Stochastic Erasure-Only List Decoding Algorithms for Reed-Solomon Codes

Chang-Ming Lee and Yu T. Su, *Senior Member, IEEE*

***Abstract*—We present a novel stochastic decoding algorithm for Reed-Solomon codes. We apply an iterative Monte Carlo based approach called the Cross-Entropy method to produce, in every iteration, a set of random error locator vectors, each indicates $n-k$ possible erasure positions within a received word. We associate each error locator vector with a candidate codeword by erasures-only decoding the received word, using the error locator vector to locate the erasures. Each iteration results in a new elite set that contains the best E candidate codewords. To increase the search radius and enhance the decoder performance we use the randomly drawn samples to generate what we call virtual received words from which extra candidate codewords and thus candidate elite members can be obtained. The proposed algorithms offer both complexity and performance advantages over some existing algebraic decoding algorithms for high rate RS codes.**

***Index Terms*—Cross-Entropy method, list decoding, Reed–Solomon code.**

I. Introduction

THE class of Reed–Solomon (RS) codes [1] is a powerful error control code that has been used in a wide variety of applications, ranging from data storage systems to deep-space and wireless communications. These codes are usually decoded by hard-decision decoding (HDD) algorithms using the Berlekamp-Massey (BM) algorithm [2] and the Euclidean algorithm [3].

Many soft-decision decoding (SDD) algorithms that use reliability information to assist HDD have been developed. Earlier proposals include Forney's generalized minimum distance (GMD) decoding algorithm [5], the Chase II algorithm [6], and the combined Chase II-GMD algorithm [7]. These algorithms give a moderate performance improvement over HDD solutions with reasonable complexity. Guruswami and Sudan (GS) [4] invented an algebraic list decoding algorithm which corrects beyond half the minimum distance. Koetter and Vardy (KV) [8] proposed an algebraic SDD algorithm based on a multiplicity assignment scheme to improve the GS algorithm. The KV algorithm can significantly outperform HDD for low rate RS codes. However, to achieve large coding gain, the complexity can be prohibitively large.

Manuscript received February 13, 2009; revised May 01, 2009. First published May 12, 2009; current version published June 09, 2009. This work was supported in part by Taiwan National Science Council under Grant 97-2221-E009-082-MY3. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yimin Zhang.

The authors are with the Department of Communications Engineering, National Chiao Tung University, Hsinchu 30056, Taiwan (e-mail: cmlee.cm89g@nctu.edu.tw, ytsu@mail.nctu.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2009.2022792

In this letter, we apply the Cross-Entropy (CE) method [9] to develop a Monte Carlo based iterative SDD algorithm which renders an improved algebraic SDD decoding performance. The CE method is an elegant practical principle for simulating rare events which approximates the probability of the rare event by means of a family of parameterized probabilistic models. Our stochastic erasure-only list decoding (SEOLD) algorithm uses the extended CE method for optimization problem by considering an optimal event as a rare event.

The rest of this letter is organized as follows. Some preliminaries are given in Section II. In Section III, the general stochastic SDD algorithm based on the CE method is introduced. One of the proposed stochastic decoding algorithms is presented in Section IV. Some simulation results and discussions are presented in Section V.

II. Preliminary

Let $\mathbf{C}$ be an (n,k) RS code over $GF(2^m)$ with minimum Hamming distance $d_{\min} = n-k+1$. Let $\mathbf{c}=(c_0,\ldots,c_{n-1})$ be a codeword in $\mathbf{C}$ with the binary expansion $\bar{\mathbf{c}}=(\bar{c}_0,\bar{c}_1,\ldots,\bar{c}_{nm-1})$. Using binary phase-shift-keying (BPSK), the transmitter maps the binary imaged codeword $\bar{\mathbf{c}}$ into the bipolar vector

$$\Psi(\bar{\mathbf{c}})=\bar{\mathbf{x}}=(\bar{x}_0,\ldots,\bar{x}_{nm-1}),\quad \bar{x}_j=\Psi(\bar{c}_j)=(-1)^{\bar{c}_j} \tag{1}$$

and sends it over an additive white Gaussian noise (AWGN) channel with zero mean and power spectral density $N_0/2$. The received sequence at the output of the matched filter is $\bar{\mathbf{y}}=(\bar{y}_0,\ldots,\bar{y}_{nm-1})$ where $\bar{y}_j=\bar{x}_j+\bar{w}_j$ and $\bar{w}_j$'s are statistically independent Gaussian random variables with zero mean and variance $N_0/2$.

Let $\bar{\mathbf{z}}=(\bar{z}_0,\ldots,\bar{z}_{nm-1})$ be the hard decision binary vector of the received bit sequence $\bar{\mathbf{y}}$, i.e.,

$$\bar{z}_j=\begin{cases}0, & \bar{y}_j>0\\ 1, & \text{otherwise}\end{cases} \tag{2}$$

and $\mathbf{z}=(z_0,\ldots,z_{n-1})$ be the corresponding symbol vector. Denoted by $\bar{\Gamma}=(\bar{\gamma}_1,\ldots,\bar{\gamma}_{nm-1})$ the reliability vector of $\bar{\mathbf{y}}$ in which $\bar{\gamma}_j$ is the magnitude of the log-likelihood ratio (LLR) associated with the corresponding hard-limited bit $\bar{z}_j$

$$L(\bar{c}_j)=\log\frac{P(\bar{c}_j=0|\bar{\mathbf{y}})}{P(\bar{c}_j=1|\bar{\mathbf{y}})} \tag{3}$$

and define the symbol reliability vector $\Gamma=(\gamma_0,\ldots,\gamma_{n-1})$ of $\mathbf{z}$ by

$$\gamma_i=\min_j\bar{\gamma}_j,\ \ j\in\{im,\ldots,(i+1)m-1\} \tag{4}$$

III. Stochastic List Decoding Algorithm

A. Algebraic Erasures-Only (EO) Decoding

It is well-known that RS codes are maximum-distance separable (MDS) which implies that any k coordinates (symbols) in an RS codeword can be used to determine the remaining $n-k$ symbols. Hence it is sufficient to decide k correct (message) or $n-k$ incorrect (error) coordinates of a codeword. Let $\mathbf{E}_L$ be the collection of all combinations of $n-k$ error coordinates,

$$\mathbf{E}_L = \left\{ \mathbf{s} = (s_0, \ldots, s_{n-1}) \,\middle|\, s_i \in \{0,1\}, \sum_i s_i = n-k \right\} \tag{5}$$

where $s_i = 1$ if the ith coordinate is in error. Then a straightforward decoding schedule is given as below:

(a) For all $\mathbf{s} \in \mathbf{E}_L$, erase the corresponding $n-k$ error coordinates of the received word $\mathbf{z}$ and decode by the erasures-only (EO) decoder. The resulting codeword set is denoted by $\mathbf{C}_\mathbf{z}$.

(b) Choose the codeword from $\mathbf{C}_\mathbf{z}$ with the best score, e.g., the one whose Euclidean distance from the received word is the smallest, as the decoder output.

It can be easily confirmed that for any $\mathbf{c} \in \mathbf{C}_\mathbf{z}$, $d_H(\mathbf{c},\mathbf{z}) \leq n-k$, where $d_H(\mathbf{c},\mathbf{z})$ is the Hamming distance between $\mathbf{c}$ and $\mathbf{z}$. Therefore, the transmitted codeword belongs to $\mathbf{C}_\mathbf{z}$ if the number of error symbols is less than $d_{\min}$. Furthermore, **(b)** is equivalent to the following minimization problem

$$\arg\min_{\mathbf{c}} d(\Psi(\bar{\mathbf{c}}), \bar{\mathbf{y}}) \text{ subject to } \mathbf{c} \in \mathbf{C}_\mathbf{z} \tag{6}$$

where $\Psi(\cdot)$ is defined by (1) and $d(\bar{\mathbf{a}}, \bar{\mathbf{b}})$ is the Euclidean distance (ED) between the nm-ary real vectors $\bar{\mathbf{a}}$ and $\bar{\mathbf{b}}$.

B. A Stochastic List Decoding Idea

Each *error locator vector* (ELV) $\mathbf{s} \in \mathbf{E}_L$ represents a particular set of $n-k$ possible error coordinates and has a corresponding codeword $\mathbf{c_s}$ that belongs to $\mathbf{C}_\mathbf{z}$. We denote the latter relationship by $\mathbf{s} \sim \mathbf{c_s}$. Although more than one ELV may be associated with the same codeword, the complexity of searching for the optimal solution $\mathbf{c}^*$ in the error location domain $\mathbf{E}_L$ is still extremely high because the cardinality of $\mathbf{E}_L$ is $\binom{n}{k}$ and only a few (or one) elements in $\mathbf{E}_L$, depending on the number and locations of the received errors, can be used to reconstruct $\mathbf{c}^*$.

Suppose we model the selection of the ELV $\mathbf{s}$ from $\mathbf{E}_L$ as a stochastic (vector-valued) experiment governed by a family of parameterized distributions $\{f(\mathbf{s};\mathbf{u})\}$ with $\mathbf{u} \in \nu$ being a real-valued parameter vector. Usually $f(\mathbf{s};\mathbf{u})$ is assumed to be uniformly distributed due to the lack of priori information whence the search in (6) is exhaustive unless some algebraic properties of the code are used. One way to solve (6) efficiently is to find a parameter $\mathbf{v}^*$ such that $f(\mathbf{s};\mathbf{v}^*) = \delta(\mathbf{s}-\mathbf{s}^*)$ where $\mathbf{s}^* \sim \mathbf{c}^*$. Then drawing one sample from $f(\mathbf{s};\mathbf{v}^*)$ is sufficient to obtain the optimal solution $\mathbf{c}^*$. To get around the difficulty that $\mathbf{c}^*$ is not known, one notices that the optimization problem (6) is related to the estimation of the probability $P(d(\Psi(\bar{\mathbf{c}}_s), \bar{\mathbf{y}}) \leq \eta | \mathbf{s} \sim \mathbf{c_s})$, which is a rare event when $\eta = \eta^* = d(\Psi(\bar{\mathbf{c}}^*), \bar{\mathbf{y}})$. The connection comes from the fact that efficient estimation of a rare event can be achieved by the method of importance sampling and in this case the optimal importance density is $f(\mathbf{s};\mathbf{v}^*)$.

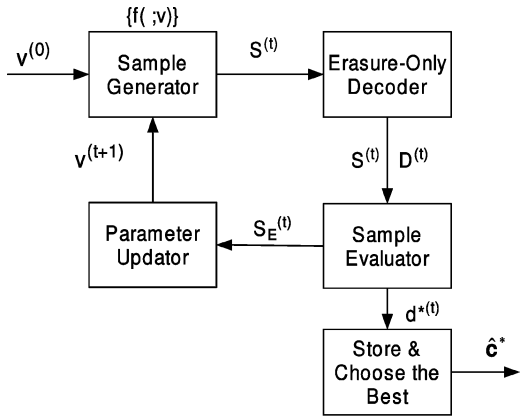

Fig. 1. Flow chart of a stochastic decoder for RS codes.

Without the knowledge of the threshold η^*, we start with a proper importance density $f(\mathbf{s};\hat{\mathbf{v}})$ to generate samples of $\mathbf{s}$ and compute an initial estimate $\hat{\eta}$ for η. Ideally, we can use those drawn samples which satisfy $d(\Psi(\bar{\mathbf{c}}_s), \bar{\mathbf{y}}) \leq \hat{\eta}$ to obtain new parameter value $\hat{\mathbf{v}}'$ such that $f(\mathbf{s};\hat{\mathbf{v}}')$ is closest to $f(\mathbf{s};\mathbf{v}^*)$ in the Kullback-Leibler (KL) sense, i.e., the CE between $f(\mathbf{s};\hat{\mathbf{v}}')$ and $f(\mathbf{s};\mathbf{v}^*)$ is minimized. Since $\mathbf{v}^*$ is unknown, we choose $\hat{\mathbf{v}}'$ such that $f(\mathbf{s};\hat{\mathbf{v}}')$ is closest to the empirical distribution of $\mathbf{s}$ in those samples that are generated by $f(\mathbf{s};\hat{\mathbf{v}})$ and satisfy $d(\Psi(\bar{\mathbf{c}}_s), \bar{\mathbf{y}}) \leq \hat{\eta}$ for this empirical distribution is likely to be a good approximation of $f(\mathbf{s};\mathbf{v}^*)$. New samples of $\mathbf{s}$ are then produced by the updated importance density $f(\mathbf{s};\hat{\mathbf{v}}')$ to compute new estimate $\hat{\eta}'$. This iterative procedure continues until $|\hat{\eta} - \hat{\eta}'|$ is less than a predetermined threshold.

The above method is known as the CE method [9] which is an iterative procedure that consists of the following two phases in each iteration.

- Generate samples from the specified importance density given by the parameters from the previous iteration.
- Update the parameters for next iteration according to the order of the score values associated with the drawn samples and the minimizing CE criterion.

Based on the above discussion, we propose a generic Monte Carlo based SDD algorithm as shown in Fig. 1 and in Table I with some detailed description given in Section IV.

C. Convergence and Complexity

Different convergence conditions and results have been discussed for the deterministic CE method and its extensions in [10] where it is also proved that convergence in distribution or η can be guaranteed but needs a proper tuning of the parameters of the algorithm such as the number of samples N, number of elites E, and smoothing factor ρ. Convergence to the global minimum is ensured only if a large sample size N is used. On the other hand, the computing complexity is related to N and is given by $O(N(n-k)^2)$. It is obvious that the decoding performance can be improved by using a larger N. As we retain the the best sample at the end of each iteration, the decoding performance is also improved by increasing the iteration number T. As an early-stopping at any iteration produces a decoded codeword, we say the algorithm converges if the sequence of decoded codewords converges. With a modest N, we found that the decoded codewords converge to the same codeword within

TABLE I
A STOCHASTIC LIST DECODING ALGORITHM

1. Define a family of probability densities $\{f(\cdot;\mathbf{v}), \mathbf{v}\in\nu\}$ on the search space $\mathbb{R}^{nm}$. Initialize $\mathbf{v}^{(0)}$. Set $t=1$.
2. Generate a sample set $\mathbf{S}^{(t)}$ whose N random vector samples are drawn from $f(\cdot;\mathbf{v}^{(t)})$. Regard the magnitudes as bit LLRs and convert them into symbol reliabilities. Erase the $n-k$ least reliable symbols and decode the received word by EO decoding.
3. Evaluate Euclidean distances between the decoded codewords and the received word. Select the E vector samples with best metrics as the new elite set $\mathbf{S}_E^{(t)}\subset\mathbf{S}^{(t)}$ and store the best decoded codeword $\mathbf{d}^{*(t)}$ in $\mathbf{D}^{(t)}$.
4. Evaluate the new parameter $\mathbf{v}^{(t+1)}$ by solving $\mathbf{v}^{(t+1)}=\arg\max_{\mathbf{v}}\frac{1}{|\mathbf{S}_E^{(t)}|}\sum_{\mathbf{s}_\ell^{(t)}\in\mathbf{S}_E^{(t)}}\ln f(\mathbf{s}_\ell^{(t)};\mathbf{v})$ [9]. Update $\mathbf{v}^{(t+1)}$ via $\mathbf{v}^{(t+1)}=\rho\mathbf{v}^{(t+1)}+(1-\rho)\mathbf{v}^{(t)}$ where $0<\rho<1$.
5. Terminate decoding if the stopping criterion is met. Choose the best codeword from the list $\{\mathbf{d}^{*(t)};\ \forall\, t\}$, say $\hat{\mathbf{c}}^*$, as the decoder output. Otherwise increase t by 1 and return to step 2.

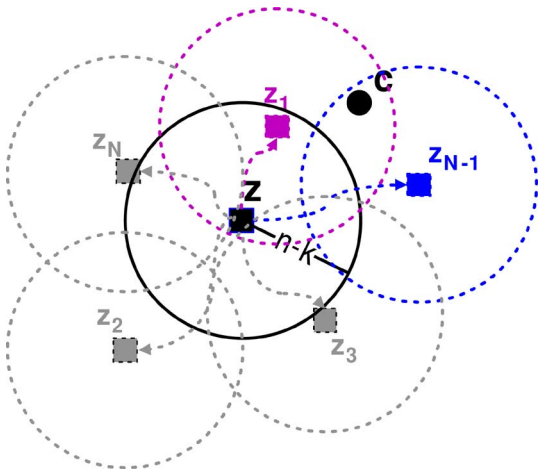

Fig. 2. Virtual received words are generated around the received LLR vector $\bar{\Gamma}$ by hard-limiting the sample vectors generated by an importance probability density whose parameter values evolved according to the CE principle.

ten iterations in most cases. Our algorithm yields good performance although uniform convergence in distribution or η within a limited iterations is not guaranteed.

IV. LIST DECODING WITH VIRTUAL RECEIVED WORDS

In Step 2 of Table I we try to find the most likely message/error coordinates such that the associated EO-decoded codeword is closest to the received vector. Note that the random samples are used to determine the erasure locations only, and the searching sphere of the algebraic list decoding described in Section III-A is always centered at the hard-limited received word $\mathbf{z}$ with radius equals to $n-k$. To increase our search range and improve decoding performance, we include some extra codewords which lie statistically in a small neighborhood of the received word in our expanded search, such that some of them may in fact be closer in ED to the true transmitted codeword $\mathbf{c}$; see Fig. 2. More specifically, the expansion is accomplished by eliminating the requirement that the search be centered at $\mathbf{z}$. Instead, we randomized the search center by EO-decoding the hard-decision versions of the drawn sample vectors which we refer to as virtual received words. If the importance density does converge to the desired density $\delta(\mathbf{s}-\mathbf{s}^*)$, such an expanded search will eventually contract and converge to the true transmitted codeword.

A. Importance Density and Sample Format

Let $\bar{\mathbf{s}}=(\bar{s}_0,\ldots,\bar{s}_{nm-1})$ be a random vector where $\bar{s}_0,\ldots,\bar{s}_{nm-1}$ are independent Gaussian random variables with means $\mu_0,\ldots,\mu_{nm-1}$ and variances $\sigma_0^2,\ldots,\sigma_{nm-1}^2$. We write $\bar{\mathbf{s}}\sim\mathcal{N}(\vec{\mu},\vec{\sigma})$, where $\vec{\mu}=(\mu_0,\ldots,\mu_{nm-1})$ and $\vec{\sigma}=(\sigma_0,\ldots,\sigma_{nm-1})$ are initialized by

$$\mu_j^{(0)}=\bar{\gamma}_j \tag{7}$$

$$\sigma_j{}^{(0)}=\sqrt{|\bar{\gamma}_j|}. \tag{8}$$

At the tth iteration, N random samples $\bar{\mathbf{s}}_1^{(t)},\bar{\mathbf{s}}_2^{(t)},\ldots,\bar{\mathbf{s}}_N^{(t)}$ are drawn from $\mathcal{N}\left(\vec{\mu}^{(t)},\vec{\sigma}^{(t)}\right)$ to form the sample set $\bar{\mathbf{S}}^{(t)}$. Each sample vector represents the bit reliabilities of an associated virtual received word. By using (4) to convert the bit reliabilities into symbol reliability, the $n-k$ coordinates with smallest symbol reliabilities are erased; the remaining bit positions are hard-limited, mapped into symbol decisions and the resulting virtual received word is then decoded by an EO decoder.

B. Update Parameters

Let $\mathbf{d}_1^{(t)},\ldots,\mathbf{d}_N^{(t)}$ be the output codewords of the EO decoder. We compute the ED between each candidate codeword and the received word $\mathbf{y}$ and sort the corresponding random vectors according to the descending order of their associated EDs. Define an elite set $\bar{\mathbf{S}}_E^{(t)}$ which includes E vectors with the smallest EDs to $\mathbf{y}$, i.e., the corresponding codewords are more likely to have been transmitted. We always store the best one in $\bar{\mathbf{S}}_E^{(t)}$ up to the current iteration for the final decision when the maximum number of iteration is reached.

Then the two sets of parameters $\vec{\mu}^{(t+1)}$ and $\vec{\sigma}^{(t+1)}$ are updated by [9]

$$\mu_j^{(t+1)}=(1-\lambda)\mu_j^{(t)}+\lambda\cdot\frac{\sum_{\bar{\mathbf{s}}_\ell^{(t)}\in\bar{\mathbf{S}}_E^{(t)}}\bar{s}_{\ell,j}^{(t)}}{E} \tag{9}$$

and

$$\sigma_j^{(t+1)}=(1-\eta)\sigma_j^{(t)}+\eta\cdot\frac{\sum_{\bar{\mathbf{s}}_\ell^{(t)}\in\bar{\mathbf{S}}_E^{(t)}}\left(\bar{s}_{\ell,j}^{(t)}-\mu_j^{(t+1)}\right)^2}{E} \tag{10}$$

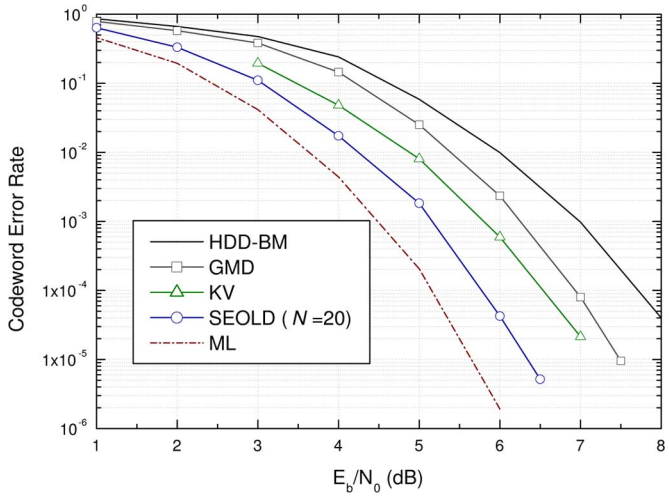

Fig. 3. Codeword error probability performance of the (15,11) Reed-Solomon code; ten iterations.

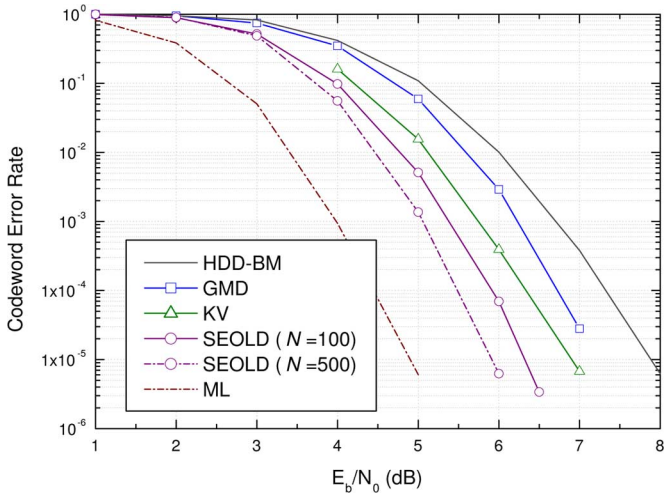

Fig. 4. Codeword error probability performance of the (31,25) Reed-Solomon code; ten iterations.

where λ and η are real values between (0,1) used to smooth the variation of these parameters. The algorithm described in this section is called the stochastic erasures-only listing decoding (SEOLD) algorithm.

V. Simulation Results and Discussion

In this section, some simulated performance of the SEOLD algorithm is presented and compared with that of other well known RS decoding algorithms. A standard binary input AWGN channel is assumed over which the BPSK modulated codewords are transmitted. We model the receive matched filter output as the sum of a ± 1- valued sequence and Gaussian sequence with zero-mean i.i.d. components. The average performance bounds on the ML error probability of RS codes over an AWGN channel developed in [11] are used as the performance lower limits.

Due to the complexity and the decoding delay considerations, the SEOLD algorithm will not terminate until convergence is assured. Instead, we limit our decoding procedure to T iterations in all simulations.

Figs. 3 and 4 show respectively the codeword error rate (CER) performance of the (15,11) and (31,25) RS codes over an AWGN channel. HDD-BM refers to the performance of a hard decision bounded minimum distance decoder such as the BM algorithm. Curves labelled by GMD and KV are the performance obtained by using Forney's GMD algorithm and the algebraic soft decision decoding algorithm of Koetter and Vardy, respectively. Note that the performance of the KV algorithm shown here represents the lower bound obtained by using an infinite interpolation multiplicity. For the case of (15,11) RS codes, the size of the sample set N and the size of the elite set E at every iteration are set to be 20 and 6, respectively. With $T = 10$ iterations, the SEOLD algorithm obtains 1.2 dB and 1.0 dB coding gains with respect to the GMD and the KV algorithms at CER $= 10^{-5}$. For the case of (31,25) RS codes, the SEOLD algorithm also outperform the other two algorithms with reasonable complexity of $N = 100$, $E = 10$ and $N = 500$, $E = 50$ in 10 iterations. The SEOLD algorithm has about 0.6 dB and 1.0 dB coding gains over the KV algorithm when N is equal to 100 and 500, respectively. In conclusion, the proposed decoding algorithm is capable of offering good performance with modest complexity for short high rate RS codes. Its performance can be further improved by increasing the sample size N and/or the maximum iteration number T at the cost of increased decoding complexity.

References

[1] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Ind. Appl. Math*, vol. 8, pp. 300–304, 1960.

[2] E. B. Berlekamp, *Algebraic Coding Theory*. New York: McGraw-Hill, 1968.

[3] Y. Sugiyama, M. kasahara, S. Hirasawa, and T. Namekawa, "A method for solving key equation for decoding goppa codes," *Inform. Contr.*, vol. 27, pp. 87–99, 1975.

[4] V. Guruswami and M. Sudan, "Decoding of Reed-Solomon codes beyond the error-correction bound," *J. Complexity*, vol. 13, pp. 180–193, 1997.

[5] G. D. Forney, Jr, "Generalized minimum distance decoding," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 125–131, Apr. 1966.

[6] D. Chase, "A class of algorithms for decoding block codes with channel measurement information," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 170–182, Jan. 1972.

[7] H. Tang, Y. Liu, M. Fosorier, and S. Lin, "On combining chase-2 and GMD decoding algorithms for nonbinary block codes," *IEEE Commun. Lett.*, vol. 5, pp. 209–211, May 2001.

[8] R. Köetter and A. Vardy, "Algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2809–2825, Nov. 2003.

[9] R. Rubinstein and D. P. Kroese, *The Cross-Entropy Method. A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. Berlin, Germany: Springer, Information Science and Statistics, , 2004.

[10] F. Dambreville, Cross-Entropy Method: Convergence Issues for Extended Implementation [Online]. Available: http://www.Frederic-Dambreville.com

[11] M. EI-Khamy and R. J. McEliece, "Bounds on the average binary minimum distance and the maximum likelihood performance of Reed-Solomon codes," in *42nd Allerton Conf. Communication, Control, and Computing*, 2004.