large as $q - 1$. The evaluation points of the constructed RS-code and the proof are very similar to the first construction.

Proposition 2.3. *Let $\mathbb{F}_q$ be a finite field of characteristic $p > 2$ and let $\mathcal{A} \subseteq \mathbb{F}_q^*$ be a subset of size n. Let γ be a root of a degree 3 irreducible polynomial over $\mathbb{F}_q$, and let the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ be some ordering of the n elements $\delta + \delta^2 \cdot \gamma, \delta \in \mathcal{A}$. Then, the $[n, 2]$ RS-code defined over $\mathbb{F}_{q^3}$ with the evaluation vector $\boldsymbol{\alpha}$ can correct any $n - 3$ insdel errors.*

Proof. As before, assume towards a contradiction that the claim is false, then by Proposition 1.6 there exist two vectors of distinct evaluation points $(\beta_1, \beta_2, \beta_3)$ and $(\beta_4, \beta_5, \beta_6)$, that agree on at most one coordinate, such that

$$\left| \begin{pmatrix} 1 & \beta_1 & \beta_4 \\ 1 & \beta_2 & \beta_5 \\ 1 & \beta_3 & \beta_6 \end{pmatrix} \right| = 0 \ .$$

Equivalently,

$$(\beta_1 - \beta_2)(\beta_5 - \beta_6) - (\beta_2 - \beta_3)(\beta_4 - \beta_5) = 0 \ .$$

Write $\beta_i := \delta_i + \delta_i^2 \cdot \gamma$ and observe that the LHS is a polynomial in γ of degree less than 3 over $\mathbb{F}_q$. Namely, $p_0(\boldsymbol{\delta}) + p_1(\boldsymbol{\delta}) \cdot \gamma + p_2(\boldsymbol{\delta}) \cdot \gamma^2 = 0$, where,

$$p_0(\boldsymbol{\delta}) = (\delta_1 - \delta_2)(\delta_5 - \delta_6) - (\delta_2 - \delta_3)(\delta_4 - \delta_5)$$
$$p_1(\boldsymbol{\delta}) = (\delta_1 - \delta_2)(\delta_5^2 - \delta_6^2) + (\delta_1^2 - \delta_2^2)(\delta_5 - \delta_6)$$
$$\qquad - (\delta_2^2 - \delta_3^2)(\delta_4 - \delta_5) - (\delta_2 - \delta_3)(\delta_4^2 - \delta_5^2)$$
$$p_2(\boldsymbol{\delta}) = (\delta_1^2 - \delta_2^2)(\delta_5^2 - \delta_6^2) - (\delta_2^2 - \delta_3^2)(\delta_4^2 - \delta_5^2) \ .$$

Next, by the definition of γ, $p_i(\boldsymbol{\delta}) = 0$ for $i = 0, 1, 2$. $p_0(\boldsymbol{\delta}) = 0$ implies that,

$$(\delta_1 - \delta_2)(\delta_5 - \delta_6) = (\delta_2 - \delta_3)(\delta_4 - \delta_5) \neq 0, \tag{6}$$

where the inequality follows since the coordinates of each of the vectors $(\delta_1, \delta_2, \delta_3)$ and $(\delta_4, \delta_5, \delta_6)$ are distinct. Substituting (6) in $p_2(\boldsymbol{\delta}) = 0$ and $p_1(\boldsymbol{\delta}) = 0$ gives,

$$(\delta_1 + \delta_2)(\delta_5 + \delta_6) = (\delta_2 + \delta_3)(\delta_4 + \delta_5) \ , \tag{7}$$

and

$$\delta_1 + \delta_6 = \delta_3 + \delta_4 \ , \tag{8}$$

respectively. Summing (6) and (7) and as the characteristic is different than 2 we get that,

$$\delta_5(\delta_1 - \delta_3) = \delta_2(\delta_4 - \delta_6) \ ,$$

which with (8) implies that $\delta_2 = \delta_5$. Plugging it in (7), and taking (8) into consideration, we get that,

$$\delta_1 \delta_6 = \delta_3 \delta_4 \ . \tag{9}$$

By (8) and (9), we get that $\{\delta_1, \delta_6\}$ and $\{\delta_3, \delta_4\}$ are the solutions to the quadratic equation $x^2 - (\delta_3 + \delta_4)x + \delta_3\delta_4 = 0$, and therefore $\{\delta_1, \delta_6\} = \{\delta_3, \delta_4\}$, which as before leads to a contradiction. $\qquad \square$

References

[BGZ17]   Joshua Brakensiek, Venkatesan Guruswami, and Samuel Zbarsky. Efficient low-redundancy codes for correcting multiple deletions. *IEEE Transactions on Information Theory*, 64(5):3403–3410, 2017.

[CGHL21]   Kuan Cheng, Venkatesan Guruswami, Bernhard Haeupler, and Xin Li. Efficient linear and affine codes for correcting insertions/deletions. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1–20. SIAM, 2021.

[CJL+23]   Kuan Cheng, Zhengzhong Jin, Xin Li, Zhide Wei, and Yu Zheng. Linear insertion deletion codes in the high-noise and high-rate regimes. *arXiv preprint arXiv:2303.17370*, 2023.

[CJLW18]   Kuan Cheng, Zhengzhong Jin, Xin Li, and Ke Wu. Deterministic document exchange protocols, and almost optimal binary codes for edit errors. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 200–211. IEEE, 2018.

[CST22]   Roni Con, Amir Shpilka, and Itzhak Tamo. Explicit and efficient constructions of linear codes against adversarial insertions and deletions. *IEEE Transactions on Information Theory*, 68(10):6516–6526, 2022.

[CST23]   Roni Con, Amir Shpilka, and Itzhak Tamo. Reed–solomon codes against adversarial insertions and deletions. *IEEE Transactions on Information Theory*, 2023.

[CZ21]   Bocong Chen and Guanghui Zhang. Improved Singleton bound on insertion-deletion codes and optimal constructions. *arXiv preprint arXiv:2105.02004*, 2021.

[DLTX19]   Tai Do Duc, Shu Liu, Ivan Tjuawinata, and Chaoping Xing. Explicit Constructions of Two-Dimensional Reed-Solomon Codes in High Insertion and Deletion Noise Regime. *arXiv preprint arXiv:1909.03426*, 2019.

[GH21]   Venkatesan Guruswami and Johan Håstad. Explicit two-deletion codes with redundancy matching the existential bound. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 21–32. SIAM, 2021.

[GHS20]   Venkatesan Guruswami, Bernhard Haeupler, and Amirbehshad Shahrasbi. Optimally resilient codes for list-decoding from insertions and deletions. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 524–537, 2020.

[GW16]   Venkatesan Guruswami and Mary Wootters. Repairing reed-solomon codes. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 216–226, 2016.

[GW17]   Venkatesan Guruswami and Carol Wang. Deletion codes in the high-noise and high-rate regimes. *IEEE Transactions on Information Theory*, 63(4):1961–1970, 2017.

[Hae19]   Bernhard Haeupler. Optimal document exchange and new codes for insertions and deletions. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 334–347. IEEE, 2019.

[HMG19]   Reinhard Heckel, Gediminas Mikutis, and Robert N Grass. A characterization of the DNA data storage channel. *Scientific reports*, 9(1):1–12, 2019.

[HS17]   Bernhard Haeupler and Amirbehshad Shahrasbi. Synchronization strings: codes for insertions and deletions approaching the Singleton bound. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 33–46. ACM, 2017.

[HS21]   Bernhard Haeupler and Amirbehshad Shahrasbi. Synchronization strings and codes for insertions and deletions - A survey. *IEEE Trans. Inf. Theory*, 67(6):3190–3206, 2021.

[LSWZY19]   Andreas Lenz, Paul H Siegel, Antonia Wachter-Zeh, and Eitan Yaakobi. Coding over sets for dna storage. *IEEE Transactions on Information Theory*, 66(4):2331–2351, 2019.

[LT21]   Shu Liu and Ivan Tjuawinata. On 2-dimensional insertion-deletion Reed-Solomon codes with optimal asymptotic error-correcting capability. *Finite Fields and Their Applications*, 73:101841, 2021.

[LTX23]   Shu Liu, Ivan Tjuawinata, and Chaoping Xing. Explicit construction of q-ary 2-deletion correcting codes with low redundancy. *arXiv preprint arXiv:2306.02868*, 2023.

[LX22]   Shu Liu and Chaoping Xing. Bounds and constructions for insertion and deletion codes. *IEEE Transactions on Information Theory*, 2022.

[MS81]   Robert J. McEliece and Dilip V. Sarwate. On sharing secrets and Reed-Solomon codes. *Communications of the ACM*, 24(9):583–584, 1981.

[SH+22]   Ilan Shomorony, Reinhard Heckel, et al. Information-theoretic foundations of dna data storage. *Foundations and Trends® in Communications and Information Theory*, 19(1):1–106, 2022.

[SNW02]   Reihaneh Safavi-Naini and Yejing Wang. Traitor tracing for shortened and corrupted fingerprints. In *ACM workshop on Digital Rights Management*, pages 81–100. Springer, 2002.

[Soo08]   Tan Jin Soon. QR code. *Synthesis Journal*, 2008:59–78, 2008.

[SWZGY17]   Clayton Schoeny, Antonia Wachter-Zeh, Ryan Gabrys, and Eitan Yaakobi. Codes correcting a burst of deletions or insertions. *IEEE Transactions on Information Theory*, 63(4):1971–1985, 2017.

[TB14]   Itzhak Tamo and Alexander Barg. A family of optimal locally recoverable codes. *IEEE Transactions on Information Theory*, 60(8):4661–4676, 2014.

[TSN07]   Dongvu Tonien and Reihaneh Safavi-Naini. Construction of deletion correcting codes using generalized Reed–Solomon codes and their subcodes. *Designs, Codes and Cryptography*, 42(2):227–237, 2007.

[WMSN04]   Yejing Wang, Luke McAven, and Reihaneh Safavi-Naini. Deletion correcting using generalized Reed-Solomon codes. In *Coding, Cryptography and Combinatorics*, pages 345–358. Springer, 2004.