A Rate-Distortion Exponent Approach to 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*—Algorithms based on multiple decoding attempts of Reed-Solomon (RS) codes have recently attracted new attention. Choosing decoding candidates based on rate-distortion theory, as proposed previously by the authors, currently provides the best performance-versus-complexity trade-off. In this paper, an analysis based on the rate-distortion exponent is used to directly minimize the exponential decay rate of the error probability. This enables rigorous bounds on the error probability for finite-length RS codes and leads to modest performance gains. As a byproduct, a numerical method is derived that computes the rate-distortion exponent for independent non-identical sources. Analytical results are given for errors/erasures decoding.**

I. Introduction

The design of a computationally efficient soft-decision decoding algorithm for Reed-Solomon (RS) codes has been the topic of significant research interest for the past several years. Currently, there are several soft-decision decoding algorithms for RS codes which exhibit a wide range of trade-offs between computational complexity and error performance.

Among such decoding methods is a class of algorithms called multiple errors-and-erasures decoding. The algorithms belonging to this class first construct a set of erasure patterns based on the available soft information and then run an errors-and-erasures decoding algorithm, such as the Berlekamp-Massey (BM) algorithm, multiple times. Each time one erasure pattern in the set is used for decoding. By doing this, the algorithm outputs a list of candidate codewords and then chooses the best codeword from the list. Several algorithms of this type, including the popular generalized minimum distance (GMD) decoding algorithm, are discussed in [1], [2], [3], [4].

In [4], the authors proposed a rate-distortion (RD) approach for constructing the set of erasure patterns. The main idea is to choose an appropriate distortion measure so that the decoding is successful if and only if the distortion between the error pattern and erasure pattern is smaller than a fixed threshold. After that, a set of erasure patterns is generated randomly (similar to a random codebook generation) in order to minimize the expected minimum distortion. The approach was also extended to analyze multiple-decoding for decoding schemes beyond conventional errors-and-erasures decoding.

This material is based upon work supported by the National Science Foundation under Grant No. 0802124. The work of P. Nguyen was also supported in part by a Vietnam Education Foundation fellowship. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

One of the drawbacks in the RD approach is that the mathematical framework is only valid as the block-length goes to infinity. Therefore, we also consider the natural extension to a rate-distortion *exponent* (RDE) approach that studies the behavior of the probability, p_e, that the transmitted codeword is not on the list as a function of the block-length N. The overall error probability can be approximated by p_e because the probability that the transmitted codeword is on the list but not chosen is very small compared to p_e. Hence, our new approach essentially focuses on investigating the exponent at which the error probability decays as N goes to infinity.

The proposed RDE approach can also be considered as the generalization of the RD approach since the latter is a special case of the former when the RDE function tends to zero. Using the RDE analysis, our proposed approach also helps answer the following two questions: (i) What is the maximum rate-distortion exponent achievable at or below a given number of decoding attempts (or a given size of the set of erasure patterns)? (ii) What is the minimum number of decoding attempts required to achieve a rate-distortion exponent at or above a given level?

The paper is organized as follows. In Section II, we review multiple errors-and-erasures decoding algorithms and highlight the connection between multiple errors-and-erasures decoding and rate-distortion. Then, in Section III, we propose a RDE approach to construct a good set of erasure patterns for a finite length codewords. Next, we discuss how to compute the RDE function which is required in the proposed approach. Finally, simulation results are presented in Section V and conclusion is provided in Section VI.

II. Multiple errors-and-erasures Decoding

In this section, we discuss several multiple errors-and-erasures decoding algorithms. While each algorithm uses a different set of erasure patterns, the common trend is that one either erases or tries several different candidates for each symbol in the least reliable positions (LRPs). One focuses on the LRPs because the hard-decision made at these positions are more likely to be incorrect.

Let $\mathbb{F}_m$ be the Galois field with m elements denoted as $\alpha_1, \alpha_2, \ldots, \alpha_m$. We consider an (N,K) RS code of length N and dimension K over $\mathbb{F}_m$. Assume that we send a codeword $\mathbf{c} = (c_1, c_2, \ldots, c_N)$ over some channel and $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ is the received vector. A well-known decoding threshold states that a single attempt of errors-and-erasures decoding succeeds if and only if

$$2v + e < d_{min} = N - K + 1 \quad (1)$$

where e is number of erased symbols and v is the number of errors in unerased positions. A multiple errors-and-erasures decoding is considered to succeed if the decoding threshold (1) is satisfied for at least one attempt of decoding. Intuitively, the best case is when one erases an error and the worst case is when ones wastes an erasure on a hard-decision symbol that turns out be correct.

The first algorithm of this type is called Generalized Minimum Distance (GMD) decoding [1] where the set of erasure patterns is obtained by successively erasing the $0, 2, 4, \ldots, d_{min} - 1$ LRPs (with the assumption that the minimum distance d_{min} is odd). Recent work by Lee and Kumar [2] proposes a soft-information successive (multiple) error-and erasure decoding (SED) which constructs the set of erasure patterns in a more exhaustive way. Specifically, $\text{SED}(l, f)$ tries to erase all possible combinations of an even number less than or equal to f of positions within the l LRPs. The SED algorithm hence yields better performance but at increased complexity.

In an attempt to answer the question how to build a good set of erasure patterns in terms of performance-versus-complexity, in [4], we proposed a probabilistic method based on rate-distortion theory and random coding arguments instead of the deterministic methods which had been used in previously proposed algorithms. Basically, after defining x^N and $\hat{x}^N$ as an error pattern and an erasure pattern whose letters x_i's and $\hat{x}_i$'s are in the alphabets $\mathcal{X}$ and $\hat{\mathcal{X}}$ respectively, a letter-by-letter distortion measure $\delta : \mathcal{X} \times \hat{\mathcal{X}} \to \mathbb{R}_{\geq 0}$ is chosen properly so that the condition (1) can be reduced to the form

$$d\left(x^N, \hat{x}^N\right) < N - K + 1 \quad (2)$$

where the distortion between an error pattern and an erasure pattern $d\left(x^N, \hat{x}^N\right) = \sum_{i=1}^{N} \delta(x, \hat{x}_i)$ is smaller than a fixed threshold. In general, an appropriate distortion measure $\delta(j,k)$ for every $j \in \mathcal{X}$ and $k \in \hat{\mathcal{X}}$ should be specified.

Example 1: Consider a specific class of multiple errors-and-erasures (Berlekamp-Massey) top-ℓ decoding (mBM-ℓ) for an positive integer ℓ smaller than the field size m where at each codeword index, up to the ℓ-th most likely symbols are taken care of. In this case, $\mathcal{X} = \hat{\mathcal{X}} = \mathbb{Z}_{l+1}$ and $x^N \in \mathcal{X}^N$ where at each index i, $x_i = 0$ implies that using up to the ℓ-th most likely symbols as the hard-decision all gives an error, $x_i = j$ implies that the j-th most likely symbol is correct for $j = 1, 2, \ldots, \ell$; $\hat{x}^N \in \hat{\mathcal{X}}^N$ where at each index i, $\hat{x}_i = 0$ implies that an erasure is applied and $\hat{x}_i = k$ implies that the k-th most likely symbol is used as the hard-decision for $k = 1, 2, \ldots, \ell$. For example, mBM-1 is the case of multiple conventional errors-and-erasures decoding. The letter-by-letter distortion measure for mBM-1 is chosen in the following way

$$\begin{array}{ll} \delta(0,0) = 1 & \delta(0,1) = 2 \\ \delta(1,0) = 1 & \delta(1,1) = 0. \end{array} \quad (3)$$

It is also possible to choose appropriate distortion measures that work for $\ell > 1$ and other decoding schemes such as algebraic soft-decision (ASD) decoding. Still, the main idea is to convert the decoding threshold of the corresponding decoding scheme into the form of (2).

Thus, by viewing x^N (resp. $\hat{x}^N$) as a source sequence (resp. reproduction sequence) and choosing a suitable distortion measure, the task of designing a good set of erasure patterns turns out to be how to best approximate the source sequence with a minimum number of reproduction sequences. In other words, it becomes a covering problem where one wants to cover the most-likely error patterns with the fewest number of balls whose centers are erasure patterns. The main steps in the RD based algorithm are given here briefly, but more detail can be found in [4].

Step 1: Empirically compute the reliability matrix whose entries are $\Pr(c_i = \alpha_j | r_i)$ for $i = 1, 2, \ldots N$ and $j = 1, 2, \ldots m$. From this, determine probability matrix $\mathbf{P}$ where $p_{i,j} = \Pr(x_i = j)$ for $i = 1, 2, \ldots, N$ and $j \in \mathcal{X}$.

Step 2: Compute the RD function using $\mathbf{P}$. Determine the test-channel input-distribution matrix $\mathbf{Q}$ where $q_{i,k} = \Pr(\hat{x}_i = k)$ for $i = 1, 2 \ldots, N$ and $k \in \hat{\mathcal{X}}$ that achieves a point on the RD curve corresponding to a chosen rate R.

Step 3: Randomly generate a set $\mathcal{B}$ of 2^R erasure patterns using the distribution matrix $\mathbf{Q}$ in the correct reliability order of the codeword positions.

Step 4: Run multiple attempts of the corresponding decoding scheme using the set $\mathcal{B}$ to produce a list of candidate codewords.

Step 5: Use Maximum-Likelihood (ML) decoding to pick the best codeword on the list.

III. RATE-DISTORTION EXPONENT APPROACH

In the RD approach, the set $\mathcal{B}$ of $2^{N\bar{R}}$ (or 2^R) erasure patterns can be generated randomly so that[1]

$$\lim_{N \to \infty} \frac{1}{N} E_{x^N, \mathcal{B}} [\min_{\hat{x}^N \in \mathcal{B}} d(x^N, \hat{x}^N)] < \bar{D}.$$

Thus, for large enough N, with high probability we have $\min_{\hat{x}^N \in \mathcal{B}} d(x^N, \hat{x}^N) < N\bar{D} = D$. Basically, [4] focuses on minimizing the average minimum distortion with little knowledge of how the tail of the distribution behaves. In this paper, we instead focus on directly minimizing the probability that the minimum distortion is not less than the pre-determined threshold $D = N - K + 1$ (due to the condition (2)) with the help of an error-exponent analysis. The exact probability of interest is $p_e = \Pr(x^N : \min_{\hat{x}^N \in \mathcal{B}} d(x^N, \hat{x}^N) > D)$ that reflects how likely the decoding threshold (1) is going to fail.

In other words, every error pattern x^N can be covered by a sphere centered at an erasure pattern $\hat{x}^N$ except for a set of error patterns of probability p_e. The RDE analysis shows that p_e decays exponentially as $N \to \infty$ and the maximum exponent attainable is the RDE function. In our context, we have a source sequence x^N of N independent non-identical source components. We denote the rate-distortion exponent by $F(R,D)$ using unnormalized quantities (i.e., without dividing by N) and note that exponent used by other authors in [5], [6], [7] is often the normalized version $\bar{F}(R,D) \triangleq \frac{F(R,D)}{N}$.

The original RDE function $F(R,D)$, defined in [5] for a single source x, is given by[2]

[1]We denote the rate and distortion by R and D, respectively, using unnormalized quantities, i.e., $R = N\bar{R}$ and $D = N\bar{D}$.

[2]All logarithms are taken to base 2.

$$F(R,D) = \max_{\mathbf{w}} \min_{\tilde{\mathbf{p}} \in \mathcal{P}_{R,D}} \sum_j \tilde{p}_j \log \frac{\tilde{p}_j}{p_j}$$

where $p_j \triangleq \Pr(x=j)$, $w_{k|j} \triangleq \Pr(\hat{x}=k|x=j)$, and

$$\mathcal{P}_{R,D} = \left\{ \tilde{\mathbf{p}} \left| \begin{array}{c} \sum_j \sum_k \tilde{p}_j w_{k|j} \log \frac{w_{k|j}}{\sum_j \tilde{p}_j Q_{k|j}} \geq R \\ \sum_j \sum_k \tilde{p}_j w_{k|j} \delta_{jk} \geq D \end{array} \right. \right\}.$$

The RDE was first extensively discussed in [5], [6] and their results show that there exists a set $\mathcal{B}$ of roughly $2^{N\bar{R}}$ codewords, generated randomly using the test-channel input distribution matrix $\mathbf{Q}$, that achieves $\bar{F}(R,D)$. This gives the upper bound that for every $\varepsilon > 0$, we have

$$p_e \leq 2^{-N[\bar{F}(R,D)-\varepsilon]}. \tag{4}$$

for N large enough (see [8, p. 229]). An exponentially tight lower bound for p_e can also be obtained for N large enough (see [8, p. 236]) and this gives

$$\lim_{N\to\infty} -\frac{1}{N} \log p_e = \bar{F}(R,D).$$

Proposed algorithm: In the RDE approach proposed here, instead of computing the RD function, we need to compute the RDE function $F(R,D)$ along with the optimal test-channel input distribution matrix $\mathbf{Q}$ (see Section IV). This distribution serves as a replacement for the distribution used in Step 2 of the RD based algorithm given in the previous section. Apart from this, the other steps of that algorithm are unchanged for the proposed RDE-based algorithm.

Remark 1: The RDE approach possesses several advantages. First, it can help one estimate the smallest number of decoding attempts to get to a RDE of F (or get to an error probability of roughly $2^{-N\bar{F}}$) or, similarly, allow one to estimate the RDE (and error probability) for a fixed number of decoding attempts. Second, it provides a converse based on the sphere-packing bound lower bound for p_e. This implies that, given an arbitrary set $\mathcal{B}$ of roughly $2^{N\bar{R}}$ erasure patterns and any $\varepsilon > 0$, the probability p_e cannot be made lower than $2^{-N[\bar{F}(R,D)+\varepsilon]}$ for N large enough. Thus, no matter how one chooses the set $\mathcal{B}$ of erasure patterns, the difference between the induced probability of error and the p_e for the RDE approach becomes negligible for N large enough.

Remark 2: It is interesting to note that the RDE approach for ASD decoding schemes contains the special case where the codebook has only one entry (i.e., ASD decoding is run one time). In this case, the distribution of the codebook that maximizes the exponent implicitly generates the optimal multiplicity matrix. This is similar to the line of work [9], [10], [11] where various researchers tried to find the multiplicity matrix that optimizes the error-exponent obtained by either applying a Chernoff bound [9], [10] or using Sanov's theorem [11].

IV. COMPUTING THE RDE FUNCTION

In this section, we first present an extension of Arimoto's numerical method for computing the RDE function [12] that works for any chosen discrete distortion measure. Next, we consider some special case where we can give an analytical treatment of the function.

A. Numerical computation of RDE function

For each discrete source component x_i, given two parameters $s \geq 0$ and $t \leq 0$, the Arimoto algorithm given in [12] computes the RDE function numerically as follows.

- Step 1: Choose an arbitrary all-positive distribution vector $\underline{q}^{(0)} = \left(q_1^{(0)}, q_2^{(0)}, \ldots, q_{|\hat{\mathcal{X}}|}^{(0)}\right)$.
- Step 2: Iterate the following steps at $\tau = 0, 1, \ldots$

$$w_{k|j}^{(\tau)} = \frac{q_k^{(\tau)} 2^{t\delta_{jk}}}{\sum_k q_k^{(\tau)} 2^{t\delta_{jk}}}$$

$$q_k^{(\tau+1)} = \frac{\left\{\sum_j p_j 2^{-st\delta_{jk}} (w_{k|j}^{(\tau)})^{(1+s)}\right\}^{\frac{1}{1+s}}}{\sum_k \left\{\sum_j p_j 2^{-st\delta_{jk}} (w_{k|j}^{(\tau)})^{(1+s)}\right\}^{\frac{1}{1+s}}}.$$

for $j \in \mathcal{X}$ and $k \in \hat{\mathcal{X}}$.

It is shown by Arimoto that $w_{k|j}^{(\tau)} \to w_{k|j}^{\star}$ and $q_k^{(\tau)} \to q_k^{\star}$ as $\tau \to \infty$. Using the resulting $w_{k|j}^{\star}$ and $q_k^{\star}$, we can compute

$$F = \sum_j \tilde{p}_j^{\star} \log \frac{\tilde{p}_j^{\star}}{p_j} \tag{5}$$

$$R = \sum_j \sum_k \tilde{p}_j^{\star} w_{k|j}^{\star} \log \frac{w_{k|j}^{\star}}{\sum_j \tilde{p}_j^{\star} w_{k|j}^{\star}} \tag{6}$$

$$D = \sum_j \sum_k \tilde{p}_j^{\star} w_{k|j}^{\star} \delta_{jk} \tag{7}$$

where $\tilde{p}_j^{\star} = \frac{p_j (\sum_k q_k^{\star} 2^{t\delta_{jk}})^{-s}}{\sum_j p_j (\sum_k q_k^{\star} 2^{t\delta_{jk}})^{-s}}$.

However, in the context we consider, the source (error pattern) x^N comprises independent but not necessarily identical source components x_i's. The complexity is a problem if we consider a group of source letters $(j_1, j_2, \ldots, j_N)$ as a supper-source letter $\mathcal{J}$, a group of reproduction letters $(k_1, k_2, \ldots, k_N)$ as a super-reproduction letter $\mathcal{K}$ and apply the Arimoto algorithm straightforwardly . Instead, we can avoid this computational obstacle by choosing the initial distribution still arbitrarily but following a factorization rule $q_{\mathcal{K}}^{(0)} = \prod_{i=1}^N q_{k_i}^{(0)}$. Then, we can verify that this factorization rule still holds for $w_{\mathcal{K}|\mathcal{J}}^{(\tau)}$ and $q_{\mathcal{K}}^{(\tau)}$ after every step of the Arimoto algorithm. This leads to

$$w_{\mathcal{K}|\mathcal{J}}^{\star} = \prod_{i=1}^N w_{k_i|j_i}^{\star} \quad \text{and} \quad q_{\mathcal{J}}^{\star} = \prod_{i=1}^N q_{k_i}^{\star}.$$

Combining with $\delta_{\mathcal{J}\mathcal{K}} = \sum_{i=1}^N \delta_{j_i k_i}$ and $p_{\mathcal{J}} = \prod_{i=1}^N p_{j_i}$, we have

$$\tilde{p}_{\mathcal{J}}^{\star} = \prod_{i=1}^N \tilde{p}_{j_i}^{\star}.$$

This gives the following proposition.

Proposition 1: (Factored Arimoto algorithm for RDE function) Consider a discrete source x^N of independent but non-identical source components x_i's. Given parameters $s \geq 0$ and $t \leq 0$, the exponent, rate and distortion are given by

$$F|_{s,t} = \sum_{i=1}^N F_i|_{s,t}, \ R|_{s,t} = \sum_{i=1}^N R_i|_{s,t}, \ D|_{s,t} = \sum_{i=1}^N D_i|_{s,t}$$

where the components $F_i|_{s,t}, R_i|_{s,t}, D_i|_{s,t}$ are computed parametrically by the Arimoto algorithm.

B. Analytical computation of RDE function

In this subsection, we consider the case m-BM1 whose distortion measure is given in (3). We study the setup that RS codewords defined over Galois field $\mathbb{F}_m$ are transmitted over the m-ary symmetric channel (m-SC) which for each parameter p can be modeled as

$$\Pr(r|c) = \begin{cases} p & \text{if } r = c \\ (1-p)/(m-1) & \text{if } r \neq c. \end{cases}$$

Here, c (resp. r) is the transmitted (resp. received) symbol and $r, c \in \mathbb{F}_m$. With this channel model, we consider p not too small so that $p > (1-p)/(m-1)$. Therefore, at each index i of the codeword, the hard-decision is also the received symbol and then it is correct with probability p. Thus, we have $p_{i,1} \triangleq \Pr(x_i = 1) = p$ for every index i of the error pattern x^N. That means, in this context we have a source x^N with i.i.d. binary components x_i. Since the components x_i are i.i.d we can treat each x_i as a binary source X with $\Pr(X=1) \triangleq p$ and $\Pr(X=0) = 1-p \triangleq \bar{p}$ and first compute the RDE function for this source X.

According to [5], for any admissible (R,D) pair we can find two parameters $s \geq 0$ and $t \leq 0$ so that $F(R,D)$ can be parametrically evaluated as

$$\begin{aligned} F(R,D) &= sR - stD + \max_{q_1}\left(-\log f(q_1)\right) \\ &= sR - stD - \log \min_{q_1} f(q_1) \end{aligned}$$

where

$$f(q_1) = \bar{p}\left(\sum_k q_k 2^{t\delta_{0k}}\right)^{-s} + p\left(\sum_k q_k 2^{t\delta_{1k}}\right)^{-s}$$

and R, D are given in terms of optimizing $q^\star$ which we will discuss later.

For the distortion measure in (3) and note that $q_0 = 1 - q_1$, we have

$$f(q_1) = \bar{p}\left((1-q_1)2^t + q_1 2^{2t}\right)^{-s} + p\left((1-q_1)2^t + q_1\right)^{-s}$$

which is a convex function in q_1. We then see that

$$\frac{\partial f}{\partial q_1} = 0 \Leftrightarrow q_1^\star = \frac{1+2^t}{1-2^t}\left(\frac{1}{1+2^t} - \frac{\bar{p}^{\frac{1}{s+1}}}{2^{\frac{st}{s+1}} p^{\frac{1}{s+1}} + \bar{p}^{\frac{1}{s+1}}}\right) \triangleq \beta.$$

In order to minimize $f(q_1)$ over $q_1 \in [0,1]$, we consider three following cases where the optimal $q_1^\star$ is either on the boundary or at a point with zero gradient.

- Case 1: $0 \leq p \leq \frac{2^t}{1+2^t}$ then $\beta \leq 0$. Since f convex, it is non-decreasing in the interval $[\beta, \infty)$ and therefore in the interval $[0,1]$. Thus, the optimal $q_1^\star = 0$ and we can also compute from (5), (6), (7) that

$$D = 1; \quad R = 0; \quad F = 0 = D_{KL}(u||p)$$

where in this case $u = p$.

- Case 2: $1 \geq p \geq \frac{1}{1+2^{t(2s+1)}}$ then $\beta \geq 1$. Since f convex, it is non-increasing in the interval $(-\infty, \beta]$ and therefore in the interval $[0,1]$. Thus, the optimal $q_1^\star = 1$ and similarly we get

$$D = \tfrac{2\bar{p}}{p 2^{2ts} + \bar{p}}; \quad R = 0; \quad F = D_{KL}(u||p)$$

where in this case $u = 1 - \frac{D}{2}$. We can further see that $D \in [2(1-p), 1]$ and $u \in [1-D, p]$.

- Case 3: $\frac{2^t}{1+2^t} < p < \frac{1}{1+2^{t(2s+1)}}$ then $\beta \in [0,1]$. In this case, the optimal $q_1^\star = \beta$. We then can find $w_{k|j}^\star = \frac{q_k^\star 2^{t\delta_{jk}}}{\sum_k q_k^\star 2^{t\delta_{jk}}}$ according to [5] and plug in (5), (6), (7) to get[3]

$$\begin{aligned} D &= \frac{2^t}{1+2^t} + 1 - u \\ R &= H(u) - H(u + D - 1) \\ F &= D_{KL}(u||p) \end{aligned}$$

where $u = \frac{2^{\frac{st}{s+1}} p^{\frac{1}{s+1}}}{2^{\frac{st}{s+1}} p^{\frac{1}{s+1}} + \bar{p}^{\frac{1}{s+1}}}$. With this notation of u, we can express $q_1^\star = \frac{1-D}{3-2(u+D)}$ and $q_0^\star = \frac{2(1-u)-D}{3-2(u+D)}$. We can see that $D \in (1-p, 1)$. It can also be verified that, in this case, by varying s and t, u spans $(1-D, 1-D/2)$ and R spans $(0, H(1-D))$.

Based on the above analysis, we obtain the following lemmas and theorems.

Lemma 1: Let $h(u) = H(u) - H(u+D-1)$ map $u \in [1-D, 1-D/2)$ to R. Then, the inverse mapping of h,

$$h^{-1} : (0, H(1-D)] \to [1-D, 1-D/2),$$

is well-defined and maps R to u.

Proof: We first notice that $h(u)$ is strictly decreasing since the derivative is negative over $[1-D, 1-D/2)$, hence the mapping $h : [1-D, 1-D/2) \to (0, H(1-D)]$ is one-to-one. From the analysis above, one can also see that h is onto. ■

Theorem 1: Using mBM-1 with 2^R decoding attempts where $R \in (0, NH(1-\frac{D}{N})]$, the maximum rate-distortion exponent that can be achieved is

$$F = N D_{KL}\left(h^{-1}(R/N)\,||\,p\right). \tag{8}$$

Proof: First, note that in our context where we have a source sequence x^N of N i.i.d. source components, the rate and exponent for each source component is now $\frac{R}{N}$ and $\frac{F}{N}$. From Case 3 in the analysis above and from Lemma 1, we have

$$F/N = D_{KL}(u||p) = D_{KL}\left(h^{-1}(R/N)\,||\,p\right)$$

and the theorem follows. ■

Lemma 2: Let $g(u) = D_{KL}(u||p)$ map $u \in [1-D, p]$ to F. Then, the inverse mapping of g,

$$g^{-1} : [0, D_{KL}(1-D\,||\,p)] \to [1-D, p]$$

is well-defined and maps F to u.

Proof: We first see that $g(u)$ is a strictly convex function and achieved minimum value at $u = p$ and therefore $g(u)$ is strictly decreasing over $[1-D, p]$. Thus, the mapping $g : [1-D, p] \to [0, D_{KL}(1-D\,||\,p)]$ is one-to-one. From the analysis above, one can also see that g is onto. ■

Theorem 2: In order to achieve a rate-distortion exponent of $F \in [0, N D_{KL}(1-D\,||\,p)]$, the minimum number of decoding attempts required for mBM-1 is 2^R where

$$R = N\left[H\left(g^{-1}(F/N)\right) - H\left(g^{-1}(F/N) + D/N - 1\right)\right]^+$$

Proof: We also note that the rate, distortion and exponent for each source component is $\frac{R}{N}, \frac{D}{N}$ and $\frac{F}{N}$ respectively. Combining all the cases in the above analysis, we have

$$R/N = \left[H\left(g^{-1}(F/N)\right) - H\left(g^{-1}(F/N) + D/N - 1\right)\right]^+$$

[3]The binary entropy function is $H(u) \triangleq -u\log u - (1-u)\log(1-u)$ and the Kullback-Leibler divergence is $D_{KL}(u||p) \triangleq u \log \frac{u}{p} + (1-u)\log\frac{1-u}{1-p}$.

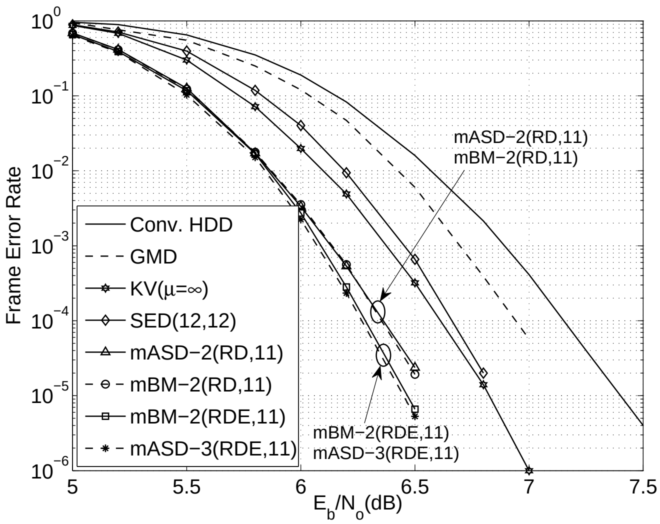

Figure 1. Performance of various decoding algorithms for the (255,239) RS code over an AWGN channel.

and the theorem follows. ■

V. SIMULATION

Simulations of the proposed algorithm were conducted for the (255,239) RS code over an AWGN channel with BPSK as the modulation format. In Fig. 1, the mBM-2(RD,11) curve belongs to the algorithm mBM-2 using RD approach proposed in [4] while the mBM-2(RDE,11) one corresponds to the algorithm mBM-2 using RDE approach proposed in this paper. The label SED(12,12) denotes the algorithm presented in [2]. While all these three algorithms use the same number of 2^{11} erasure patterns, at a FER of 10^{-4}, the mBM2(RDE,11) provides a performance gain of roughly 0.4 dB over the SED(12,12) and outperforms the mBM2(RD,11) by about 0.1 dB. The conventional HDD and the GMD algorithms have modest performance since they use only one or a few decoding attempts. Compared to the conventional HDD, the proposed algorithm mBM-2(RDE,11) gives approximately a 0.9 dB gain. It also outperforms the Koetter-Vardy (KV) algorithm [13] with infinite multiplicity ($\mu = \infty$). The performance of mBM-2(RDE,11) is roughly the same as the performance of mASD-3(RDE,11). This implies that, for this setup, algorithms based on multiple trials of BM decoding perform as good as algorithms based on running the more complicated ASD decoding multiple times. In Fig. 2, we simulate the performance mBM-1(RDE,11) for the same RS code over an *m*-SC channel. One curve reflects the simulated frame-error rate (FER) and the other is the approximation derived from 2^{-F} where F is given in (8) with $R = 11$.

VI. CONCLUSION

A RDE-based algorithm has been proposed for multiple decoding attempts of RS codes. The RDE analysis shows that this approach has several advantages. Firstly, the RDE approach achieves a near optimal performance-versus-complexity trade-off among algorithms that consider running a decoding scheme multiple times (see Remark 1). Secondly, it can help one estimate the error probability using exponentially tight bounds for N large enough. Simulations are used to confirm that algorithms using this approach achieve a better trade-off than previously known algorithms. Along with this, a numerical method is given to compute the required RDE function.

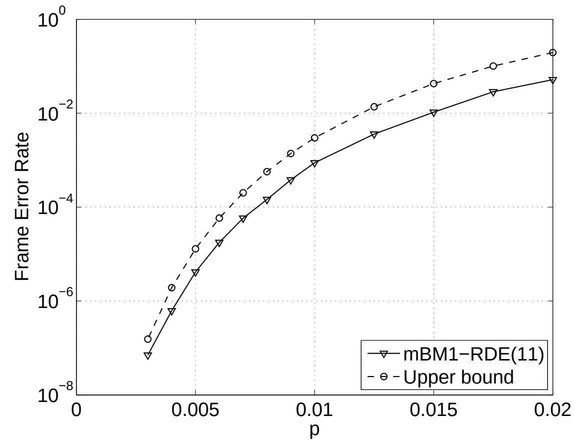

Figure 2. Performance of mBM-1(RDE,11) and its approximation 2^{-F} where F is given in (8) for the (255,239) RS code over an m-SC(p) channel.

Our future work focuses on extending this approach to analyze multiple decoding attempts for ISI channels. In this case, it makes sense for the decoder to consider multiple candidate error-events during decoding. Extending the RD approach directly gives a RD problem for Markov sources in the large distortion regime. Some work is required though because this is a well-known open problem.

REFERENCES

[1] G. D. Forney, Jr., "Generalized minimum distance decoding," *IEEE Trans. Inform. Theory*, vol. 12, no. 2, pp. 125–131, April 1966.

[2] 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.

[3] 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, July 2006, pp. 2037–2041.

[4] P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "A rate-distortion perspective on multiple decoding attempts for Reed-Solomon codes," in *Proc. 47th Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Oct. 2009.

[5] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 4, pp. 405–417, July 1974.

[6] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 197–199, March 1974.

[7] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*. Akademiai Kiado, Budapest, Hungary, 1981.

[8] R. E. Blahut, *Principles and practice of information theory*. Addison-Wesley, 1987, iSBN-0201107090.

[9] M. El-Khamy and R. J. McEliece, "Interpolation multiplicity assignment algorithms for algebraic soft-decision decoding of Reed-Solomon codes," *AMS-DIMACS volume on Algebraic Coding Theory and Information Theory*, vol. 68, pp. 99–120, 2005.

[10] N. Ratnakar and R. Koetter, "Exponential error bounds for algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3899–3917, 2005.

[11] H. Das and A. Vardy, "Multiplicity assignments for algebraic soft-decoding of Reed-Solomon codes using the method of types," in *Proc. IEEE Int. Symp. Information Theory*, Seoul, Korea, June 2009, pp. 1248–1252.

[12] S. Arimoto, "Computation of random coding exponent functions," *IEEE Trans. Inform. Theory*, vol. 22, no. 6, pp. 665–671, Nov. 1976.

[13] 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.