Repairing Reed-Solomon Codes via Subspace Polynomials

Hoang Dau, Member, IEEE, Dinh Thi Xinh, Han Mao Kiah, Member, IEEE, Tran Thi Luong, and Olgica Milenkovic, Fellow, IEEE

Abstract—We propose new repair schemes for Reed-Solomon codes that use subspace polynomials and hence generalize previous works in the literature that employ trace polynomials. The Reed-Solomon codes are over $\mathbb{F}_{q^m}$ and have redundancy $r = n - k \geq q^m$, $1 \leq m \leq \ell$, where $n$ and $k$ are the code length and dimension, respectively. In particular, for one erasure, we show that our schemes can achieve optimal repair bandwidths whenever $n = q^s$ and $r = q^m$, for all $1 \leq m \leq \ell$. For two erasures, our schemes use the same bandwidth per erasure as the single erasure schemes, for $r/m$ is a power of $q$, and for $\ell/m$ is a power of $q$, and for $\ell \geq q^s$, $m = q^s - 1 > 1$ ($a \geq b \geq 1$), and for $m \geq \ell/2$ when $\ell$ is even and $q$ is a power of two.

I. INTRODUCTION

The repair bandwidth is a crucial performance metric of erasure codes when deployed in distributed storage systems [2], [3]. In such systems, for an underlying finite field $\mathbb{F}$, e.g. $\mathbb{F} = \text{GF}(256)$, a data vector in $\mathbb{F}^k$ is transformed into a codeword vector in $\mathbb{F}^n$, whose components are subsequently stored at different storage nodes. When a node fails, the codeword symbol stored at that node is erased (lost). A replacement node (RN) has to recover the content stored at the failed node by downloading relevant information from the remaining operational nodes. The repair bandwidth refers to the total amount of information (in bits) that the RN has to download in order to complete the repair process. If multiple erasures occur, different RNs may also exchange information in a distributed manner, and we are interested in the bandwidth used per erasure. Alternatively, multiple erasures can be recovered by a centralized entity, which, however, is not the focus of this work.

Reed-Solomon codes [4], the most practically used maximum distance separable codes [5], have been deployed in major distributed storage systems such as the Google File System II, Quantcast File System, Yahoo Object Store, Facebook f4 Storage System, Baidu Atlas Cloud Storage, Backblaze Vaults, and HDFS (see [6] Table I). However, they perform poorly as erasure codes under the repair bandwidth metric. For instance, to repair a data chunk of size 256 MB, the default repair scheme for the Reed-Solomon code (14,10) employed by Facebook’s f4 [7] implementation requires a repair bandwidth of 2.56 GB. As observed earlier in [8], the bandwidth used for repairing Reed-Solomon coded data in a Facebook analytics cluster amounts to 10%-20% of the total network traffic within the cluster.

There has been a considerable effort by the research community to improve and optimize the repair bandwidth of Reed-Solomon codes [2]–[16]. Several extensions to the case of multiple erasures were also studied [6], [16]–[20]. The optimal repair bandwidth of Reed-Solomon codes is generally unknown, except for some full-length codes [1], [10], [11] and for codes with exponentially large subpacketizations [15], [16]. Constructions of Reed-Solomon codes and repair schemes that trade-off the repair bandwidth and subpacketization size were also investigated in [21]–[23]. Another line of relevant research has focused on the I/O cost (the number of bits accessed at helper nodes) of repairing Reed-Solomon codes [24]–[26].

The focus of this work is on constructions of repair schemes for Reed-Solomon codes over $\mathbb{F}_{q^m}$ with redundancy $r \geq q^m$ using subspace polynomials in $\mathbb{F}_{q^m}[x]$ whose root sets form $m$-dimensional $\mathbb{F}_q$-subspaces of $\mathbb{F}_{q^m}$ (treated as a vector space over $\mathbb{F}_q$). For a single erasure, we show that the proposed repair scheme uses a repair bandwidth of $(n - 1)(\ell - m)/\log_2 q$ bits per erasure, which is optimal when $n = q^s$ and $r = q^m$ for all $1 \leq m \leq \ell$, based on a newly derived lower bound that (slightly) improves upon that in [10], [11]. Our scheme generalizes the method introduced in [10], [11] which employs trace polynomials and only works when $r \geq q^m$ and $(\ell - m)$ divides $\ell$. Note that a trace polynomial is a special subspace polynomial and the divisibility condition is imposed by the property that the set of values of the trace polynomial must lie in a subspace of $\mathbb{F}_{q^m}$. This constraint is relaxed in our construction as only subspaces are required.

We also develop distributed schemes that repair two erasures for Reed-Solomon codes. When $r \geq q^m$, it has been shown by Dau et al. [6], [17] and by Zhang and Zhang [20] that two erasures can be repaired using a bandwidth of $(n - 1)\log_2 q$ bits per erasure via trace polynomial approaches. Our goal is to consider the more general case when $r \geq q^m$, $1 \leq m \leq \ell$. We describe constructions of several schemes repairing two erasures for Reed-Solomon codes with a bandwidth of $(n - 1)(\ell - m)/\log_2 q$ bits per erasure, the same bandwidth as required for a single erasure. More specifically, our constructions apply when $\ell/m$ is a power of $q$, and when $\ell = q^a$ and $m = q^b - 1 > 1$ (for all $a \geq b \geq 1$), and when $m \geq \ell/2$ and $\ell$ is even and $q$ is a power of two. In this setting we also make use of subspace polynomials: However, while any subspace polynomial may be used for the repair scheme of a single erasure, this is not the case for two erasures. There, subspace polynomials satisfying certain additional properties are needed.
The remainder of the paper is organized as follows. We first provide relevant definitions and terminologies and then discuss the Guruswami-Wootters repair scheme for Reed-Solomon codes in Section II. The improved lower bound on the repair bandwidth of a single erasure is presented in Section III. We introduce new repair schemes for a single erasure and for two erasures in Section IV and Section V, respectively. We provide concluding remarks in Section VI.

II. PRELIMINARIES

A. Definitions and Notations

Let \([n]\) denote the set \(\{1, 2, \ldots, n\}\) and \([m, n]\) the set \(\{m, m+1, \ldots, n\}\). Let \(F_q\) be the finite field of \(q\) elements, for some prime power \(q\). Let \(F_{q^r}\) be an extension field of \(F_q\), where \(r \geq 1\). We refer to the elements of \(F_{q^r}\) as symbols and the elements of \(F_q\) as subscripts. The field \(F_{q^r}\) may also be viewed as a vector space of dimension \(r\) over \(F_q\), i.e., \(F_{q^r} \cong F_q^r\), and hence each symbol in \(F_{q^r}\) can be represented as a vector of length \(r\) over \(F_q\). We use \(\text{span}_{F_q}(U)\) to denote the \(F_q\)-subspace of \(F_{q^r}\) spanned by a set of elements \(U \subseteq F_{q^r}\). We use \(\dim_{F_q}(\cdot)\) and \(\text{rank}_{F_q}(\cdot)\) to denote the dimension of a subspace and the rank of a set of vectors over \(F_q\), respectively. The (field) trace of any symbol \(\alpha \in F_{q^r}\) over \(F_q\) is defined as \(\text{Tr}_{F_{q^r}/F_q}(\alpha) = \sum_{i=0}^{r-1} \alpha^q^i\). When clear from the context, we omit the subscripts \(F_{q^r}/F_q\).

A linear \([n, k]\) code \(C\) over \(F_{q^r}\) is an \(F_{q^r}\)-subspace of \(F_{q^r}^n\) of dimension \(k\). Each element of a code is referred to as a codeword. Each element \(c_j\) of a codeword \(c = (c_1, c_2, \ldots, c_n) \in C \subseteq F_{q^r}^n\) is referred to as a codeword symbol. The dual \(C^\perp\) of a code \(C\) is the orthogonal complement of \(C\) in \(F_{q^r}^n\) and has dimension \(r = n-k\).

Definition 1. Let \(F_{q^r}[x]\) denote the ring of polynomials over \(F_{q^r}\). A Reed-Solomon code \(RS(A, k) \subseteq F_{q^r}^n\) of dimension \(k\) over a finite field \(F_{q^r}\) with evaluation points \(A = \{\alpha_j\}_{j=1}^n \subseteq F_{q^r}\) is defined as

\[
RS(A, k) = \left\{ \left(f(\alpha_1), \ldots, f(\alpha_n)\right) : f \in F_{q^r}[x], \deg(f) < k \right\}.
\]

The Reed-Solomon code is full length if \(n = q^k\), i.e., \(A \equiv F_{q^r}\).

A generalized Reed-Solomon code, \(GRS(A, k, \lambda)\), where \(\lambda = (\lambda_1, \ldots, \lambda_n) \in F^n\), is defined similarly to a Reed-Solomon code, except that the codeword corresponding to a polynomial \(f\) is defined as \((\lambda_1 f(\alpha_1), \ldots, \lambda_n f(\alpha_n))\), where \(\lambda_j \neq 0\) for all \(j \in [n]\). It is well known that the dual of a Reed-Solomon code \(RS(A, k)\), for any \(n \leq |F|\), is a generalized Reed-Solomon code \(GRS(A, n-k, \lambda)\), for some multiplier vector \(\lambda\) (see [8], Chp. 10).

Whenever clear from the context, we use \(f(x)\) to denote a polynomial of degree at most \(k-1\), which corresponds to a codeword of the Reed-Solomon code \(C = RS(A, k)\), and \(g(x)\) to denote a polynomial of degree at most \(r-1 = n-k-1\), which corresponds to a codeword of the dual code \(C^\perp\). Since \(\sum_{j=1}^n g(\alpha_j) (\lambda_j f(\alpha_j)) = 0\), we also refer to the polynomial \(g(x)\) as a check polynomial for \(C\). Note that when \(n = q^k\), we have \(\lambda_j = 1\) for all \(j \in [n]\). In general, as the column multipliers \(\lambda_j\) do not play any role in evaluating the repair bandwidth, they are often omitted to simplify the notation (see also Remark 1).

B. Trace repair framework

First, note that each element of \(F_{q^r}\) can be recovered from its \(\ell\) independent traces. More precisely, given a basis \(\{\beta_i\}_{i=1}^\ell\) of \(F_{q^r}\) over \(F_q\), any \(\alpha \in F_{q^r}\) can be uniquely determined given the values of \(\text{Tr}(\beta_i \alpha)\) for \(i \in [\ell]\), i.e., \(\alpha = \sum_{i=1}^\ell \text{Tr}(\beta_i \alpha) \beta_i^\perp\), where \(\{\beta_i^\perp\}_{i=1}^\ell\) is the dual (trace-orthogonal) basis of \(\{\beta_i\}_{i=1}^\ell\) (see, e.g., [27], Ch. 2, Def. 2.30).

Let \(C\) be an \([n, k]\) linear code over \(F_{q^r}\) and \(C^\perp\) its dual. The repair scheme for \(c_j\), based on \(\ell\) dual codewords \(\vec{g}^{(1)}, \ldots, \vec{g}^{(\ell)}\), where \(\dim_{F_q}(S_j) = \ell\), incurs a repair bandwidth of \(\sum_{j \neq j_*} \dim_{F_q}(S_j)\) subsymbols in \(F_{q^r}\), where \(S_j\) are column spaces defined as in (2).

Lemma 1 (Guruswami-Wootters [10]). Let \(C\) be an \([n, k]\) linear code over \(F_{q^r}\) and \(C^\perp\) its dual. The repair scheme for \(c_j\), based on \(\ell\) dual codewords \(\vec{g}^{(1)}, \ldots, \vec{g}^{(\ell)}\), where \(\dim_{F_q}(S_j) = \ell\), incurs a repair bandwidth of \(\sum_{j \neq j_*} \dim_{F_q}(S_j)\) subsymbols in \(F_{q^r}\), where \(S_j\) are column spaces defined as in (2).
Remark 1. When \( \mathcal{C} = \text{RS}(A, k) \) is a Reed-Solomon code with 
\( A = \{ \alpha_j \}_{j=1}^n \subset \mathbb{F}_q^* \), its dual codewords are of the form 
\( \tilde{g} = (\lambda_1 g(\alpha_1), \lambda_2 g(\alpha_2), \ldots, \lambda_n g(\alpha_n)) \), where \( g(x) \in \mathbb{F}_q[x] \) are polynomials of degrees at most \( r - 1 = n - k + 1 \) and \( \lambda_j \in \mathbb{F}_q^* \) are fixed column multipliers. A repair scheme for \( c_j \), is based on \( \ell \) polynomials \( g_j(x), \ldots, g_r(x) \). Since 
\( S_j = \lambda_j \text{span}_{\mathbb{F}_q} \langle (g_i(\alpha_j))_{i=1}^\ell \rangle \),
we have \( \dim_{\mathbb{F}_q}(S_j) = \dim_{\mathbb{F}_q} \text{span}_{\mathbb{F}_q} \langle (g_i(\alpha_j))_{i=1}^\ell \rangle \). Therefore, the multiplier \( \lambda_j \) are irrelevant for determining the repair bandwidth of the repair scheme based on \( g_1(x), \ldots, g_r(x) \). We slightly abuse the notation and henceforth ignore \( \lambda_j \) and referring to 
\( S_j = \text{span}_{\mathbb{F}_q} \langle (g_i(\alpha_j))_{i=1}^\ell \rangle \) as the column space of the repair scheme for a Reed-Solomon code. Another way to view this simplification is that as recovering \( f(\alpha_j) \) is equivalent to recovering \( \lambda_j f(\alpha_j) \), one can safely ignore \( \lambda_j \) and focus only on \( g_i(x) \) in the construction of low-bandwidth repair schemes for Reed-Solomon codes.

| \( j \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----|---|---|---|---|---|---|---|---|
| \( g^{(1)} \) | 1 | \( \xi^1 \) | \( \xi^2 \) | \( \xi^3 \) | \( \xi^4 \) | \( \xi^5 \) | \( \xi^6 \) | \( \xi^7 \) |
| \( g^{(2)} \) | \( \xi^2 \) | \( \xi^3 \) | \( \xi^4 \) | \( \xi^5 \) | \( \xi^6 \) | \( \xi^7 \) | \( \xi^8 \) | \( \xi^0 \) |
| \( g^{(3)} \) | \( \xi^3 \) | 1 | \( \xi^4 \) | \( \xi^5 \) | \( \xi^6 \) | \( \xi^7 \) | \( \xi^8 \) | . |
| \( \dim_{\mathbb{F}_q}(S_{j=1}) \) | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

TABLE 1: A list of three dual codewords used to repair the first codeword symbol \( c_1 \) of an \([8, 6]\) Reed-Solomon code over \( \mathbb{F}_8 \). These dual codewords must be known to all nodes in advance. Here, a dot “.” stands for a zero entry. It suffices for the replacement node to download two bits from each available node, for instance, \( \text{Tr}(\xi^2 c_2) \) and \( \text{Tr}(\xi^2 c_3) \) from the node storing \( c_2 \), or \( \text{Tr}(\xi^2 c_4) \) and \( \text{Tr}(c_1) \) from the node storing \( c_4 \). Thus, the scheme has a repair bandwidth of \( 14 = 7 \times 2 \) bits.

Example 1. Consider a repair scheme of the first codeword symbol \( c_1 \) of an \([8, 6]\) Reed-Solomon code over \( \mathbb{F}_8 \) that is based on \( \ell = 3 \) dual codewords \( g^{(1)}, g^{(2)}, \) and \( g^{(3)} \) as given in Table 1. We have \( q = 2, n = 8, k = 6, \) and \( j^* = 1 \) in this case. This repair scheme can be constructed using Construction I developed in Section [7] (see Example 2). First, this set of three dual codewords corresponds to a repair scheme for \( c_1 \) because

\[ \dim_{\mathbb{F}_q}(S_1) = \text{rank}_{\mathbb{F}_q}(\langle 1, \xi, \xi^2 \rangle) = 3 = \ell, \]

where \( \xi \) is a primitive element of \( \mathbb{F}_8 \) satisfying \( \xi^3 + \xi + 1 = 0 \). For instance, for the codeword \( \tilde{c} = ([?], 1, \xi^2, \xi^1, 0, 0, 0, 0) \) where \( c_1 \) has been erased, the three repair equations are given below.

\[
\begin{align*}
\text{Tr}(c_1) &= -\text{Tr}(\xi^2 c_2) - \text{Tr}(\xi^6 c_4) - \text{Tr}(\xi^5 c_5) = 1, \\
\text{Tr}(\xi c_1) &= -\text{Tr}(\xi^4 c_2) - \text{Tr}(c_3) - \text{Tr}(\xi^6 c_4) - \text{Tr}(\xi^5 c_5) = 0, \\
\text{Tr}(\xi^2 c_1) &= -\text{Tr}(\xi^2 c_2) - \text{Tr}(\xi^3 c_3) - \text{Tr}(\xi^2 c_4) - \text{Tr}(\xi^3 c_5) = 0.
\end{align*}
\]

As \([1, \xi^2, \xi] \) is a dual basis of \([1, \xi^2, \xi^2] \), we can recover \( c_1 \) as follows.

\[ c_1 = \text{Tr}(c_1) + \text{Tr}(\xi c_1)\xi^2 + \text{Tr}(\xi^2 c_1)\xi = 1. \]

Next, the repair bandwidth is equal to the sum of the dimensions (over \( \mathbb{F}_q \)) of the column spaces \( S_j \), or equivalently, the sum of ranks (over \( \mathbb{F}_q \)) of the entries in columns \( j \) of Table 1.

III. AN IMPROVED LOWER BOUND ON THE REPAIR BANDWIDTH OF A SINGLE ERASURE

In order to evaluate their proposed single erasure repair scheme, Guruswami and Wootters [10] established a lower bound on the repair bandwidth for Reed-Solomon codes. We start our exposition by improving their bound. The result of this derivation also suggests the number of subsymbols that needs to be downloaded from each available node using an optimal repair scheme. Consequently, the bound allows one to perform a theoretical/numerical search for optimal repair schemes in a simple manner.

Proposition 1. Any linear repair scheme for Reed-Solomon codes \( \text{RS}(A, k) \) over \( \mathbb{F}_q \) that uses the base field \( \mathbb{F}_q \) requires a bandwidth of at least 
\[ t \log_q \left( \frac{(n-1)q^t}{(r-1)(q^t - 1) + (n-1)} \right), \]

subsymbols over \( \mathbb{F}_q \), where \( n = |A| \leq q^t \), and where \( b_{\text{AVE}} \) and \( t \) are defined as

\[ b_{\text{AVE}} = \log_q \left( \frac{(n-1)q^t}{(r-1)(q^t - 1) + (n-1)} \right), \]

and \( \hat{t} \equiv n - 1 \) if \( b_{\text{AVE}} \in \mathbb{Z} \), and \( \hat{t} \equiv T - (n-1)q^{-\lceil \log_q \left( q^{-1}b_{\text{AVE}} \right) \rceil} \) otherwise. Here,

\[ T = \frac{(r-1)(q^t - 1) + (n-1)}{q^{b_{\text{AVE}}} - q^{-1}b_{\text{AVE}}}. \]

Proof. The first part of the proof proceeds along the same lines as the proof of [10] Thm. 6. But once the optimization problem is solved to arrive at a fractional lower bound, rather than allowing the number of subsymbols downloaded from each available node to be real-valued, we perform a rounding procedure which leads to an improved integral lower bound.
Fix an \( \alpha^* \in A \) and consider an arbitrary exact linear repair scheme of Reed-Solomon codes for the node storing \( f(\alpha^*) \) that uses \( b \) subsymbols from \( \mathbb{F}_q \). By [10] Thm. 4, there is a set of \( \ell \) polynomials \( g_1(x), \ldots, g_\ell(x) \) such that \( \text{rank}_{\mathbb{F}_q}(\{g_1(\alpha^*), \ldots, g_\ell(\alpha^*)\}) = \ell \) and \( \text{rank}_{\mathbb{F}_q}(\{g_1(\alpha), \ldots, g_\ell(\alpha)\}) = b_\alpha \), for all \( \alpha \in A \setminus \{\alpha^*\} \), where \( b = \sum_{\alpha \in A \setminus \{\alpha^*\}} b_\alpha \). For each \( \alpha \in A \), \( \ell \)

\[
S_\alpha \triangleq \{ \tilde{s} = (s_1, \ldots, s_\ell) \in \mathbb{F}_q^\ell : \sum_{i=1}^\ell s_i g_i(\alpha) = 0 \}.
\]

As \( \text{rank}_{\mathbb{F}_q}(\{g_1(\alpha), \ldots, g_\ell(\alpha)\}) = b_\alpha \), we have \( \dim_{\mathbb{F}_q}(S_\alpha) = \ell - b_\alpha \). Averaging over all nonzero vectors \( \tilde{s} \in \mathbb{F}_q^\ell \), we obtain

\[
\frac{1}{q^\ell - 1} \sum_{\tilde{s} \in \mathbb{F}_q^\ell \setminus \{0\}} |\{ \alpha \in A \setminus \{\alpha^*\} : \tilde{s} \in S_\alpha \}| = \frac{1}{q^\ell - 1} \sum_{\alpha \in A \setminus \{\alpha^*\}} (q^{\ell-b_\alpha} - 1) = E. \tag{3}
\]

Therefore, there exists some \( \tilde{s}^* = (s_1^*, \ldots, s_\ell^*) \in \mathbb{F}_q^\ell \setminus \{0\} \) so that \( |\{ \alpha : \tilde{s}^* \in S_\alpha \}| \geq E \). Let \( g(x) = \sum_{i=1}^\ell s_i^* g_i(x) \). By the choice of \( \tilde{s}^* \), \( g(x) \) vanishes on at least \( E \) points of \( A \setminus \{\alpha^*\} \). Also, since \( \tilde{s}^* \neq 0 \), \( g(\alpha^*) = \sum_{i=1}^\ell s_i^* g_i(\alpha^*) \neq 0 \). Therefore, \( g(x) \) corresponds to a nonzero codeword in the dual code \( C^\perp \) and hence can have at most \( r - 1 - \ell \) roots. Thus,

\[
\frac{1}{q^\ell - 1} \sum_{\alpha \in A \setminus \{\alpha^*\}} (q^{\ell-b_\alpha} - 1) = E \leq r - 1,
\]

or equivalently,

\[
\sum_{\alpha \in A \setminus \{\alpha^*\}} q^{\ell-b_\alpha} \leq \left((r-1)(q^\ell - 1) + (n-1)\right)/q^\ell = \ell: \tag{4}
\]

Let

\[
b_{\min} \triangleq \min_{b_\alpha \in \{0, 1, \ldots, \ell\}} \sum_{\alpha \in A \setminus \{\alpha^*\}} b_\alpha, \quad \text{subject to (4)}. \tag{5}
\]

Then, any feasible repair scheme has to have \( b \geq b_{\min} \). To solve the optimization problem (5), the authors of [10] Thm. 6] relaxed the condition that \( b_\alpha \) are integer-valued and arrived at a lower bound that reads as \( n - k - 1 \leq b_{\min} \) where \( b_{\min} \triangleq \log_q(n - 1)/T \). But one can still solve (5) for \( b_\alpha \in \{0, 1, \ldots, \ell\} \) and arrive at a closed form expression for \( b_{\min} \). To see how to accomplish this analysis, we first let \( \{b_1, \ldots, b_{n-1}\} \) refer to \( \{b_\alpha : \alpha \in A \setminus \{\alpha^*\}\} \). We then claim that

\[
b_1^* = \cdots = b_\ell^* = [b_{\min}], \quad b_{\ell+1}^* = \cdots = b_{n-1}^* = [b_{\min}],
\]

where \( t \) is the largest integer satisfying \( \sum_{i=1}^{n-1} q^{-b_i^*} \leq T \), is an optimal solution of (5). To this end, if \( \{b_1, \ldots, b_{n-1}\} \) is an optimal solution of (5), and \( b_i - b_j \geq 2 \) for some \( i \) and \( j \), we may decrease \( b_i \) by one and increase \( b_j \) by one, and retain an optimal solution. Repeating this “balancing” procedure for as many times as possible, we obtain an optimal solution for which \( |b_i - b_j| \leq 1 \), \( i, j \in [n-1] \). If \( \min_i b_i < [b_{\min}] \), then \( \{b_1, \ldots, b_{n-1}\} \) cannot be a feasible solution. Therefore, \( \min_i b_i \geq [b_{\min}] \). Because of the way \( t \) was chosen, we always have \( \sum_{i=1}^{n-1} b_i \geq \sum_{i=1}^{t-1} b_i^* \), which establishes the optimality of \( \{b_1^*, \ldots, b_{n-1}^*\} \). Finally, \( t \) may be easily computed as follows. If \( b_{\min} \in \mathbb{Z} \) then \( t = n - 1 \), otherwise

\[
t = \left\lfloor \frac{T - (n - 1)q^{-[b_{\min}]} - q^{-[b_{\min}]}}{q^{-[b_{\min}]} - q^{-[b_{\min}]}} \right\rfloor.
\]

Corollary 1. When \( n = q^t \) and \( r = q^m \), for some \( m \in [\ell] \), any linear repair scheme over the subfield \( \mathbb{F}_q^\ell \) of a Reed-Solomon code \( RS(A, k) \) defined over \( \mathbb{F}_q^\ell \) requires a bandwidth of at least \( (n - 1)(\ell - m) \) subsymbols over \( \mathbb{F}_q^\ell \).

Proof. In this case, \( b_{\min} = \ell - m \in \mathbb{Z} \) and \( t = n - 1 \), which according to Proposition [1] give the desired bound.

Note that the integral bound of Corollary [1] and the Guruswami-Wootters fractional bound coincide. However, in many other cases, the integral bound strictly outperforms the fractional bound. Consider as an example the Facebook RS(14,10) code defined over GF(256). If the code is repaired over the subfield GF(16), the fractional bound results in at least 28 downloaded bits, while our integral bound asserts that a download of at least 44 bits is needed. It is also apparent that the fractional bound does not depend on the base field that the code is repaired over, while the integral bound does. In general, the bigger the order of the base field (given that the coding field is fixed), the larger the gap between the two bounds.

Also, one may assume that if a repair scheme that achieves the bound of Corollary [1] were to exist, it would require that the replacement node downloads \( \ell - m \) subsymbols from each available node. This intuitive reasoning will be highly valuable for the design of optimal repair schemes for Reed-Solomon codes.

IV. REPAIR ONE ERASURE FOR REED-SOLOMON CODES VIA SUBSPACE POLYNOMIALS

We now describe a way to select dual codewords \( \{\tilde{q}^{(i)}\}_{i=1}^\ell \) for \([n, k]\) Reed-Solomon codes with \( r = n - k \geq q^m \), \( 1 \leq m \leq \ell \), so that the corresponding repair scheme for one erasure incurs a low repair bandwidth. The key ingredient of our procedure are subspace polynomials [5], [27], [28]. When the code is full length, i.e., \( n = q^t \), and \( r = q^m \), the proposed repair scheme indeed achieves the minimum repair bandwidth, according to Proposition [1].

We henceforth use \( f(x) \in \mathbb{F}_q^\ell[x] \) to denote a polynomial of degree at most \( k - 1 \), which corresponds to a codeword of the Reed-Solomon code \( C = RS(A, k) \), and \( g(x) \in \mathbb{F}_q^\ell[x] \) to denote a polynomial of degree at most \( r - 1 \), which corresponds to a codeword of the dual code \( C^\perp \). In the remainder of this section, we assume that \( f(\alpha^*) \) is the erased codeword symbol, where \( \alpha^* \in A \) is an evaluation point of the code. Applying the trace repair framework to the dual codewords generated by \( \{g_i(x)\}_{i=1}^\ell \), we arrive at the following \( \ell \) repair equations

\[
\text{Tr}(g_i(\alpha^*) f(\alpha^*)) = -\sum_{\alpha \in A \setminus \{\alpha^*\}} \text{Tr}(g_i(\alpha) f(\alpha)), \quad i \in [\ell]. \tag{6}
\]

Note that in [6] we ignore the column multipliers to simplify the notation without affecting the bandwidth results of the repair scheme (see Remark [1]). The column spaces of the repair scheme based on \( \{g_i(x)\}_{i=1}^\ell \) are

\[
S_\alpha \triangleq \text{span}_{\mathbb{F}_q}(g_1(\alpha), \ldots, g_\ell(\alpha)), \quad \alpha \in A. \tag{7}
\]

In order to form a repair scheme, it is required that \( \dim_{\mathbb{F}_q}(S_\alpha) = \ell \). Moreover, the scheme uses a bandwidth of \( b = \sum_{\alpha \in A \setminus \{\alpha^*\}} \dim_{\mathbb{F}_q}(S_\alpha) \) subsymbols in \( \mathbb{F}_q \).
Construction 1. Let \( \{\beta_1, \ldots, \beta_t\} \) be an \( \mathbb{F}_q \)-basis of \( \mathbb{F}_{q^f} \) and \( W \) an arbitrary \( \mathbb{F}_q \)-subspace of dimension \( m \) of \( \mathbb{F}_{q^f} \). Set \( L_{W}(x) = \prod_{w \in W}(x-w) \). The check polynomials used to repair \( f(\alpha^*) \) are
\[
g_i(x) = L_{W}(\beta_i(x - \alpha^*))/(x - \alpha^*), \quad i \in [\ell].
\]

Note that \( L_W(x) \) constructed as in Construction 1 is referred to as a subspace polynomial, which is known to be a special type of linearized polynomials over \( \mathbb{F}_{q^f} \) (see [5, Ch. 4] and [28, p. 4]). The properties of \( L_W(x) \) and \( g_i(x) \) are captured in Lemma 2. Note that for an \( \mathbb{F}_q \)-linear mapping \( L: \mathbb{F}_q \to \mathbb{F}_{q^f} \), we use \( \ker(L) = \{ \alpha \in \mathbb{F}_q : L(\alpha) = 0 \} \) and \( \text{im}(L) = \{ \alpha \in \mathbb{F}_{q^f} : \alpha \in \mathbb{F}_q \} \) to denote the kernel and the image of \( L \), respectively.

Lemma 2. Let \( W \) be an \( m \)-dimensional \( \mathbb{F}_q \)-subspace of \( \mathbb{F}_{q^f} \) and \( L_W(x) \) and \( g_i(x) \) defined as in Construction 1. Then the following statements hold.

(a) \( L_W(\cdot) \) is an \( \mathbb{F}_q \)-linear mapping from \( \mathbb{F}_q \) to itself. Moreover, \( \ker(L_W) = W \) and \( \text{dim}_q(\text{im}(L_W)) = \ell - m \).

(b) \( \deg(g_i) = q^m - 1 \) for all \( i \in [\ell] \).

(c) \( g_i(\alpha^*) = \tau_W \beta_i \) for all \( i \in [\ell] \), where \( \tau_W = \prod_{w \in W \backslash \{0\}} w \).

(d) \( \text{dim}_q(S_{\alpha^*}) = \ell \).

(e) \( \text{dim}_q(S_{\alpha}) \leq \text{dim}_q(\text{im}(L_W)) = \ell - m \) for \( \alpha \neq \alpha^* \).

Proof of Lemma 2 A proof of Part (a) can be found in [28, p. 4]. As \( |W| = q^m \), \( L_W(x) \) has degree \( q^m \), which implies Part (b). To prove (c), we first write
\[
L_W(x) = \prod_{w \in W \backslash \{0\}} (x-w) = \tau_W x + x^2 M(x),
\]
for some polynomial \( M(x) \) of degree \( q^m - 2 \), noting that \( (-1)^{q^m-2} = 1 \) for both \( 1 \) and \( \alpha \). Therefore,
\[
L_W(\beta_i(x - \alpha^*)) = \tau_W \beta_i(\alpha^*) + (\beta_i(x - \alpha^*))^2 M(x).
\]

Hence, \( g_i(\alpha^*) = \tau_W \beta_i \) for \( i \in [\ell] \). For (d), as \( \tau_W \neq 0 \) and \( \{\beta_1, \ldots, \beta_t\} \) is an \( \mathbb{F}_q \)-basis of \( \mathbb{F}_{q^f} \), it follows that the set \( \{g_1(\alpha^*), \ldots, g_t(\alpha^*)\} \) has rank \( \ell \) over \( \mathbb{F}_q \), or equivalently, \( \text{dim}_q(S_{\alpha^*}) = \ell \).

It remains to prove Part (e). For \( \alpha \neq \alpha^* \), set \( \gamma_i = \beta_i(\alpha - \alpha^*) \) so that we have \( g_i(\alpha) = \frac{1}{\alpha - \alpha^*} L_W(\gamma_i) \). Hence, using Part (a),
\[
\text{rank}_q(\{g_1(\alpha), \ldots, g_t(\alpha)\}) = \text{rank}_q(\{L_W(\gamma_1), \ldots, L_W(\gamma_t)\}) \leq \dim_q(\text{im}(L_W)) = \ell - m,
\]
or equivalently,
\[
\text{dim}_q(S_{\alpha}) \leq \text{dim}_q(\text{im}(L_W)) = \ell - m \text{ for } \alpha \neq \alpha^*.
\]

Theorem 1. Let \( q^f \geq n \geq r \geq q^m \), where \( 1 \leq m \leq \ell \). The repair scheme based on check polynomials generated by Construction 1 can repair a codeword symbol \( f(\alpha^*) \) of an \( [n,k] \) Reed-Solomon code using a bandwidth of at most \((n-1)(\ell-m)\) subsymbols in \( \mathbb{F}_q \). When \( n = q^f \) and \( r = q^m \), this bandwidth is optimal.

Proof. We use the properties of \( g_i(x) \) described in Lemma 2 to prove the result. Part (b) of the lemma implies that the polynomials \( g_i(x) \) indeed correspond to the dual codewords of the Reed-Solomon code \( C \) with \( r = n - k \geq q^m \). Part (c) and Part (d) guarantee that the scheme based on \( \{g_i(x)\}_{i=1}^t \) can repair \( f(\alpha^*) \). Part (e) gives an upper bound on the bandwidths used by the helper nodes that equals
\[
b = \sum_{\alpha \in A \backslash \{\alpha^*\}} \dim_q(S_{\alpha}) \leq (n-1)(\ell-m).
\]
A. One-Round Repair Schemes for Two Erasures

We first show that if there is an $\mathbb{F}_q$-subspace $W$ of $\mathbb{F}_q^n$ of dimension $m$ satisfying certain properties then there exists a one-round repair scheme for two erasures with low bandwidth. We then demonstrate that such a subspace always exists when $\ell$ is even, $m \geq \ell/2$, and $q$ is a power of two.

**Theorem 2.** Consider a Reed-Solomon code of full length $n = q^\ell$ and $r \geq q^m$ over $\mathbb{F}_q$. Suppose there exists an $\mathbb{F}_q$-subspace $W$ of $\mathbb{F}_q^n$ of dimension $m$ satisfying

P1: $\tau_W := \{w \in \mathbb{F}_q^n | w \in W \}$, and

P2: $L^2(F_q) = \{0\}$, or equivalently, $L(W) \subseteq W = ker(L\_{\tau_W})$, where $L(W)$ is defined as in Construction I.

Then there exists a scheme that can repair two arbitrary erasures for this code using a bandwidth of at most $(n - 1)(\ell - m)$ subsymbols over $\mathbb{F}_q$ per erasure.

Note that a linearized polynomial $L(x) \in \mathbb{F}_q[x]$ satisfying $L^{\otimes 2}(F_q) = \{0\}$, or equivalently, $L(x) \equiv 0 \pmod{x^q - x}$, is referred to as a 2-nilpotent linearized polynomial (see [29]). However, such a polynomial may not even be a subspace polynomial. We first modify Construction I to cope with two erasures.

**Construction II.** Let $W$ be the $m$-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_q^n$ satisfying (P1) and (P2). Let $\{\beta_1, \ldots, \beta_m, \beta_i\}$ be an $\mathbb{F}_q$-basis of $W/(\alpha - \alpha^*)$, where $\beta_{m+1}, \ldots, \beta_{\ell}$ are chosen so that $\{\beta_1, \ldots, \beta_{\ell}\}$ forms an $\mathbb{F}_q$-basis of $W$, but are otherwise arbitrary. The check polynomials used to repair $f(\alpha^*)$ and $f(\alpha^*)$ are, respectively,

- $g_i(x) = L_W(\beta_i(x - \alpha^*))/ (x - \alpha^*)$, $i \in [\ell]$,
- $h_i(x) = L_W(\beta_i(x - \alpha))/ (x - \alpha)$, $i \in [\ell]$.

The following properties of $g_i$ and $h_i$ are needed to show that Construction II ensures the claimed results.

**Lemma 3.** Let $W$, $g_i$, $h_i$ be defined as in Construction II. Then the following statements hold.

a. $\deg(g_i(x)) = \deg(h_i(x)) = q^m - 1 - r - 1$, for all $i \in [\ell]$.

b. $g_i(\alpha^*) = h_i(\alpha^*) \notin \tau_W$, for all $i \in [\ell]$.

c. $g_i(\alpha) = h_i(\alpha) = 0$, for all $i \in [m]$.

d. $g_i(\alpha^*)$ and $h_i(\alpha^*)$ belong to $\text{span}_\mathbb{F}_q\{\tau_W\beta_1, \ldots, \tau_W\beta_m\}$, for $i \in [m + 1, \ell]$.

Note that (a) implies that $g_i$ and $h_i$ are check polynomials for the code, (b) guarantees that $\{g_i(x)\}_{i=1}^\ell$ and $\{h_i(x)\}_{i=1}^\ell$ form repair schemes for $f(\alpha^*)$ and $f(\alpha^*)$, respectively, while (b) and (c) together imply that $g_i(x), \ldots, g_m(x)$ involve $f(\alpha^*)$ but not $f(\alpha^*)$, and $h_i(x), \ldots, h_{m}(x)$ involve $f(\alpha^*)$ but not $f(\alpha^*)$. Finally, (d) allows the RNs to help each other using the information obtained from the Download Phase (we will discuss this point in detail later).

**Proof of Lemma 3.** As $|W| = q^m$, $L_W$ has degree $q^m$, which implies Statement (a). To prove (b), note that $L_W(\beta_i(x - \alpha^*)) = \tau_W\beta_i(x - \alpha^*) + (\beta_i(x - \alpha^*))^2M(x)$, for some polynomial $M(x)$, and therefore, $g_i(\alpha^*) = \tau_W\beta_i$. The same argument works for $h_i(x)$. Next, for $i \in [m]$, as $\beta_i(\alpha) = \alpha$, it holds that $L_W(\beta_i(\alpha - \alpha^*)) = 0$, which implies that $g_i(\alpha^*) = 0$. Similarly, $h_i(\alpha^*) = 0$ for all $i \in [m]$, which proves (c). Finally, to establish (d), note that (P2) implies that $L_W(\mathbb{F}_q^n) \subseteq ker(L_W) = W$. Therefore, $g_i(\alpha^*) = L_W(\beta_i(\alpha^*)/(\alpha^*) = \text{span}_{\mathbb{F}_q}\{\beta_1, \ldots, \beta_m\}$ as $\tau_W \in \mathbb{F}_q$, for all $i \in [m + 1, \ell]$. Similar argument holds for $h_i(\alpha^*)$, noting that $W/(\alpha^*) \equiv W/(\alpha^*)$.

Fig. 1: Illustration of the two-round repair scheme using check polynomials from Construction II. The arrows mean the target traces available at an RN after the Download Phase can then be used to generate (as linear combinations) the repair traces to reconstruct the target traces at the other RN in the Collaboration Phase ($\tau_W$ is ignored for notational simplicity).

**Proof of Theorem 2.** We describe below the two phases of the repair scheme based on the two sets of check polynomials $\{g_i(x)\}_{i=1}^\ell$ and $\{h_i(x)\}_{i=1}^\ell$ obtained in Construction II.

**Download Phase.** In this phase, each RN contacts $n - 2$ available nodes to download repair traces and also recovers their target traces. By Lemma 3(e), to obtain all the repair traces $\{g_i(\alpha)\}_{i=1}^\ell$ where $\alpha \in A' \setminus \{\alpha^*, \alpha\}$, the RN for $f(\alpha^*)$ needs to download at most $\ell - m$ traces/subsymbols over $\mathbb{F}_q$ from the helper node storing $f(\alpha^*)$. Hence, it uses a bandwidth of at most $(n - 2)(\ell - m)$ symbols in this phase. Next, the RN for $f(\alpha^*)$ uses the first $m$ check polynomials $g_1, \ldots, g_m$, which do not involve $f(\alpha^*)$ according to Lemma 3(c), to construct the following $m$ repair equations.

$$g_i(\alpha^*)f(\alpha^*) = - \sum_{\alpha \in A' \setminus \{\alpha^*, \alpha\}} \tau_W(\beta_i(\alpha f(\alpha^*))$$

As the result, it can reconstruct $m$ target traces $g_i(\alpha^*)f(\alpha^*) = \tau_W(\beta_i(\alpha f(\alpha^*))$, $i \in [m]$, of $f(\alpha^*)$. Similarly, the RN for $f(\alpha^*)$ uses the $m$ repair equations

$$\tau_W(\beta_i(\alpha f(\alpha^*)) = - \sum_{\alpha \in A' \setminus \{\alpha^*, \alpha\}} \tau_W(\beta_i(\alpha f(\alpha^*))$$

As the result, it can reconstruct $m$ target traces $\tau_W(\beta_i(\alpha f(\alpha^*)) = \tau_W(\beta_i(\alpha f(\alpha^*))$, $i \in [m]$, of $f(\alpha^*)$. The bandwidth spent is at most $(n - 2)(\ell - m)$ subsymbols over $\mathbb{F}_q$.

**Collaboration Phase.** In this phase, the two RNs exchange information to help each other recover the last $\ell - m$ missing target traces. To this end, the following two sets of $\ell - m$ repair equations each for $f(\alpha^*)$ based on $g_{m+1}, \ldots, g_\ell$ and for $f(\alpha^*)$ based on $h_{m+1}, \ldots, h_\ell$ can be used. For $i \in [m + 1, \ell]$,

$$g_i(\alpha^*)f(\alpha^*) + g_i(\alpha f(\alpha^*)) = - \sum_{\alpha \in A' \setminus \{\alpha^*, \alpha\}} g_i(\alpha f(\alpha^*))$$

(9)

$$h_i(\alpha^*)f(\alpha^*) + h_i(\alpha f(\alpha^*)) = - \sum_{\alpha \in A' \setminus \{\alpha^*, \alpha\}} h_i(\alpha f(\alpha^*))$$

(10)
It is clear that from the repair traces collected in the Download Phase that the right-hand-sides of (9) and (10) can be determined. However, to determine the target traces Tr$(g_i(\alpha^s) f(\alpha^s))$, the RN for $f(\alpha^s)$ also needs to know the repair traces Tr$(g_i(\beta) f(\beta))$, $i \in [m + 1, \ell]$. It turns out that these repair traces can be deduced from the target traces Tr$(h_i(\alpha^s) f(\alpha^s))$, $i \in [m]$, which are already available at the RN for $f(\alpha^s)$ (see Fig. [1] for an illustration). Indeed, by Lemma 3 (d), $g_i(\alpha^s) \in \text{span}_F \{\tau_W x_1, \ldots, \tau_W x_m\}$ for $i \in [m + 1, \ell]$. Hence, the RN for $f(\alpha^s)$ can compute Tr$(g_i(\beta) f(\beta))$, $i \in [m + 1, \ell]$, as linear combinations of Tr$(\tau_W x_1 f(\beta)) = \text{Tr}(h_i(\alpha^s) f(\beta))$, $i \in [m]$, and send these repair traces over to the RN for $f(\alpha^s)$. Likewise, the RN for $f(\alpha^s)$ can compute Tr$(h_i(\alpha^s) f(\alpha^s))$, $i \in [m + 1, \ell]$, based on Tr$(\tau_W x_1 f(\alpha^s)) = \text{Tr}(g_i(\alpha^s) f(\alpha^s))$, $i \in [m]$, and send these repair traces to the RN for f(\beta). Once all repair traces are available, each RN can recover the missing $\ell - m$ target traces and then the corresponding codeword symbol.

The bandwidth used in the Collaboration Phase is $\ell - m$ subsymbols over $\mathbb{F}_q$ per erasure. Combining with that in the Download Phase, we conclude that this repair scheme incurs a bandwidth of at most $(n - 1)(\ell - m)$ subsymbols over $\mathbb{F}_q$.

An important question to ask is whether there exists a subspace $W$ satisfying both properties (P1) and (P2) listed in Theorem 2. In the remainder of this subsection, we establish the existence of such a subspace when $q$ and $\ell$ are even and $m \geq \ell/2$.

We first describe our key idea. Consider even $q$ and $\ell$, and $m \geq \ell/2$. Note that it is necessary that $m \geq \ell/2$ for (P2) to hold. Indeed, (P2) implies that $\text{im}(L_W) \subseteq \ker(L_W) = W$, which means that $\ell - m = \dim_F(\text{im}(L_W)) \leq \dim_F(\ker(L_W)) = m$, or $m \geq \ell/2$. It seems quite difficult to construct directly a subspace $W$ satisfying both (P1) and (P2). Our strategy is to first construct a subspace satisfying (P1) and then turn it into a subspace satisfying both (P1) and (P2). The construction of $W$ is broken down into two steps.

- **Step 1** Constructing a subspace $U$ of dimension $m - \ell/2$ in $\mathbb{F}_{q^s}/\mathbb{F}_q$ satisfying $\tau_U = \pm 1 \in \mathbb{F}_q$.
- **Step 2** Using $U$ and $\mathbb{F}_{q^s}/\mathbb{F}_q$ to construct $W$, which satisfies both (P1) and (P2) (Lemma 5).

**Lemma 4.** For all $m \in [\ell]$, there always exists an $m$-dimensional $\mathbb{F}_q$-subspace $U$ of $\mathbb{F}_{q^s}$ satisfying $\tau_U = \pm 1 \in \mathbb{F}_q$.

**Proof.** We prove the following two claims, which together establish Lemma 4.

**Claim 1.** For $1 \leq m \leq \ell$ and $s = \text{gcd}(m,\ell)$, there exists an $\mathbb{F}_q$-subspace $U_0$ of $\mathbb{F}_{q^s}$ of dimension $m$ over $\mathbb{F}_q$ satisfying $\tau_{U_0} = \prod_{\alpha \in U_0 \setminus \{0\}} u_\alpha = \pm \zeta^x(q^{s-1} - 1)$, for some integer $x$, where $\zeta$ is a primitive element of $\mathbb{F}_{q^s}$.

To prove Claim 1, take $U_0$ to be an $m/s$-dimensional $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^s}$. Such a subspace always exists because $m/s \in \mathbb{Z}$ and $\ell/s \in \mathbb{Z}$. Then $\dim_{\mathbb{F}_q}(U_0) = m$. Define a relation $\sim$ in $U_0 = U_0 \setminus \{0\}$ as follows: for $u, v \in U_0$, $u \sim v$ if $u/v \in \mathbb{F}_{q^s}$. One can verify that this is an equivalence relation and its equivalence classes are multiplicative cosets of $\mathbb{F}_{q^s}$, which are of the form $u\mathbb{F}_{q^s}$, $u \in U_0$. Let $\{u_i\mathbb{F}_{q^s} : i = 1, \ldots, \frac{m}{s} - 1\}$ be the set of all disjoint multiplicative cosets of $\mathbb{F}_{q^s}$ in $U_0$, each of which is of size $q^s - 1$. Then,

$$U_0 = \bigcup_{i=1}^{\frac{m}{s} - 1} u_i\mathbb{F}_{q^s},$$

which implies that

$$\tau_{U_0} = \left(\prod_{i=1}^{\frac{m}{s} - 1} u_i\right)^{q^{s-1} - 1}.$$

Note that $\tau_{U_0} = -1$. Therefore, choosing an integer $x$ such that $\zeta^x = \prod_{i=1}^{\frac{m}{s} - 1} u_i$, we have $\tau_{U_0} = \pm \zeta^x(q^{s-1} - 1)$, as claimed.

**Claim 2.** Suppose that there exists an $\mathbb{F}_q$-subspace $U_0$ of $\mathbb{F}_{q^s}$ of dimension $m$ over $\mathbb{F}_q$ satisfying $\tau_{U_0} = \pm \zeta^x(q^{s-1} - 1)$ for some integer $x$, where $s = \text{gcd}(m,\ell)$. Then there exists an $m$-dimensional $\mathbb{F}_q$-subspace $U$ of $\mathbb{F}_{q^s}$ satisfying $\tau_U = \pm 1$.

To prove Claim 2, note that as $\text{gcd}(m,\ell) - 1 = q^s - 1$ (based on Euclid’s algorithm), there exist integers $y$ and $z$ satisfying

$$x(q^s - 1) + y(q^m - 1) = z(q^\ell - 1).$$

Set $\gamma = \zeta^y$ and $U \triangleq \gamma U_0$. Then

$$\tau_U = \gamma q^m - \tau_{U_0} = \pm \zeta^y(q^{m-1} - 1) \zeta^x(q^{s-1} - 1) = \pm \zeta^{yq^m} = \pm 1.$$

Claim 1 and Claim 2 prove Lemma 4.

**Lemma 5** is referred to as the **Reduction Lemma** because it reduces the existence of a subspace of $\mathbb{F}_{q^s}$ satisfying (P1) and (P2) to the existence of a subspace of $\mathbb{F}_{q^s/2}$ satisfying (P1) only.

**Lemma 5 (Reduction Lemma).** Suppose that $q$ and $\ell$ are even and $m \in [\ell/2 + 1, \ell]$. If there exists an $\mathbb{F}_q$-subspace $U$ of $\mathbb{F}_{q^s/2}$ of dimension $m - \ell/2$ satisfying (P1), i.e., $\tau_U \in \mathbb{F}_q$, then there exists an $\mathbb{F}_q$-subspace $W$ of $\mathbb{F}_{q^s/2}$ of dimension $m$ satisfying both (P1) and (P2), i.e., $\tau_W \in \mathbb{F}_q$ and $\mathbb{F}_{q^s/2}(\mathbb{F}_q) = \{0\}$.

**Proof.** Denote by $\sigma$ the trace function $\text{Tr}_{\mathbb{F}_{q^s/2}/\mathbb{F}_q}$, which is an onto function (see [27], Thm. 2.23). Then $\sigma(v) = v^{q^{s/2}} + v$. As $\mathbb{F}_q$ has characteristic two, $\sigma(v) = v^{q^{s/2}} - v = \prod_{\beta \in \mathbb{F}_{q^s/2}} (\beta - v)$. This means that $\sigma(x)$ is also the subspace polynomial of $\mathbb{F}_{q^s/2}$. Therefore, $\ker(\sigma) = \mathbb{F}_{q^s/2}$.

Our proof consists of two steps. First, based on $U$, we construct an $\mathbb{F}_q$-subspace $V$ of $\mathbb{F}_{q^s}$ of dimension $m - \ell/2$ such that $V \cap \mathbb{F}_{q^s/2} = \{0\}$ and $\sigma(V) = U$. Then, $W = V \oplus \mathbb{F}_{q^s/2}$ is the desired subspace. Indeed, we have $\dim_{\mathbb{F}_q}(W) = \ell/2 + (m - \ell/2) = m$. Moreover, as

$$W^* = \mathbb{F}_{q^s/2}^* \cup \left(\bigcup_{v \in V^*} (v + \mathbb{F}_{q^s/2})\right),$$

we have

$$\tau_W = \tau_{U_0} \prod_{v \in V^*} \left(\prod_{\beta \in \mathbb{F}_{q^s/2}} (\beta + v)\right) = \prod_{v \in V^*} \left(\prod_{\beta \in \mathbb{F}_{q^s/2}} (\beta - v)\right) = \prod_{v \in V^*} \sigma(v) = \tau_U \in \mathbb{F}_q.$$
where the last equality is due to the fact that \( \text{im}(\sigma) = \mathbb{F}_{q^2}/\mathbb{F}_q \).

Since \( W = \mathbb{F}_{q^2} \oplus V \supseteq \mathbb{F}_{q^2}/\mathbb{F}_q \), we have \( L_W(x) = H(x)|\sigma(x) \), where \( H(x) \in \mathbb{F}_{q^2}[x] \) and \( \sigma(x) \), as defined earlier, is the subspace polynomial of \( \mathbb{F}_{q^2}/\mathbb{F}_q \). Since \( L_W(\alpha) \in \mathbb{F}_{q^2}/\mathbb{F}_q \) and \( \ker(\sigma) = \mathbb{F}_{q^2}/\mathbb{F}_q \), we deduce that

\[
L_W(L_W(\alpha)) = H(L_W(\alpha))\sigma(L_W(\alpha)) = H(L_W(\alpha)) \times 0 = 0.
\]

Thus, \( W \) satisfies (P2) as well.

We now discuss the construction of \( V \), which has dimension \( m - \ell/2 \) and satisfies \( V \cap \mathbb{F}_{q^2}/\mathbb{F}_q = \{0\} \) and \( \sigma(V) = U \). Let \( \{u_{j}\}_{j=1}^{m-\ell/2} \) be an \( \mathbb{F}_q \)-basis of \( U \subseteq \mathbb{F}_{q^2}/\mathbb{F}_q \). As \( \sigma \) is onto, there exists \( m - \ell/2 \) elements \( v_1, \ldots, v_{m-\ell/2} \) satisfying \( \sigma(v_j) = u_j \) for all \( j \in [m - \ell/2] \). We claim that the set \( \{v_j\}_{j=1}^{m-\ell/2} \) is \( \mathbb{F}_q \)-linearly independent. Indeed, suppose there exist \( a_1, \ldots, a_{m-\ell/2} \) in \( \mathbb{F}_q \) so that \( 0 = \sum_{j=1}^{m-\ell/2} a_j v_j \). Applying \( \sigma \) to both sides of this equation, we obtain \( 0 = \sum_{j=1}^{m-\ell/2} a_j u_j \), which implies that \( a_j = 0 \) for all \( j \in [m - \ell/2] \).

Set \( V = \text{span}_{\mathbb{F}_q} \{v_j\}_{j=1}^{m-\ell/2} \). Then \( \dim_{\mathbb{F}_q}(V) = m - \ell/2 \) and \( \sigma(V) = U \). Moreover, \( \mathbb{F}_{q^2}/\mathbb{F}_q \cap V = \{0\} \). Indeed, as \( \sigma(V) = U \) and \( \dim_{\mathbb{F}_q}(V) = \dim_{\mathbb{F}_q}(U) \), the only element in \( V \) that is mapped to 0 by \( \sigma \) is 0, while \( \sigma(\mathbb{F}_{q^2}/\mathbb{F}_q) = \{0\} \). Hence, \( \mathbb{F}_{q^2}/\mathbb{F}_q \cap V = \{0\} \).

Combining Lemma 4 and Lemma 5, we can show that Construction II works for even \( q \) and \( \ell \geq 2 \).

**Corollary 2.** Suppose that \( 2 \leq \ell, m \geq \ell/2 \), and \( q = 2^s, s \geq 1 \). Then there exists an \( m \)-dimensional \( \mathbb{F}_q \)-subspace \( W \) of \( \mathbb{F}_{q^2} \) that satisfies the properties (P1) and (P2) in Theorem 2. Hence, there exists a distributed scheme repairing two erasures for any \( [n, k] \) Reed-Solomon code over \( \mathbb{F}_{q^2} \) with \( r = n - k \geq q^m \) using a repair bandwith of at most \((n-1)(\ell-m)/2\) subfields in \( \mathbb{F}_q \) per erasure.

**Proof.** If \( m = \ell/2 \) then set \( W = \mathbb{F}_{q^2}/\mathbb{F}_q \). Then \( \tau_W = -1 \in \mathbb{F}_q \). Hence, \( W \) satisfies (P1), as shown in the first paragraph in the proof of Lemma 5. Let \( L_W(x) = \sigma(x) = \text{Tr}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(x) \). Therefore, \( L_W(\mathbb{F}_{q^2}) = \mathbb{F}_{q^2}/\mathbb{F}_q = W = \ker(L_W) \). Equivalently, \( L_W(L_W(\mathbb{F}_{q^2})) = \{0\} \), which shows that \( W \) satisfies (P2).

Now suppose that \( m > \ell/2 \). By Lemma 3, there exists an \((m-\ell/2)\)-dimensional \( \mathbb{F}_q \)-subspace \( U \) of \( \mathbb{F}_{q^2}/\mathbb{F}_q \) with \( \tau_U \in \mathbb{F}_q \). Note that we replace \( m \) by \( m - \ell/2 \) and \( \ell \) by \( \ell/2 \) in Lemma 3. Then by Lemma 5, there exists an \( \mathbb{F}_q \)-subspace \( W \) of \( \mathbb{F}_{q^2} \) satisfying both (P1) and (P2). Applying Theorem 2 to \( W \), we conclude that there exists a repair scheme for the Reed-Solomon code with the desired bandwidth.

We summarize below the steps to construct a subspace \( W \) satisfying both (P1) and (P2) which will prove Theorem 2 for \( m > \ell/2 \). Note that when \( m = \ell/2 \), we set \( W = \mathbb{F}_{q^2}/\mathbb{F}_q \).

1. **Step 1.** Let \( U_0 \) be an \( \frac{m-\ell/2}{2} \)-dimensional \( \mathbb{F}_q \)-subspace of \( \mathbb{F}_{q^2}/\mathbb{F}_q \), where \( s = \gcd(m - \ell/2, 2) \).

2. **Step 2.** Compute \( \tau_{U_0} = \prod_{u \in U_0 \setminus \{0\}} u = \pm \xi^{|x(q^e-1)} \), where \( \xi \) is a primitive element of \( \mathbb{F}_{q^2}/\mathbb{F}_q \) and \( x \in \mathbb{Z} \).

3. **Step 3.** Set \( U = \sigma(U_0) \), where \( \gamma = \xi^y \) and \( y \in \mathbb{Z} \) such that \( x(q^e-1) + y(q^e-2) = 1 \) for some \( e \in \mathbb{Z} \).

4. **Step 4.** Let \( U \) be an \((m-\ell/2)\)-dimensional \( \mathbb{F}_q \)-vector space of \( \mathbb{F}_{q^2} \) constructed as follows. For an \( \mathbb{F}_q \)-basis \( \{u_i\}_{i=1}^{m-\ell/2} \) of \( U \), choose a set \( \{v_i\}_{i=1}^{m-\ell/2} \in \mathbb{F}_{q^2} \) so that \( \text{Tr}_{\mathbb{F}_{q^2}/\mathbb{F}_q}(v_i) = u_i \) for all \( i \in [m - \ell/2] \). Set \( V = \text{span}_{\mathbb{F}_q} \{v_i\}_{i=1}^{m-\ell/2} \).

5. **Step 5.** Set \( W = \mathbb{F}_{q^2}/\mathbb{F}_q \oplus V \).
involve multiple rounds of communications. This makes sense intuitively because when \( m \) is small compared to \( \ell \), the amount of information (i.e., the number of erasures) each RN knows about its erased codeword symbol after the Download Phase is insufficient to help the other RN recover its content in only one round of communication.

**Theorem 3.** Consider a Reed-Solomon code of full length \( n = q^\ell \) and \( r \geq q^m \) over \( \mathbb{F}_q \). Suppose there exists an \( \mathbb{F}_q \)-subspace \( W \) of \( \mathbb{F}_q^r \) of dimension \( m \) satisfying (with \( t = \ell \mod m \))

(P1) \( \tau_W := \prod_{w \in W} w \in \mathbb{F}_q \),

(P3) \( \text{im}(L_W(\gamma_{\lfloor \ell/m \rfloor})) \cap W \) has dimension at least \( t \) over \( \mathbb{F}_q \),

(P4) \( \text{im}(L_W(\gamma_{\lfloor \ell/m \rfloor}^\ell)) \supseteq W \).

Then there exists a scheme that can repair two arbitrary erasures for this code using a bandwidth of at most \((n-1)(\ell-m)\) subsymbols over \( \mathbb{F}_q \) per erasure.

Construction III generates special check polynomials that allow a batch-by-batch reconstruction of traces, the exact meaning of which will be made clear in the description of the Collaboration Phase.

**Construction III.** Suppose \( W \) is an \( \mathbb{F}_q \)-subspace of \( \mathbb{F}_q^r \) of dimension \( m \) satisfying (P1), (P3), and (P4). By (P3), we can select a set of \( t \) \( \mathbb{F}_q \)-linearly independent elements \( \gamma_1, \ldots, \gamma_t \) from the intersection \( \text{im}(L_W(\gamma_{\lfloor \ell/m \rfloor})) \cap W \). Moreover, due to (P4), we can find \( \gamma_{t+1}, \ldots, \gamma_m \) in \( W \subseteq \text{im}(L_W(\gamma_{\lfloor \ell/m \rfloor}^\ell)) \), so that \( \{\gamma_1, \ldots, \gamma_m\} \) forms an \( \mathbb{F}_q \)-basis of \( W \). By definitions of \( \gamma_1, \ldots, \gamma_m \), there exist \( \gamma_{m+1}, \ldots, \gamma_{\ell t} \) satisfying the Chain Property defined as follows.

\[ \begin{align*}
\gamma_j &= L_W(\gamma_{m+j}) = L_W(\gamma_{2m+j}) = \cdots = L_W(\gamma_{\ell m+j}) \quad \text{for } j \in [t], \\
\gamma_j &= L_W(\gamma_{m+j}) = L_W(\gamma_{2m+j}) = \cdots = L_W(\gamma_{\ell m+j}) \quad \text{for } j \in [t+1, m].
\end{align*} \]

In other words, \( \gamma_{m+1}, \ldots, \gamma_{\ell t} \) are chosen so that \( L_W(\gamma_i) = \gamma_i \) for all \( i \in [m+1, \ell] \). Finally, we set \( \beta_i = \gamma_i / (\alpha - \alpha^*) \) for \( i \in [\ell] \) and choose the two sets of check polynomials as before.

\[ \begin{align*}
 g_i(x) &= L_W(\beta_i(\alpha - x^*)) / (x - \alpha^*), \quad i \in [\ell], \\
h_i(x) &= L_W(\beta_i(x - \alpha^*)) / (x - \alpha^*), \quad i \in [\ell].
\end{align*} \]

For example, when \( \ell = 8 \), \( m = 3 \), \( t = 2 \), we have

\[ \begin{align*}
 \gamma_1 &= L_W(\gamma_4) = L_W^2(\gamma_7), \\
 \gamma_2 &= L_W(\gamma_5) = L_W^2(\gamma_8), \\
 \gamma_3 &= L_W(\gamma_6).
\end{align*} \]

**Lemma 6.** Let \( W, \gamma_i, \beta_i, g_i, h_i \) be defined as in Construction III. Then the following statements hold.

(a) \( \deg(g_i(x)) = \deg(h_i(x)) = q^m - 1 \leq r - 1 \), for all \( i \in [\ell] \).

(b) \( g_i(x^*) = h_i(x^*) = \tau_W \beta_i \), for all \( i \in [\ell] \).

(c) \( g_i(\alpha^*) = h_i(\alpha^*) = 0 \), for all \( i \in [m] \).

(d) \( \{\gamma_i\}_{i=1}^m \) and \( \{\beta_i\}_{i=1}^m \) are \( \mathbb{F}_q \)-bases of \( \mathbb{F}_q^r \).

(e) \( g_i(\alpha) = h_i(\alpha) = \beta_i \) for \( i \in [m+1, \ell] \).

**Proof of Lemma 6** The first three statements follow in the same way as those in the proof of Lemma 3. For (d) to hold, it suffices to show that the set \( \{\gamma_i\}_{i=1}^m \) is \( \mathbb{F}_q \)-linearly independent. We prove this by induction.

First, \( \gamma_1, \ldots, \gamma_m \) are \( \mathbb{F}_q \)-linearly independent by definition. Suppose that \( \gamma_1, \ldots, \gamma_s \), where \( s \in [m, \ell - 1] \), are linearly independent. We aim to show that \( \gamma_1, \ldots, \gamma_{s+1} \) are also \( \mathbb{F}_q \)-linearly independent. Assume that we can write \( \gamma_{s+1} = \sum_{i=1}^s a_i \gamma_i \), for some \( a_i \in \mathbb{F}_q \). Applying \( L_W \) to both sides of this equation and noting that \( L_W \) is \( \mathbb{F}_q \)-linear, we have

\[ L_W(\gamma_{s+1}) - \sum_{i=1}^s a_i L_W(\gamma_i) = 0. \]

Note that \( L_W(\gamma_i) = 0 \) for \( i \in [m] \) as such \( \gamma_i \) belongs to \( W \) and moreover, \( L_W(\gamma_i) = \gamma_{i-m} \) for \( i \in [m+1, \ell] \). Hence, the above equation implies that there exists a nontrivial \( \mathbb{F}_q \)-linear combination of \( \gamma_1, \ldots, \gamma_s \) equal to zero, which contradicts our induction hypothesis. Therefore, \( \gamma_1, \ldots, \gamma_{s+1} \) must also be \( \mathbb{F}_q \)-linearly independent.

Finally, to prove (e), again using the fact that \( L_W(\gamma_i) = \gamma_{i-m} \), for \( i \in [m+1, \ell] \), we have

\[ g_i(\alpha^*) = L_W(\beta_i(\alpha - \alpha^*)) / (\alpha - \alpha^*) = \gamma_{i-m} / (\alpha - \alpha^*) = \beta_{i-m}. \]

Similarly,

\[ h_i(\alpha^*) = L_W(\beta_i(\alpha - \alpha^*)) / (\alpha - \alpha^*) = L_W(\gamma_{i-m}) / (\alpha - \alpha^*) = \gamma_{i-m} / (\alpha - \alpha^*) = \beta_{i-m}. \]

This completes the proof of Lemma 6. 

**Proof of Theorem 3** We use a two-phase repair scheme based on the check polynomials produced by Construction III that can repair \( f(\alpha^*) \) and \( f(\alpha) \) with a bandwidth of \((n-1)(\ell-m)\) subsymbols per erasure. The Download Phase is the same as that in the proof of Theorem 2 and is hence omitted. We note that when the Download Phase is completed,

- the RN for \( f(\alpha^*) \) has obtained \( m \) target traces of \( f(\alpha^*) \):

\[ \text{Tr}(g_i(\alpha^*) f(\alpha^*)) = \text{Tr}(\tau_W \beta_i f(\alpha^*)), \quad i \in [m]. \]

- the RN for \( f(\alpha) \) has obtained \( m \) target traces of \( f(\alpha) \):

\[ \text{Tr}(h_i(\alpha) f(\alpha)) = \text{Tr}(\tau_W \beta_i f(\alpha)), \quad i \in [m]. \]

Each RN has used a bandwidth of \((n-2)(\ell-m)\) subsymbols.

The Collaboration Phase consists of \([\ell/m] \) rounds. In the first round, by Lemma 3 (c), based on the target traces obtained in the Download Phase, the two RNs can construct and exchange the following \( m \) repair traces \((i \in [m+1, 2m])\)

\[ \text{Tr}(g_i(\alpha^*) f(\alpha)) = \text{Tr}(h_{i-m} f(\alpha)) / \tau_W \quad \text{Tr}(h_{i-m} f(\alpha^*)) = \text{Tr}(g_i(\alpha^*) f(\alpha^*)) / \tau_W, \]

which subsequently enable them to determine the target traces \( \text{Tr}(\beta_i f(\alpha^*)) \) and \( \text{Tr}(\beta_i f(\alpha)) \), \( i \in [m+1, 2m] \), respectively.
using the corresponding repair equations. Subsequent rounds are carried out in a similar manner, each of which allows each RN to construct and exchange \( m \) repair traces based on the target traces recovered in the previous round. These repair traces in turn will allow the RNs to recover a batch of new \( m \) target traces. An exception is when \( t = \ell \mod m > 0 \), as the batch of traces recovered in the last round consists of \( t \) traces instead of \( m \).

Continuing the example for \( \ell = 8, m = 3, \) and \( t = 2 \), the traces repaired in the two phases are illustrated in Fig. \[2\] We now describe two sets of parameters \( \ell \) and \( m \) for which Construction III is feasible.

**Corollary 3.** Suppose that \( \ell/m \in \mathbb{Z} \) is a power of \( q \). Then there exists an \( m \)-dimensional \( \mathbb{F}_q \)-subspace \( W \) of \( \mathbb{F}_{q^\ell} \) that satisfies the properties (P1), (P3), and (P4) in Theorem \[3\]. Moreover, \( L_W(x) = x^{q^m} - x \). Hence, there exists a scheme repairing two erasures for full-length Reed-Solomon code over \( \mathbb{F}_{q^\ell} \) with \( r \geq q^m \) when \( \ell/m \) is a power of \( q \). The required bandwidth is at most \( (n-1)(\ell-m) \) subspaces per erasure.

**Proof.** Set \( W = \mathbb{F}_{q^m} \). Since \( m \mid \ell \), \( W \) is an \( \mathbb{F}_q \)-subspace of \( \mathbb{F}_{q^\ell} \) and has dimension \( m \) over \( \mathbb{F}_q \). Then \( L_W(x) = x^{q^m} - x \) (Fermat’s theorem, see, for example, [5, Ch. 4, Cor. 3]). Hence, \( \tau_W = -1 \in \mathbb{F}_q \), i.e., \( W \) satisfies (P1). Since \( t = \ell \mod m = 0 \), (P3) is trivially satisfied. It remains to show that (P4) holds. As introduced in Section \[1\], let \( l(x) = x^{m-1} \) be the associate of \( L_W(x) \). Then \( l(x^{\ell/m}) \) is the associate of \( L_{\ell/m}(x) \). Using the assumption that \( \ell/m \) is a power of \( q \), we have

\[
L_{\ell/m}(x) = (x - 1)^{\ell/m - 1} = \frac{(x - 1)^{\ell/m}}{x^{m-1}} = x^{m-1} - 1.
\]

Therefore,

\[
L_{\ell/m}(x) = \sum_{i=0}^{\ell/m - 1} x^{mi} = \sum_{i=0}^{\ell/m - 1} x^{(q^m)^i} = \text{Tr}_{\mathbb{F}_{q^\ell}/\mathbb{F}_{q^m}}(x).
\]

As the trace function is onto (see, for instance, [27, Thm. 2.23]), we have \( \text{im}(L_{\ell/m}) = \mathbb{F}_{q^m} = W \). Thus, (P4) is satisfied.

**Corollary 4.** Suppose that \( \ell = q^a \) and \( m = q^b - 1 > 1 \) for some \( a \geq b \geq 1 \). Then there exists an \( m \)-dimensional \( \mathbb{F}_q \)-subspace \( W \) of \( \mathbb{F}_{q^\ell} \) that satisfies the properties (P1), (P3), and (P4) in Theorem \[3\]. Moreover, \( L_W(x) = \text{Tr}_{\mathbb{F}_{q^{m+1}}/\mathbb{F}_q}(x) = \sum_{i=0}^{q^m} x^i \). Hence, there exists a scheme repairing two erasures for full-length Reed-Solomon code over \( \mathbb{F}_{q^\ell} \) with \( r \geq q^m \) when \( \ell = q^a \) and \( m = q^b - 1 \) for some \( 1 \leq b \leq a \). The required bandwidth is at most \( (n-1)(\ell-m) \) subspaces per erasure.

We need an auxiliary result for the proof of Corollary \[4\].

**Lemma 7.** Suppose that \( \ell = q^a \) and \( m = q^b - 1 > 1 \) for some \( a \geq b \geq 1 \). Let \( t = \ell \mod m \). Then

\[
\text{Tr}_{\mathbb{F}_{q^{m+1}}/\mathbb{F}_q}(x) = \text{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(x).
\]

**Proof.** Note that the associates of \( \text{Tr}_{\mathbb{F}_{q^{m+1}}/\mathbb{F}_q}(x) \) and \( \text{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(x) \) are \( p_1(x) = \sum_{i=0}^{m} x^i \) and \( p_2(x) = \sum_{i=0}^{\ell/t-1} x^{it} \), respectively. By [27, Lem. 3.59], for (11), it suffices to show that \( p_1(x) = p_2(x) \). To this end, first let \( c = a \mod b \), so that \( t = q^c \mod (q^b - 1) = q^c \). Hence,

\[
\frac{\ell - m}{m} = \frac{\ell - t}{m} = \frac{q^c(q^c - 1)}{q^b - 1} = q^c \sum_{j=0}^{\frac{q^c-1}{q^b-1}} q^{bj}.
\]

Therefore,

\[
L_{\ell/m}(x) = \left( \sum_{i=0}^{\ell/m} x^i \right)^{\ell/m - 1} = \sum_{j=0}^{\ell/m - 1} \left( \sum_{i=0}^{q^m} x^{ij} \right)^{\ell/m - 1} = \sum_{j=0}^{\ell/m - 1} \left( \sum_{i=0}^{\ell/m} x^{ij} \right)^{\ell/m - 1} = \sum_{j=0}^{\ell/m - 1} \frac{q^{bj}}{q^{b-1} - 1} = \sum_{j=0}^{\ell/m - 1} \left( \frac{q^{bj}}{q^{b-1} - 1} \right) = \frac{q^{bj}}{q^{b-1} - 1} = \sum_{j=0}^{\ell/m - 1} x^{ij} = x^{ij}.
\]

where the second to last equality is due to the fact that the set

\[
\left\{ \sum_{j=0}^{\frac{q^c-1}{q^b-1}} i q^{jb} : 0 \leq i, b, \ldots, \frac{q^c-1}{q^b-1} \leq m \right\}
\]

comprises the representations of all integers from 0 to \( \ell/t - 1 = q^{c-1} - 1 \) in base \( q^b \). The last equality follows because \( t = q^c \).

**Proof of Corollary \[4\]** First, if we set \( W \) to be the kernel (in \( \mathbb{F}_{q^{m+1}} \)) of the trace function of \( \mathbb{F}_{q^{m+1}} \) over \( \mathbb{F}_q \), then by the rank-nullity theorem (see, e.g. [30, p. 70]), \( \dim_{\mathbb{F}_q}(W) = m \). Note that as \( m + 1 = q^b \) divides \( \ell = q^a \), \( W \subset \mathbb{F}_{q^{m+1}} \subset \mathbb{F}_{q^\ell} \). Clearly, \( L_W(x) = \text{Tr}_{\mathbb{F}_{q^{m+1}}/\mathbb{F}_q}(x) \). Since \( \tau_W \) is the same as the coefficient of \( x \) in \( L_W(x) \), we deduce that \( \tau_W = 1 \in \mathbb{F}_q \). Therefore, \( W \) satisfies (P1).

We now show that (P3) holds. By Lemma \[7\]

\[
L_{\ell/m}(x) = \text{Tr}_{\mathbb{F}_{q^{m+1}}/\mathbb{F}_q}(x) = \text{Tr}_{\mathbb{F}_{q^t}/\mathbb{F}_q}(x).
\]

Therefore, since the trace is an onto map, we have

\[
\text{im}(L_{\ell/m}) = \mathbb{F}_{q^m} = W.
\]

In order to show that \( \text{im}(L_{\ell/m}) \cap W \) has dimension at least \( t \), it suffices to prove that \( W \supseteq \mathbb{F}_q \). Equivalently, we aim to show that \( L_W(x) \) is divisible by \( L_{\mathbb{F}_q}(x) = x^{q^t} - x \). By [27, Thm. 3.62], this holds if and only if the associate \( \sum_{i=0}^{q^m} x^i \) of \( L_W(x) \) is divisible by the associate \( x^{\ell-1} \) of \( L_{\mathbb{F}_q}(x) \).

To this end, note that \( t = q^c \) divides \( m+1 = q^b \), and therefore,

\[
\sum_{i=0}^{m} x^i = \frac{x^{m+1} - 1}{x - 1} = \frac{(x^t - 1) \sum_{j=0}^{(m+1)/t-1} x^{ij}}{x - 1},
\]

which is divisible by \( x^t - 1 \) because \( \sum_{j=0}^{(m+1)/t-1} x^{ij} \) is divisible by \( x - 1 \). The latter holds since over \( \mathbb{F}_q \) we have

\[
\sum_{j=0}^{(m+1)/t-1} 1^{ij} = (m+1)/t = q^{b-c} = 0.
\]

Thus, (P3) holds.
As the last step, we demonstrate that $W$ also satisfies (P4). Our goal is to show that $W \subseteq K \triangleq \im(L_W^{\otimes \lfloor \frac{t-m}{t} \rfloor})$. Note that
\[
\dim_{F_q}(\ker(L_W^{\otimes \lfloor \frac{t-m}{t} \rfloor})) \leq \log_q(\deg(L_W^{\otimes \lfloor \frac{t-m}{t} \rfloor})) = \ell - t - m,
\]
where we used the fact that
\[
\log_q(\deg(L_W^{\otimes \lfloor \frac{t-m}{t} \rfloor})) = m \left( \frac{\ell - m}{m} \right) = m(\ell - t - 1) = \ell - t - m.
\]
Therefore,
\[
\dim_{F_q}(K) = \dim_{F_q} \left( \im \left( L_W^{\otimes \lfloor \frac{t-m}{t} \rfloor} \right) \right) = \ell - t - m - (t + m) = t + m,
\]
where $K$ denotes the kernel of $L_W$ restricted to $K$. Since $W = \ker(K(L_W))$, we deduce that $W = \ker(K(L_W)) \subseteq K = \im(L_W^{\otimes \lfloor \frac{t-m}{t} \rfloor})$. Thus, (P4) is satisfied.

Example 4. Consider the previous example (Fig. 2) for $q = 2$, $\ell = 8 = 2^3$, $m = 3 = 2^2 - 1$, and $t = \ell \mod m = 2$. We illustrate next the key steps of Construction III.

As discussed in the proof of Corollary 4, we set $W$ to be the kernel of the trace $\Tr_{F_2^4/F_2}$ in $F_{2^t}$, that is,
\[
W = \{0, 1, \xi^{85}, \xi^{170}, \xi^{17}, \xi^{34}, \xi^{68}, \xi^{134}\} \subset F_{2^t},
\]
where $\xi$ is a primitive element of $F_{2^8}$. Note that $\xi^{85}$ and $\xi^{17}$ are primitive elements of $F_{2^2}$ and $F_{2^4}$, respectively. Moreover, $L_W(x) = \Tr_{F_2^4/F_2}(x) = x^2 + x^2 + x^2 + x$ and $L_W^{\otimes 2}(x) = \Tr_{F_2^{2^8}/F_2}(x)$. Also, $L_W^{\otimes 2}(x) = \Tr_{F_2^{2^8}/F_2}(x)$ and $\im(L_W^{\otimes 2}) = \im(L_W^{\otimes 2}) = F_{2^2} \subset W$. Therefore,
\[
\im(L_W^{\otimes 2}) \cap W = F_{2^2} = \{0, 1, \xi^{85}, \xi^{170}\}.
\]
We need to find $\{\gamma_i\}_{i=1}^8$ before we can determine $\{\beta_i\}_{i=1}^8$. Set $\gamma_1, \gamma_2 = \{1, \xi^{85}\}$, which is a $F_2$-linearly independent set. Next, we pick $\gamma_3 = \xi^{17}$ to make $\gamma_1, \gamma_2, \gamma_3$ an $F_2$-basis of $W$. Then, we set $\gamma_4 = \xi^{51}, \gamma_5 = \xi^7, \gamma_6 = \xi^{13}, \gamma_7 = \xi^3,$ and $\gamma_8 = \xi^{53}$. It is easy to verify that $\Ch(P)$ is indeed a $F_2$-linearly independent set. Finally, we set $\beta_i = \gamma_i/(\overline{\alpha} - \alpha)$ for $i \in [8]$, and use these $\beta_i$'s and $L_W(x)$ for generating the checks $\{g_i(x)\}_{i=1}^8$ and $\{h_i(x)\}_{i=1}^8$ to repair $f(\alpha^c)$ and $f(\overline{\alpha})$.

VI. CONCLUSIONS

We proposed several repair schemes for a single erasure and two erasures in Reed-Solomon codes over $\mathbb{F}_{q^t}$. Our schemes were constructed using subspace polynomials offering optimal repair bandwidths of $(n - 1)(\ell - m) \log_2(q)$ bits for codes of full length $n = q^t$ and $r = q^m$, $1 \leq m \leq \ell$, for the case of one erasure. For two simultaneous erasures, our distributed schemes were shown to achieve the same repair bandwidth per erasure for certain ranges of parameters $q$, $\ell$, and $m$. It remains an open problem to construct repair schemes for two or more erasures that work for all possible parameter choices. Another interesting open problem is to further improve the lower bound on the repair bandwidth, which currently appear to be quite loose for $n << q^t$.

ACKNOWLEDGMENT

This work has been supported by the 210124 ARC DECR A grant DE180100768 and the NSF grant 1526875.

REFERENCES

[1] H. Dau and O. Milenkovic, “Optimal repair schemes for some families of Reed-Solomon codes,” in Proc. IEEE Int. Symp. Inform. Theory (ISIT), 2017, pp. 346–350.
[2] A. Dimakis, P. Godfrey, M. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” in Proc. 26th IEEE Int. Conf. Comput. Commun. (INFOCOM), 2007, pp. 2000–2008.
[3] A. Dimakis, P. Godfrey, Y. Wu, M. Wainwright, and K. Ramchandran, “Network coding for distributed storage systems,” IEEE Trans. Inform. Theory, vol. 56, no. 9, pp. 4539–4551, 2010.
[4] I. S. Reed and G. Solomon, “Polynomial codes over certain finite fields,” J. Soc. Ind. Appl. Math., vol. 8, no. 2, pp. 300–304, 1960.
[5] F. J. MacWilliams and N. J. A. Sloane, The Theory of Error-Correcting Codes, 2nd ed. Amsterdam: North-Holland, 1977.
[6] H. Dau, I. Duursma, H. M. Kiah, and O. Milenkovic, “Repairing Reed-Solomon codes with multiple erasures,” IEEE Trans. Inform. Theory, vol. 54, no. 10, pp. 6567–6582, 2008.
[7] S. Muralidhar et al., “F4: Facebook’s warm BLOB storage system,” in Proc. 11th ACM/USENIX Symp. Oper. Syst. Des. Implementation (OSDI), 2014, pp. 383–398.
[8] M. Safinamoothy, M. Astoris, D. Papailiopoulos, A. G. Dimakis, R. Vahalid, S. Chen, and D. Borthakur, “XORing elephants: Novel erasure codes for Big Data,” Proc. VLDB Endow., vol. 6, no. 5, pp. 325–336, 2013.
[9] K. Shannugam, D. S. Papailiopoulos, A. G. Dimakis, and G. Caire, “A repair framework for scalar MDS codes,” IEEE J. Selected Areas Comm. (JSAC), vol. 32, no. 5, pp. 998–1007, 2014.
[10] V. Guruswami and M. Wootters, “Reparing Reed-Solomon codes,” in Proc. Annu. Symp. Theory Comput. (STOC), 2016.
[11] ——, “Reparing Reed-Solomon codes,” IEEE Trans. Inform. Theory, vol. 63, no. 9, pp. 5684–5698, 2017.
[12] M. Ye and A. Barg, “Explicit constructions of MDS array codes and RS codes with optimal repair bandwidth,” in Proc. IEEE Int. Symp. Inform. Theory (ISIT), 2016, pp. 1202–1206.
[13] ——, “Explicit constructions of high-rate MDS array codes with optimal repair bandwidth,” IEEE Trans. Inform. Theory, vol. 63, no. 4, pp. 2001–2014, 2017.
[14] I. Duursma and H. Dau, “Low bandwidth repair of the RS(10,4) Reed-Solomon code,” in Proc. Inform. Theory Appl. Workshop (ITA), 2017.
[15] I. Tamo, M. Ye, and A. Barg, “Optimal repair of Reed-Solomon codes: Achieving the cut-set bound,” in Proc. 58th Annual IEEE Symp. Foundations Computer Sci. (FOCS), 2017.
[16] ——, “The repair problem for Reed-Solomon codes: Optimal repair of single and multiple erasures with almost optimal node size,” IEEE Trans. Inform. Theory, vol. 65, no. 5, pp. 2673–2695, 2018.
[17] H. Dau, I. Duursma, H. M. Kiah, and O. Milenkovic, “Repairing Reed-Solomon codes with two erasures,” in Proc. IEEE Int. Symp. Inform. Theory (ISIT), 2017, pp. 351–355.
[18] B. Barton and M. Wootters, “Repairing multiple failures for scalar MDS codes,” in Proc. 55th Annual Allerton Conf. Comm Control Comput. (Allerton), 2017.
[19] J. Mardia, B. Barton, and M. Wootters, “Repairing multiple failures for scalar MDS codes,” IEEE Trans. Inform. Theory, vol. 65, no. 5, pp. 2661–2672, 2018.
[20] Y. Zhang and Z. Zhang, “An improved cooperative repair scheme for Reed-Solomon codes,” in Proc. 19th Int. Symp. Comm. Inform. Tech. (ISCIT), 2019, pp. 525–530.

[21] A. Chowdhury and A. Vardy, “Improved schemes for asymptotically optimal repair of MDS codes,” in Proc. 55th Annual Allerton Conf. Comm Control Comput. (Allerton), 2017.

[22] W. Li, Z. Wang, and H. Jafarkhani, “A tradeoff between the sub-packetization size and the repair bandwidth for Reed-Solomon code,” in Proc. 55th Annual Allerton Conf. Comm. Control Comput. (Allerton), 2017, pp. 942–949.

[23] ——, “On the sub-packetization size and the repair bandwidth of Reed-Solomon codes,” IEEE Trans. Inform. Theory, vol. 65, no. 9, pp. 5484–5502, 2019.

[24] H. Dau, I. Duursma, and H. Chu, “On the I/O costs of some repair schemes for full-length Reed-Solomon codes,” in Proc. IEEE Int. Symp. Inform. Theory (ISIT), 2018, pp. 1700–1704.

[25] H. Dau and E. Viterbo, “Repair schemes with optimal I/O costs for full-length Reed-Solomon codes with two parities,” in Proc. IEEE Inform. Theory Workshop (ITW), 2018, pp. 590–594.

[26] W. Li, H. Dau, Z. Wang, H. Jafarkhani, and E. Viterbo, “On the I/O costs in repairing short-length Reed-Solomon codes,” in Proc. IEEE Int. Symp. Inform. Theory (ISIT), 2019, pp. 1087–1091.

[27] R. Lidl and H. Niederreiter, Introduction to Finite Fields and Their Applications. Cambridge University Press, 1986.

[28] D. Goss, Basic Structures of Function Field Arithmetics. Springer-Verlag Berlin Heidelberg, 1996.

[29] L. Reis, “Nilpotent linearized polynomials over finite fields and applications,” Finite Fields and Their Applications, vol. 50, pp. 279–292, 2018.

[30] S. H. Friedberg, A. J. Insel, and L. E. Spence, Linear Algebra. Pearson Education, 2014.