A Rate-Distortion Perspective on Multiple Decoding Attempts for Reed-Solomon Codes

Phong S. Nguyen, Henry D. Pfister, and Krishna R. Narayanan

Department of Electrical and Computer Engineering, Texas A&M University

College Station, TX 77840, U.S.A.

{psn, hpfister, krn}@tamu.edu

Abstract—Recently, a number of authors have proposed decoding schemes for Reed-Solomon (RS) codes based on multiple trials of a simple RS decoding algorithm. In this paper, we present a rate-distortion (R-D) approach to analyze these multiple-decoding algorithms for RS codes. This approach is first used to understand the asymptotic performance-versus-complexity trade-off of multiple error-and-erasure decoding of RS codes. By defining an appropriate distortion measure between an error pattern and an erasure pattern, the condition for a single error-and-erasure decoding to succeed reduces to a form where the distortion is compared to a fixed threshold. Finding the best set of erasure patterns for multiple decoding trials then turns out to be a covering problem which can be solved asymptotically by rate-distortion theory. Next, this approach is extended to analyze multiple algebraic soft-decision (ASD) decoding of RS codes. Both analytical and numerical computations of the R-D functions for the corresponding distortion measures are discussed. Simulation results show that proposed algorithms using this approach perform better than other algorithms with the same complexity.

I. INTRODUCTION

Reed-Solomon (RS) codes are one of the most widely used error-correcting codes in digital communication and data storage systems. This is primarily due to the fact that RS codes are maximum distance separable (MDS) codes, can correct long bursts of errors, and have efficient hard-decision decoding (HDD) algorithms, such as the Berlekamp-Massey (BM) algorithm, which can correct up to half the minimum distance (d_{min}) of the code. An (n,k) RS code of length n and dimension k is known to have $d_{min} = n - k + 1$ due to its MDS nature.

Since the arrival of RS codes, people have put a considerable effort into improving the decoding performance at the expense of complexity. A breakthrough result of Guruswami and Sudan (GS) introduces a hard-decision list-decoding algorithm based on algebraic bivariate interpolation and factorization techniques that can correct errors beyond half the minimum distance of the code [1]. Nevertheless, HDD algorithms do not fully exploit the information provided by the channel output. Koetter and Vardy (KV) later

extended the GS decoder to an algebraic soft-decision (ASD) decoding algorithm by converting the probabilities observed at the channel output into algebraic interpolation conditions in terms of a multiplicity matrix [2]. Both of these algorithms however have significant computational complexity. Thus, multiple runs of error-and-erasure and error-only decoding with some low complexity algorithm, such as the BM algorithm, has renewed the interest of researchers. These algorithms essentially first construct a set of either erasure patterns [3], [4], test patterns [5], or patterns combining both [6] and then attempt to decode using each pattern. There has also been interest in lowering the complexity per decoding trial as can be seen in [7], [8], [9].

In the scope of multiple error-and-erasure decoding, there have been several algorithms using different sets of erasure patterns. After multiple decoding trials, these algorithms produce a list of candidate codewords, whose size is usually small, and then pick the best codeword on this list. The nature of multiple error-and-erasure decoding is to erase some of the least reliable symbols since those symbols are more prone to be erroneous. The first algorithm of this type is called Generalized Minimum Distance (GMD) [3] and it repeats error-and-erasure decoding while successively erasing an even number of the least reliable positions (LRPs) (assuming that d_{min} is odd). More recent work by Lee and Kumar [4] proposes a soft-information successive (multiple) error-and-erasure decoding (SED) that achieves better performance but also increases the number of decoding attempts. Literally, the Lee-Kumar's SED(l,f) algorithm runs multiple error-and-erasure decoding trials with every combination of an even number $\leq f$ of erasures within the l LRPs.

A natural question that arises is how to construct the "best" set of erasure patterns for multiple error-and-erasure decoding. Inspired by this, we first design a rate-distortion framework to analyze the asymptotic trade-off between performance and complexity of multiple error-and-erasure decoding of RS codes. The framework is also extended to analyze multiple algebraic soft-decision (ASD) decoding. Next, we propose a family of multiple-decoding algorithms based on this approach that achieve better performance-versus-complexity trade-off than other algorithms. The multiple-decoding algorithm that achieves the best trade-off turns out to be a multiple error-only decoding using the set of patterns generated by random codes combined with covering codes.

A. Outline of the paper

The paper is organized as follows. In Section II, we design an appropriate distortion measure and present a rate-distortion framework to analyze the performance-versus-complexity trade-off of multiple error-and-erasure decoding of RS codes. Also in this section, we propose a general multiple-decoding algorithm that can be applied to error-and-erasure decoding. Then, in Section III, we discuss a numerical computation of R-D function which is needed for the proposed algorithm. In Section IV, we analyze both bit-level and symbol-level ASD decoding and design distortion measures that work for the general algorithm. In Section V, we offer some extensions that help the algorithm achieve better performance and running time. Simulation results are presented in Section VI and finally, conclusions are provided in Section VII.

II. MULTIPLE ERROR-AND-ERASURE DECODING

In this section, we set up a rate-distortion framework to analyze multiple attempts of conventional hard decision error-and-erasure decoding.

Let $\mathbb{F}_{2^q}$ be the Galois field with 2^q elements denoted as $\alpha_1, \alpha_2, \ldots, \alpha_{2^q}$. We consider an (n,k) RS code of length n, dimension k over $\mathbb{F}_{2^q}$. Assume that we transmit a codeword $\mathbf{c} = (c_1, c_2, \ldots, c_n) \in \mathbb{F}_{2^q}^n$ over some channel and receive a vector $\mathbf{r} = (r_1, r_2, \ldots, r_n) \in \mathcal{Y}^n$ where $\mathcal{Y}$ is the receive alphabet for a single RS symbol. In this paper, we assume that $\mathcal{Y} = \mathbb{R}^q$ and all simulations are based on transmitting each of the q bits in a symbol using Binary Phase-Shift Keying (BPSK) on an Additive White Gaussian Noise (AWGN) channel. For each codeword index i, let $\pi_i : \{1, 2, \ldots, 2^q\} \to \{1, 2, \ldots, 2^q\}$ be the permutation given by sorting $p_{i,j} = \Pr(c_i = \alpha_j | \mathbf{r})$ in decreasing order so that $p_{i,\pi_i(1)} \geq p_{i,\pi_i(2)} \geq \ldots \geq p_{i,\pi_i(2^q)}$. Then, we can specify $y_{i,j} = \alpha_{\pi_i(j)}$ as the j-th most reliable symbol for $j = 1, \ldots, 2^q$ at codeword index i. To obtain the reliability of the codeword positions (indices), we construct the permutation $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ given by sorting the probabilities $p_{i,\pi_i(1)}$ of the most likely symbols in increasing order. Thus, codeword position $\sigma(i)$ is the i-th LRP. These above notations will be used throughout this paper.

Example 1: Consider $n = 3$ and $q = 2$. Assume that we have the probability $p_{i,j}$ written in a matrix form as follows

$$\mathbf{P} = \begin{pmatrix} 0.01 & 0.01 & \mathbf{0.93} \\ \mathbf{0.94} & 0.03 & 0.04 \\ 0.03 & \mathbf{0.49} & 0.01 \\ 0.02 & 0.47 & 0.02 \end{pmatrix} \text{ where } p_{i,j} = [\mathbf{P}]_{j,i}$$

then $\pi_1(1,2,3,4) = (2,3,4,1)$, $\pi_2(1,2,3,4) = (3,4,2,1)$, $\pi(1,2,3,4) = (1,2,4,3)$ and $\sigma(1,2,3) = (2,3,1)$.

Condition 1: (Classical decoding threshold, see [10], [11]): If e symbols are erased, a conventional hard-decision error-and-erasure decoder such as the BM algorithm is able to correct v errors in unerased positions if

$$2v + e < n - k + 1. \tag{1}$$

A. Conventional error and erasure patterns.

Definition 1: (Conventional error and erasure patterns) We define $x^n \in \mathbb{Z}_2^n \triangleq \{0,1\}^n$ and $\hat{x}^n \in \mathbb{Z}_2^n$ as an error pattern and an erasure pattern respectively, where $x_i = 0$ means that an error occurs (i.e. the most likely symbol is incorrect) and $\hat{x}_i = 0$ means that an erasure occurs at index i.

Example 2: $\{111111\ldots, 001111\ldots, 000011\ldots, \ldots\}$ is the set of erasure patterns for the GMD algorithm assuming d_{min} is odd. For the SED(3,2) algorithm, the set of erasure patterns has the form $\{111111\ldots, 001111\ldots, 010111\ldots, 100111\ldots\}$. Here, in each erasure pattern the letters are written in increasing reliability order of the codeword positions.

Let us revisit the question how to construct the best set of erasure patterns for multiple error-and-erasure decoding. First, it can be seen that a multiple error-and-erasure decoding succeeds if the condition (1) is satisfied during at least one round of decoding. Thus, our approach is to design a distortion measure that converts the condition (1) into a form where the distortion between an error pattern x^n and an erasure pattern $\hat{x}^n$, denoted as $d(x^n, \hat{x}^n)$, is less than a fixed threshold.

Definition 2: Given a *letter-by-letter* distortion measure δ, the distortion between an error pattern x^n and an erasure pattern $\hat{x}^n$ is defined by

$$d(x^n, \hat{x}^n) = \sum_{i=1}^n \delta(x_i, \hat{x}_i).$$

Proposition 1: If we choose the *letter-by-letter* distortion measure $\delta : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{R}_{\geq 0}$ as follows

$$\begin{array}{ll} \delta(0,0) = 1 & \delta(0,1) = 2 \\ \delta(1,0) = 1 & \delta(1,1) = 0 \end{array} \tag{2}$$

then the condition (1) for a successful error-and-erasure decoding reduces to the form where the distortion is less than a fixed threshold

$$d(x^n, \hat{x}^n) < n - k + 1.$$

Proof: First, we define $\chi_{s,t} \triangleq |\{i \in \{1, 2, \ldots, n\} : x_i = s, \hat{x}_i = t\}|$ to count the number of $(x_i, \hat{x}_i)$ pairs equal to (s,t) for every $s, t \in \{0, 1\}$. Noticing that $e = \chi_{0,0} + \chi_{1,0}$ and $v = \chi_{0,1}$, the condition (1) for one error-and-erasure decoding attempt to succeed becomes $2\chi_{0,1} + \chi_{0,0} + \chi_{1,0} < n - k + 1$. By seeing that $d(x^n, \hat{x}^n) = 2\chi_{0,1} + \chi_{0,0} + \chi_{1,0}$ we conclude the proof. ∎

Next, we try to maximize the chance that this successful decoding condition is satisfied by at least one of the decoding attempts (i.e. $d(x^n, \hat{x}^n) < n - k + 1$ for at least one erasure patterns $\hat{x}^n$). Mathematically, we want to build a set $\mathcal{B}$ of no more than 2^R erasure patterns $\hat{x}^n$ in order to

$$\max_{\mathcal{B}: |\mathcal{B}| \leq 2^R} \Pr\{\min_{\hat{x}^n \in \mathcal{B}} d(x^n, \hat{x}^n) < n - k + 1\}. \tag{3}$$

The exact answer to this problem is difficult to find. However, one can see it as a covering problem where one wants to cover the space of error patterns using a minimum number of balls centered at the chosen erasure patterns.

Error pattern □
Erasure pattern •
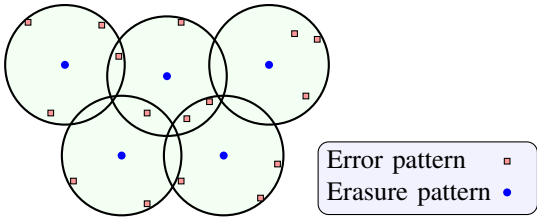
This view leads to an asymptotic solution of the problem based on rate-distortion theory. More precisely, we view the error pattern x^n as a source sequence and the erasure pattern $\hat{x}^n$ as a reproduction sequence. Rate-distortion theory shows that the set $\mathcal{B}$ of 2^R reproduction sequences can be generated randomly so that

$$\lim_{n \to \infty} E\left[\min_{\hat{x}^n \in \mathcal{B}} d(x^n, \hat{x}^n)\right] \leq D$$

where the distortion D is minimized for a given rate R. Thus, for large enough n, we have

$$\min_{\hat{x}^n \in \mathcal{B}} d(x^n, \hat{x}^n) \leq D$$

with high probability. Here, R and D are closely related to the complexity and the performance, respectively, of the decoding algorithm. Therefore, we characterize the trade-off between those two aspects using the relationship between R and D.

B. Generalized error and erasure patterns

In this subsection, we consider a generalization of the conventional error and erasure patterns under the same framework to make better use of the soft information. At each index of the RS codeword, beside erasing the symbol we can try to decode using not only the most likely symbol but also other ones as the hard decision (HD) symbol. To handle up to the l most likely symbols at each index i, we let $\mathbb{Z}_{l+1} \triangleq \{0, 1, \ldots, l\}$ and consider the following definition.

Definition 3: (Generalized error patterns and erasure patterns) Consider a positive integer $l < 2^q$. Let us define $x^n \in \mathbb{Z}_{l+1}^n$ as the generalized error pattern where, at index i, $x_i = j$ implies that the j-th most likely symbol is correct for $j \in \{1, 2, \ldots l\}$, and $x_i = 0$ implies none of the first l most likely symbols is correct. Let $\hat{x}^n \in \mathbb{Z}_{l+1}^n$ be the generalized erasure pattern used for decoding where, at index i, $\hat{x}_i = j$ implies that the j-th most likely symbol is used as the hard-decision symbol for $j \in \{1, 2, \ldots, l\}$, and $\hat{x}_i = 0$ implies that an erasure is used at that index.

For simplicity, we will refer to x^n as the error pattern and $\hat{x}^n$ as the erasure pattern like in the conventional case. Next, we also want to convert the condition (1) to the form where $d(x^n, \hat{x}^n)$ is less than a fixed threshold. Proposition 1 is thereby generalized into the following theorem.

Theorem 1: We choose the *letter-by-letter* distortion measure $\delta : \mathbb{Z}_{l+1} \times \mathbb{Z}_{l+1} \to \mathbb{R}_{\geq 0}$ defined by $\delta(x, \hat{x}) = [\Delta]_{x, \hat{x}}$ in terms of the $(l+1) \times (l+1)$ matrix

$$\Delta = \begin{pmatrix} 1 & 2 & \ldots & 2 & 2 \\ 1 & 0 & \ldots & 2 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & \ldots & 0 & 2 \\ 1 & 2 & \ldots & 2 & 0 \end{pmatrix}.$$

Using this, the condition (1) for a successful error-and-erasure decoding becomes

$$d(x^n, \hat{x}^n) < n - k + 1.$$

Proof: The reasoning is similar to Proposition 1 using the fact that $e = \sum_{s=0}^{l} \chi_{s,0}$ and $v = \sum_{t=1}^{l} \sum_{s=0, s \neq t}^{l} \chi_{s,t}$ where $\chi_{s,t} \triangleq |\{i \in \{1, 2, \ldots, n\} : x_i = s, \hat{x}_i = t|$ for every $s, t \in \mathbb{Z}_{l+1}$. ∎

For each $l = 1, 2, \ldots, 2^q$, we will refer to this generalized case as mBM-l decoding.

Example 3: We consider the case mBM-2 decoding where $l = 2$. The distortion measure is given by following the matrix

$$\Delta = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 1 & 2 & 2 \end{pmatrix}.$$

Here, at each codeword position, we consider the first and second most likely symbols as the two hard-decision choices. This is similar to the Chase-type decoding method proposed by Bellorado and Kavcic [7].

C. Proposed General Multiple-Decoding Algorithm

In this section, we propose a general multiple-decoding algorithm for RS codes based on the rate-distortion approach. This general algorithm applies to not only multiple error-and-erasure decoding but also multiple-decoding of other decoding schemes that we will discuss later. The first step is designing a distortion measure that converts the condition for a single decoding to succeed to the form where distortion is less than a fixed threshold. After that, decoding proceeds as described below.

• *Phase I: Compute rate-distortion function.*

Step 1: Transmit τ (say $\tau = 1000$) arbitrary test RS codewords, indexed by time $t = 1, 2, \ldots, \tau$, over the channel and compute a set of τ $2^q \times n$ matrices $\mathbf{P}_1^{(t)}$ where $[\mathbf{P}_1^{(t)}]_{j,i} = p_{i, \pi_i^{(t)}(j)}^{(t)}$ is the probability of the j-th most likely symbol at position i during time t.

Step 2: For each time t, obtain the matrix $\mathbf{P}_2^{(t)}$ from $\mathbf{P}_1^{(t)}$ through a permutation $\sigma^{(t)} : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ that sorts the probabilities $p_{i, \pi_i^{(t)}(1)}^{(t)}$ in increasing order to indicate the reliability order of codeword positions. Take the entry-wise average of all τ matrices $\mathbf{P}_2^{(t)}$ to get an average matrix $\bar{\mathbf{P}}$.

Step 3: Compute the R-D function of a source sequence (error pattern) with probability of source letters derived from $\bar{\mathbf{P}}$ and the designed distortion measure (see Section III and Section V-B) . Determine the point on the R-D curve

that corresponds to a designated rate R along with the test-channel input-probability distribution vector $\underline{q}$ that achieves that point.

- *Phase II: Run actual decoder.*

Step 4: Based on the actual received signal sequence, compute $p_{i,\pi_i(1)}$ and determine the permutation σ that gives the reliability order of codeword positions by sorting $p_{i,\pi_i(1)}$ in increasing order.

Step 5: Randomly generate a set of 2^R erasure patterns using the test-channel input-probability distribution vector $\underline{q}$ and permute the indices of each erasure pattern by the permutation σ^{-1}.

Step 6: Run multiple attempts of the corresponding decoding scheme (e.g. error-and erasure decoding) using the set of erasure patterns in Step 5 to produce a list of candidate codewords.

Step 7: Use Maximum-Likelihood (ML) decoding to pick the best codeword on the list.

III. COMPUTING THE RATE-DISTORTION FUNCTION

In this section, we will present a numerical method to compute the R-D function and the test-channel input-probability distribution that achieves a specific point in the R-D curve. This probability distribution will be needed to randomly generate the set of erasure patterns in the general multiple-decoding algorithm that we have proposed.

For an arbitrary discrete distortion measure, it can be difficult to compute the R-D function analytically. Fortunately, the Blahut-Arimoto (B-A) algorithm (see details in [12], [13]) provides a technique that efficiently computes the R-D function of a single discrete source. More precisely, given a parameter $s < 0$ which represents the slope of the $R - D$ curve at a specific point and an arbitrary all-positive initial test-channel input-probability distribution vector $\underline{q}^{(0)}$, the B-A algorithm shows us how to compute the rate-distortion point (R_s, D_s) by means of computing the test-channel input-probability distribution vector $\underline{q}^\star = \lim_{t \to \infty} \underline{q}^{(t)}$ and the test-channel transition probability matrix $Q^\star = \lim_{t \to \infty} Q^{(t)}$ that achieves that point.

However, it is not straightforward to apply the B-A algorithm to compute the R-D for a discrete source sequence x^n (an error pattern in our context) of n independent but non identical source components x_i. In order to do that, we consider the group of source letters $(j_1, j_2, \ldots, j_n)$ where $j_i \in \mathcal{J}$ as a super-source letter $J \in \mathcal{J}^n$, the group of reproduction letters $(k_1, k_2, \ldots, k_n)$ where $k_i \in \mathcal{K}$ as a super-reproduction letter $K \in \mathcal{K}^n$, and the source sequence x^n as a single source. For each super-source letter J, $p_J = \Pr(x^n = J) = \prod_{i=1}^n \Pr(x_i = j_i) = \prod_{i=1}^n p_{j_i}$ follows from the independence of source components. While we could apply the B-A algorithm to this source directly, the complexity is a problem because the alphabet sizes for J and K become the super-alphabet sizes $|\mathcal{J}|^n$ and $|\mathcal{K}|^n$ respectively. Instead, we avoid this computational challenge by choosing the initial test-channel input-probability distribution so that it can be factored into

a product of n initial test-channel input-probability components, i.e. $q_K^{(0)} = \prod_{i=1} q_{k_i}^{(0)}$. Then, we see that this factorization rule still applies after every step of the iterative process. By doing this, for each parameter s we only need to compute the rate-distortion pair for each component (or index i) separately and sum them together. This is captured into the following theorem.

Theorem 2: (Factored Blahut-Arimoto algorithm) Consider a discrete source sequence x^n of n independent but non identical source components x_i. Given a parameter $s < 0$, the rate and the distortion for this source sequence are given by

$$R_s = \sum_{i=1}^n R_{i,s} \text{ and } D_s = \sum_{i=1}^n D_{i,s}$$

where the components $R_{i,s}$ and $D_{i,s}$ are computed by the B-A algorithm with the parameter s. This pair of rate and distortion can be achieved by the corresponding test-channel input-probability distribution $q_K \triangleq \Pr(\hat{x}^n = K) = \prod_{i=1}^n q_{k_i}$ where the component probability distribution $q_{k_i} \triangleq \Pr(\hat{x}_i = k_i)$.

Proof: See [14]. ∎

IV. MULTIPLE ALGEBRAIC SOFT DECISION (ASD) DECODING

In this section, we analyze and design a distortion measure to convert the condition for successful ASD decoding to a suitable form so that we can apply the general multiple-decoding algorithm to ASD decoding.

First, let us give a brief review on ASD decoding of RS codes. Considering a received vector $\mathbf{r}$, we can compute the *a posteriori* probability (APP) matrix $\mathbf{P}$ as follows.

$$[\mathbf{P}]_{j,i} = p_{i,j} = \Pr(c_i = \alpha_j | \mathbf{r}) \text{ for } 1 \le i \le n, 1 \le j \le 2^q.$$

The ASD decoding as in [2] then uses a particular multiplicity assignment scheme (MAS) to derive a $2^q \times n$ multiplicity matrix, denoted as $\mathbf{M}$, of non-negative integer entries $\{m_{i,j}\}$ from the APP matrix $\mathbf{P}$.

In this paper, we denote $m = \max_{i,j} m_{i,j}$ as the maximum multiplicity. Intuitively, higher multiplicity should be put on more likely symbols. Increasing m generally gives rise to the performance of ASD decoding. However, one of the drawbacks of ASD decoding is that its decoding complexity is roughly $\mathcal{O}(m^6)$ which sharply increases with m. Thus, in this section we will work with small m to keep the complexity affordable.

One of the main contributions of [2] is to offer a condition for successful ASD decoding represented in terms of two quantities specified as the score and the cost as follows.

Definition 4: The score $S_{\mathbf{M}}(\mathbf{c})$ with respect to a codeword $\mathbf{c}$ and a multiplicity matrix $\mathbf{M}$ is defined as $S_{\mathbf{M}}(\mathbf{c}) = \sum_{j=1}^n m_{[c_j],j}$ where $[c_j] = i$ such that $\alpha_i = c_j$. The cost $C_{\mathbf{M}}$ of a multiplicity matrix $\mathbf{M}$ is defined as $C_{\mathbf{M}} = \frac{1}{2} \sum_{i=1}^q \sum_{j=1}^n m_{i,j}(m_{i,j} + 1)$

Condition 2: (ASD decoding threshold, see [2], [15], [16]). The transmitted codeword will be on the list if

$$T(S_{\mathbf{M}}) > C_{\mathbf{M}} \quad \text{where} \quad T(S_{\mathbf{M}}) = (a+1)\left[S_{\mathbf{M}} - \frac{a}{2}(k-1)\right]$$

for any $a \in \mathbb{N}$ such that $a(k-1) < S_{\mathbf{M}} \leq (a+1)(k-1)$. (4)

To match the general framework, the ASD decoding threshold (or condition for successful ASD decoding) should be converted to the form where the distortion is smaller than a fixed threshold.

A. Bit-level ASD case

In this subsection, we consider multiple trials of ASD decoding using bit-level erasure patterns. A bit-level error pattern $b^N \in \mathbb{Z}_2^N$ and a bit-level erasure pattern $\hat{b}^N \in \mathbb{Z}_2^N$ has length $N = n \times q$ since each symbol has q bits. Similar to Definition 1 of a conventional error pattern and a conventional erasure pattern, $b_i = 0$ in a bit-level error pattern implies a bit-level error occurs and $\hat{b}_i$ in a bit-level erasure pattern implies that a bit-level erasure occurs.

From each bit-level erasure pattern we can specify entries of the multiplicity matrix $\mathbf{M}$ using the bit-level MAS proposed in [15] as follows: for each codeword position, assign multiplicity 2 to the symbol with no bit erased, assign multiplicity 1 to each of the two candidate symbols if there is 1 bit erased, and assign multiplicity zero to all the symbols if there are ≥ 2 bits erased. All the other entries are zeros by default. This MAS has a larger decoding region compared to the conventional error-and-erasure decoding scheme.

Condition 3: (Bit-level ASD decoding threshold, see [15]) For RS codes of rate $\frac{k}{n} \geq \frac{2}{3} + \frac{1}{n}$, ASD decoding using the proposed bit-level MAS will succeed (i.e. the transmitted codeword is on the list) if

$$3v_b + e_b < \frac{3}{2}(n-k+1) \quad (5)$$

where e_b is the number of bit-level erasures and v_b is the number of bit-level errors in unerased locations.

We can choose an appropriate distortion measure according to the following proposition which is a natural extension of Proposition 1 in the symbol level.

Proposition 2: If we choose the bit-level *letter-by-letter* distortion measure $\delta : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{R}_{\geq 0}$ as follows

$$\begin{matrix} \delta(0,0)=1 & \delta(0,1)=3 \\ \delta(1,0)=1 & \delta(1,1)=0 \end{matrix} \quad (6)$$

then the condition (5) becomes

$$d(b^N, \hat{b}^N) < \frac{3}{2}(n-k+1). \quad (7)$$

Proof: The proof uses the same reasoning as the proof of Proposition 1. ∎

Remark 1: We refer the the multiple-decoding of bit-level ASD as m-b-ASD.

B. Symbol-level ASD case

In this subsection, we try to convert the condition for successful ASD decoding in general to the form that suits our goal. We will also determine which multiplicity assignment schemes allow us to do so.

Definition 5: (Multiplicity type) For some codeword position, let us assign multiplicity m_j to the j-th most likely

symbol for $j = 1, 2, \ldots, l$ where $l \leq 2^q$. The remaining entries in the column are zeros by default. We call the sequence, $(m_1, m_2, \ldots, m_l)$, the column multiplicity type for "top-l" decoding.

First, we notice that a choice of multiplicity types in ASD decoding at each codeword position has the similar meaning to a choice of erasure decisions in the conventional error-and-erasure decoding. However, in ASD decoding we are more flexible and may have more types of erasures. For example, assigning multiplicity zero to all the symbols (all-zero multiplicity type) at codeword position i corresponds to the case when we have a complete erasure at that position. Assigning the maximum multiplicity m to one symbol corresponds to the case when we choose that symbol as the hard-decision one. Hence with some abuse of terminology, we also use the term (generalized) erasure pattern $\hat{x}^n$ for the multiplicity assignment scheme in the ASD context. Each erasure-letter x_i gives the multiplicity type for the corresponding column of the multiplicity matrix $\mathbf{M}$.

Definition 6: (Error and erasure patterns for ASD decoding) Consider a MAS with z multiplicity types. Let $\hat{x}^n \in \{1, 2 \ldots, z\}^n$ be an erasure pattern where, at index i, $x_i = j$ implies that multiplicity type j is used at column i of the multiplicity matrix $\mathbf{M}$. Notice that the definition of an error pattern $x^n \in \mathbb{Z}_{l+1}^n$ in Definition 3 applies unchanged here.

Rate-distortion theory gives us the intuition that in general the more multiplicity types (erasure choices) we have, the better performance of multiple ASD decoding we achieve as n becomes large. Thus, we want to find as many as possible multiplicity types for "top-l" that allow us to convert condition for successful ASD decoding to the correct form.

Example 4: Choosing $m = 2$, for example, gives four column multiplicity types for "top-2" decoding as follows: the first is $(2,0)$ where we assign multiplicity 2 to the most likely symbol $y_{i,1}$, the second is $(1,1)$ where we assign equal multiplicity 1 to the first and second most likely symbols $y_{i,1}$ and $y_{i,2}$, the third is $(0,2)$ where we assign multiplicity 2 to the second most likely symbol $y_{i,2}$, and the fourth is $(0,0)$ where we assign multiplicity zero to all the symbols at index i (i.e. the i-th column of $\mathbf{M}$ is an all-zero column). As a corollary of Theorem 3 below, the distortion matrix that converts (4) to the correct form for this case is

$$\Delta = \begin{pmatrix} 2 & 5/3 & 2 & 1 \\ 0 & 2/3 & 2 & 1 \\ 2 & 5/3 & 0 & 1 \end{pmatrix}.$$

The following definition and theorem provide a set of allowable multiplicity types that converts the condition for successful ASD decoding into the form where distortion is less than a fixed threshold.

Definition 7: The set of allowable multiplicity types for "top-l" decoding with maximum multiplicity m is defined to be[1]

$$\mathcal{A}(m,l) \triangleq \left\{ (m_1, m_2, \ldots, m_l) : \sum_{r=1}^{l} m_r \leq m \text{ and} \right.$$

---

[1] We use the convention that $\min_{r:m_r \neq 0} m_r = 0$ if $\{r : m_r \neq 0\} = \emptyset$.

$$\sum_{r=1}^{l} m_r(m-m_r) \le (m+1)\left(|\{r : m_r \ne 0\}| - 1\right) \min_{r:m_r \ne 0} m_r. \quad (8)$$

Taking the elements of this set in an arbitrary order, we let the j-th multiplicity type in the allowable set be $(m_{j,1}, m_{j,2}, \dots m_{j,l})$.

Example 5: $\mathcal{A}(3,2)$ consists of all permutations of $(3,0), (2,1), (1,1), (0,0)$. Meanwhile, $\mathcal{A}(2,2)$ comprises all the permutations of $(2,0), (1,1), (0,0)$ and we refer to the multiple ASD decoding algorithm using this set of multiplicity types as mASD-2. $\mathcal{A}(3,3)$ consists of all the permutations of $(3,0,0), (0,0,0), (1,1,0), (2,1,0), (1,1,1)$ and this case is referred as mASD-3. We also consider another case called mASD-2a that uses the set of multiplicity types $\{(2,0), (1,1), (0,0)\}$.

Theorem 3: Let $z = |\mathcal{A}(m,l)|$ be the number of multiplicity types in a MAS for "top-l" decoding with maximum multiplicity m. Let $\delta : \mathbb{Z}_{l+1} \times \mathbb{Z}_{z+1} \setminus \{0\} \to \mathbb{R}_{\ge 0}$ be a *letter-by-letter* distortion measure defined by $\delta(x,\hat{x}) = [\Delta]_{x,\hat{x}}$, where Δ is the $(l+1) \times z$ matrix

$$\Delta = \begin{pmatrix} \mu_1 & \mu_2 & \dots & \mu_z \\ \mu_1 - 2m_{1,1}/m & \mu_2 - 2m_{2,1}/m & \dots & \mu_z - 2m_{z,1}/m \\ \mu_1 - 2m_{1,2}/m & \mu_2 - 2m_{2,2}/m & \dots & \mu_z - 2m_{z,2}/m \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1 - 2m_{1,l}/m & \mu_2 - 2m_{2,l}/m & \dots & \mu_z - 2m_{z,l}/m \end{pmatrix}.$$

with $\mu_t = 1 + \sum_{r=1}^{l} \frac{m_{t,r}(m_{t,r}+1)}{m(m+1)}$. Then, the condition (4) for successful ASD decoding of a RS code with rate $\frac{k}{n} \ge \frac{1}{n} + \frac{m(m+3)}{(m+1)(m+2)}$ is equivalent to

$$d(x^n, \hat{x}^n) < n - k + 1. \quad (9)$$

Proof: See [14]. ∎

V. Some Extensions and Generalizations

A. Erasure patterns using covering codes

The R-D framework we use is most suitable when $n \to \infty$. For a finite n, the random coding approach may have problems with only a few LRPs. We can instead use good covering codes to handle these LRPs. In the scope of covering problems, one can use an l-ary t_c-covering code (e.g. a perfect Hamming or Golay code) with covering radius t_c to cover the whole space of l-ary vectors of the same length. The covering may still work well if the distortion measure is close to, but not exactly equal to the Hamming distortion.

In order take care of up to the l most likely symbols at each of the n_p LRPs of an (n,k) RS, we consider an (n_c, k_c) l-ary t_c-covering code whose codeword alphabet is $\mathbb{Z}_{l+1} \setminus \{0\} = \{1,2,\dots,l\}$. Then, we give a definition of the (generalized) error patterns and erasure patterns for this case. In order to draw similarities between this case and the previous cases, we still use the terminology "generalized erasure pattern" and shorten it to erasure pattern even if error-only decoding is used. For error-only decoding, Condition 1 for successful decoding becomes

$$v < \frac{1}{2}(n - k + 1).$$

Definition 8: (Error and erasure patterns for error-only decoding) Let us define $x^n \in \mathbb{Z}_{l+1}^n = \{0,1,\dots,l\}^n$ as an error pattern where, at index i, $x_i = j$ implies that the j-th most likely symbol is correct for $j \in \{1,2,\dots l\}$, and $x_i = 0$ implies none of the first l most likely symbols is correct. Let $\hat{x}^n \in \{1,2,\dots,l\}^n$ be an erasure pattern where, at index i, $\hat{x}_i = j$ implies that the j-th most likely symbol is chosen as the hard-decision symbol for $j \in \{1,2,\dots,l\}$.

Proposition 3: If we choose the *letter-by-letter* distortion measure $\delta : \mathbb{Z}_{l+1} \times \mathbb{Z}_{l+1} \setminus \{0\} \to \mathbb{R}_{\ge 0}$ defined by $\delta(x,\hat{x}) = [\Delta]_{x,\hat{x}}$ in terms of the $(l+1) \times l$ matrix

$$\Delta = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix} \quad (10)$$

then the condition for successful error-only decoding then becomes

$$d(x^n, \hat{x}^n) < \frac{1}{2}(n - k + 1). \quad (11)$$

Proof: It follows directly from $d(x^n, \hat{x}^n) = \sum_{t=1}^{l} \sum_{s=0, s \ne t}^{l} \chi_{s,t} = v$. ∎

Remark 2: If we delete the first row which corresponds to the case where none of the first l most likely symbols is correct then the distortion measure is exactly the Hamming distortion.

Split covering approach:: We can break an error pattern x^n into two sub-error patterns $x^{LRPs} \triangleq x_{\sigma(1)} x_{\sigma(2)} \dots x_{\sigma(n_c)}$ of n_c least reliable positions and $x^{MRPs} \triangleq x_{\sigma(n_c+1)} \dots x_{\sigma(n)}$ of $n - n_c$ most reliable positions. Similarly, we can break an erasure pattern $\hat{x}^n$ into two sub-erasure patterns $\hat{x}^{LRPs} \triangleq \hat{x}_{\sigma(1)} \hat{x}_{\sigma(2)} \dots \hat{x}_{\sigma(n_c)}$ and $\hat{x}^{MRPs} \triangleq \hat{x}_{\sigma(n_c+1)} \dots \hat{x}_{\sigma(n)}$. Let z_{n_c} be the number of positions in the n_c LRPs where none of the first l most likely symbols is correct, or $z_{n_c} = |\{i = 1,2,\dots,n_c : x_{\sigma(i)} = 0\}|$. If we assign the set of all sub-error patterns $\hat{x}^{LRPs}$ to be an (n_c, k_c) t_c-covering code then $d(x^{LRPs}, \hat{x}^{LRPs}) \le t_c + z_{n_p}$ because this covering code has covering radius t_c. Since $d(x^n, \hat{x}^n) = d(x^{LRPs}, \hat{x}^{LRPs}) + d(x^{MRPs}, \hat{x}^{MRPs})$, in order to increase the probability that the condition (11) is satisfied we want to make $d(x^{MRPs}, \hat{x}^{MRPs})$ as small as possible by the use of the R-D approach. The following proposition summarizes how to generate a set of 2^R erasure patterns for multiple runs of error-only decoding.

Proposition 4: In each erasure pattern, the letter sequence at n_c LRPs is set to be a codeword of an (n_c, k_c) l-ary t_c-covering code. The letter sequence of the remaining $n - n_c$ MRPs is generated randomly by the R-D method (see Section II-C) with rate $R_{MRPs} = R - k_c \log_2 l$ and the distortion measure in (10). Since this covering code has l^{k_c} codewords, the total rate is $R_{MRPs} + \log_2 l^{k_c} = R$.

Example 6: For a (7,4,3) binary Hamming code which has covering radius $t_c = 1$, we take care of the 2 most likely symbols at each of the 7 LRPs. We see that 1001001 is a codeword of this Hamming code and then form erasure patterns $1001001 \hat{x}_8 \hat{x}_9 \dots \hat{x}_n$ with assumption that the positions

are written in increasing reliability order. The 2^{R-4} sub-erasure patterns $\hat{x}_8 \hat{x}_9 \ldots \hat{x}_n$ are generated randomly using the R-D approach with rate $(R-4)$.

Remark 3: While it also makes sense to use a covering codes for the n_c LRPs of the erasure patterns and set the the rest to be letter 1 (i.e. chose the most likely symbol as the hard-decision), our simulation results shows that the performance can be improved by using a combination of covering codes and random (i.e., generated by the R-D approach) codes.

B. Closed form rate-distortion functions

For some simple distortion measures, we can compute the R-D functions analytically in closed form. This method converges to the solution faster than the numerical method in Section III. The following two theorems describe how to compute the R-D functions for the simple distortion measures of Proposition 1 and 2.

Theorem 4: (Conventional error-and-erasure decoding) Let $p_i \triangleq \Pr(x_i = 1)$, the overall rate-distortion function is given by $R(D) = \sum_{i=1}^{n} \left[H(p_i) - H(\tilde{D}_i) \right]^+$ where $\tilde{D}_i \triangleq D_i + p_i - 1$ and $\tilde{D}_i$ can be found be a reverse water-filling procedure:

$$\tilde{D}_i = \begin{cases} \lambda & \text{if } \lambda < \min\{p_i, 1-p_i\} \\ \min\{p_i, 1-p_i\} & \text{otherwise} \end{cases}$$

where λ should be chosen so that $\sum_{i=1}^{n} \tilde{D}_i = D + \sum_{i=1}^{n} p_i - n$. The $R(D)$ function can be achieved by the test-channel input-probability distribution

$$q_{0,i} \triangleq \Pr(\hat{x}_i = 0) = \frac{1 - p_i - \tilde{D}_i}{1 - 2\tilde{D}_i} \text{ and } q_{1,i} \triangleq \Pr(\hat{x}_i = 1) = \frac{p_i - \tilde{D}_i}{1 - 2\tilde{D}_i}.$$

Proof: See [14]. ∎

Theorem 5: (Bit-level ASD case in Proposition 2) The overall rate-distortion function in this case is given by $R(D) = \sum_{i=1}^{N} [R_i(\lambda)]^+$ where $R_i(\lambda) = H(p_i) - H(\frac{1+\lambda}{1+\lambda+\lambda^2}) + (p_i - \frac{1+\lambda}{1+\lambda+\lambda^2})H(\frac{\lambda}{1+\lambda})$ and the distortion component D_i is given by

$$D_i = \begin{cases} \frac{1+2\lambda+3\lambda^2}{1+\lambda+\lambda^2} - p_i \frac{1+2\lambda}{1+\lambda} & \text{if } R_i(\lambda) > 0 \\ \min\{1, 3(1-p_i)\} & \text{otherwise} \end{cases}$$

where $\lambda \in (0,1)$ should be chosen so that $\sum_{i=1}^{N} D_i = D$. The $R(D)$ function can be achieved by the test-channel input-probability distribution $q_{i,0} \triangleq \Pr(\hat{b}_i = 0) = \frac{(1+\lambda) - p_i(1+\lambda+\lambda^2)}{1-\lambda^2}$ and $q_{i,1} = \Pr(\hat{b}_i = 1) = \frac{p_i(1+\lambda+\lambda^2) - \lambda(1+\lambda)}{1-\lambda^2}$.

Proof: See [14]. ∎

VI. SIMULATION RESULTS

Using simulations, we consider the performance of the (255,239) RS code over an AWGN channel with BPSK as the modulation format. The mBM-1 curve corresponds to standard error-and-erasure BM decoding with multiple erasure patterns. For $l > 1$, the mBM-l curves correspond to error-and-erasure BM decoding with multiple decoding trials using both erasures and top-l symbols. The mASD-m curves
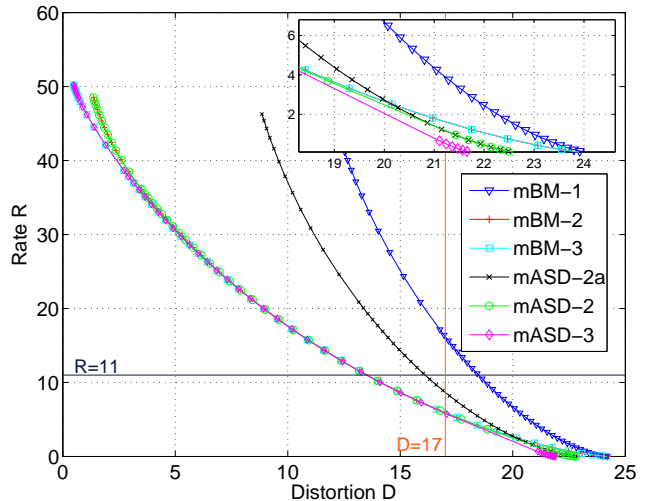

Figure 1. A realization of R-D curves at $E_b/N_0 = 5.2$dB for various decoding algorithms for the (255,239) RS code over an AWGN channel.
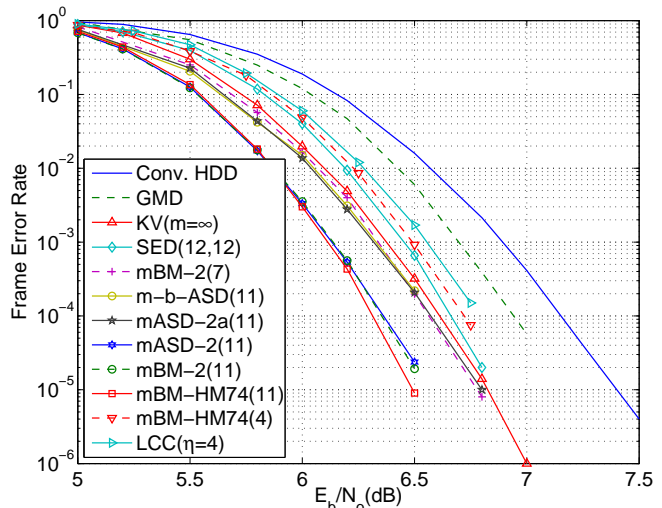

Figure 2. Performance of various decoding algorithms for the (255,239) RS code over an AWGN channel.

correspond to multiple ASD decoding trials with maximum multiplicity m. The number of trial decoding patterns is 2^R where R is denoted in parentheses in each algorithm's acronym (e.g., m-BM-2(11) uses $R = 11$).

Fig. 1 shows the R-D curves for various algorithms at $E_b/N_o = 5.2$ dB. For this code, the fixed threshold for decoding is $D = n - k + 1 = 17$. Therefore, one might expect that algorithms whose average distortion is less than 17 should have a frame error rate (FER) less than $\frac{1}{2}$. The R-D curve allows one to estimate the number of decoding patterns required to achieve this FER. Conventional BM decoding is very similar to mBM-1 decoding at rate 0. Notice that the mBM-1 algorithm at rate 0, which is very similar to conventional BM decoding, has an expected distortion of roughly 24. For this reason, the FER on conventional decoding is close to 1. The R-D curve tells us that trying roughly 2^{16} (i.e., $R = 16$) erasure patterns would reduce the FER to roughly $\frac{1}{2}$ because this is where the distortion drops down to 17. Likewise, the

mBM-2(11) algorithm has an expected distortion of less than 14. So we expect (and our simulations confirm) that the FER should be less than $\frac{1}{2}$. One weakness of this approach is that the R-D describes only the average distortion and does not directly consider the probability that the distortion is greater than 17. Still, we can make the following observations from the R-D curve. Even at high rates (e.g., $R \geq 4$), we see that the distortion D achieved by mBM-2 is roughly the same as mBM-3, mASD-2, and mASD-3 but smaller than mASD-2a and mBM-1. This implies that mBM-2 is no worse than the more complicated ASD based approaches for a wide range of rates (i.e., $4 \leq R \leq 35$).

The FER of various algorithms can be seen in Fig. 2. The focus on $R = 11$ allows us to make fair comparisons with SED(12,12). With the same number of decoding trials, mBM-2(11) outperforms SED(12,12) by 0.3 dB at an FER$= 10^{-4}$. Even mBM-2(7), with many fewer decoding trials, outperforms both SED(12,12) and the KV algorithm with $m = \infty$. Among all our proposed algorithms with rate $R = 11$, the mBM-HM74(11) achieves the best performance. This algorithm uses the Hamming (7,4) covering code for the 7 LRPs and the R-D approach for the remaining codeword positions.

Meanwhile, small differences in the performance between mBM-2(11), mBM-3(11), mASD-2(11), and mASD-3(11) suggest that: (i) taking care of the 2 most likely symbols at each codeword position is good enough for multiple decoding of high-rate RS code and (ii) multiple runs of error-and-erasure decoding is almost as good as multiple runs of ASD decoding. Recall that this result is also correctly predicted by the R-D analysis. Moreover, it is quite reasonable since we know that the gain of GS decoding, with infinite multiplicity, over the BM algorithm is negligible for high-rate RS codes.

To compare with the LCC($\eta = 4$) Chase-type approach used in [7], we also consider the mBM-HM74(4) algorithm, which uses the Hamming (7,4) covering codes for the 7 LRPs and the hard decision pattern for the remaining codeword positions. This shows that the covering code achieves better performance with the same number (2^4) decoding attempts. The comparison is not entirely fair, however, because of their low-complexity approach to multiple decoding. We believe, nevertheless, that their technique can be generalized to covering codes.

VII. CONCLUSION

A rate-distortion approach is proposed as a unified framework to analyze multiple decoding trials, with various algorithms, of RS codes in terms of performance and complexity. A connection is made between the complexity and performance (in some asymptotic sense) of these multiple-decoding algorithms and the rate and distortion of an associated R-D problem. Covering codes are also combined with the rate-distortion approach to mitigate the suboptimality of random codes when the effective block-length is not large. As part of this analysis, we also present numerical and analytical computations of the rate-distortion function for sequences of independent but non-identical sources. Finally, the simulation results show that our proposed algorithms based on the R-D approach achieve a better performance-versus-complexity trade-off than previously proposed algorithms. One key result is that, for high-rate RS codes, multiple-decoding using the standard BM algorithm is as good as multiple-decoding using more complex ASD algorithms.

In this paper, we only discuss the rate-distortion approach to solve the problem in (3). However, the performance can be further improved by focusing on the rate-distortion error-exponent. This allows us to approximately solve the covering problem for finite n rather than just as $n \to \infty$. The complexity of multiple decoding can also be decreased by using clever techniques to lower the complexity per decoding trial (e.g., [7]). We will address these two improvements in a future paper.

REFERENCES

[1] V. Guruswami and M. Sudan, "Improved decoding of Reed-Solomon and Algebraic-Geometry codes," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 1757–1767, Sept. 1999.
[2] R. Koetter and A. Vardy, "Algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Inform. Theory*, vol. 49, no. 11, pp. 2809–2825, Nov. 2003.
[3] G. D. Forney, Jr., "Generalized minimum distance decoding," *IEEE Trans. Inform. Theory*, vol. 12, no. 2, pp. 125–131, April 1966.
[4] S.-W. Lee and B. V. K. V. Kumar, "Soft-decision decoding of Reed-Solomon codes using successive error-and-erasure decoding," in *Proc. IEEE Global Telecom. Conf.*, New Orleans, LA, Nov. 2008, pp. 1–5.
[5] D. Chase, "A class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 170–182, Jan. 1972.
[6] M. F. H. Tang, Y. Liu and S. Lin, "On combining Chase-2 and GMD decoding algorithms for nonbinary block codes," *IEEE Commun. Letters*, vol. 5, no. 5, pp. 209–211, May 2001.
[7] J. Bellorado and A. Kavčić, "A low-complexity method for Chase-type decoding of Reed-Solomon codes," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, USA, 2006, pp. 2037–2041.
[8] H. Xia and J. R. Cruz, "Reliability-based forward recursive algorithms for algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Commun.*, vol. 55, no. 7, pp. 1273–1278, 2007.
[9] H. Xia, H. Wang, and J. R. Cruz, "A Chase-GMD algorithm for soft-decision decoding of Reed-Solomon codes on perpendicular channels," in *Proc. IEEE Int. Conf. Commun.*, Beijing, China, May 2008, pp. 1977–1981.
[10] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1983, iSBN 0-13-283796-X.
[11] R. E. Blahut, *Algebraic Codes for Data Transmission*. Cambridge University Press, 2003, iSBN-10 0521553741.
[12] ——, "Computation of channel capacity and rate distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, July 1972.
[13] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.
[14] P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "A rate-distortion perspective on multiple decoding attempts for Reed-Solomon codes," 2009. [Online]. Available: http://arxiv.org/abs/0908.2828
[15] J. Jiang and K. R. Narayanan, "Algebraic soft-decision decoding of Reed-Solomon codes using bit-level soft information," *IEEE Trans. Inform. Theory*, vol. 54, no. 9, pp. 3907–3928, Sept. 2008.
[16] R. J. McEliece, "The Guruswami-Sudan decoding algorithm for Reed-Solomon codes," *JPL Interplanetary Network Progress Report*, pp. 42–153, May 2003.