A rate-distortion exponent approach to multiple decoding attempts for Reed-Solomon codes

Phong S. Nguyen
(joint work with H. D. Pfister & K. R. Narayanan)

Department of Electrical & Computer Engineering
Texas A&M University

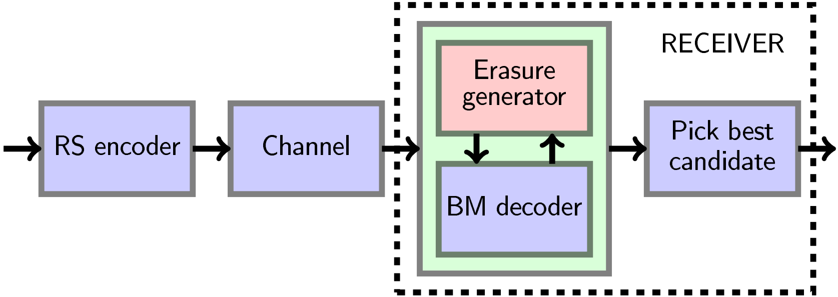
The IEEE International Symposium on Information Theory
June 15th, 2010

Reed-Solomon (RS) Codes

- RS codes are one of the most widely used error-correcting codes.
- (N, K) RS codes have rate $\frac{K}{N}$ and minimum distance $d_{min} = N - K + 1$.
- Conventional hard-decision (HD) decoding algorithms such as Berlekamp-Massey (BM) can correct up to $d_{min}/2$ errors.
- Researchers have put considerable effort into improving decoding.
  - Guruswami-Sudan (GS) - can correct errors well beyond $d_{min}/2$.
  - Koetter-Vardy (KV) - algebraic soft-decision decoding (ASD) uses soft-information to improve this.
- Both algorithms, however, are quite computationally complex.

Multiple Errors-and-Erasures Decoding (1)

- Recently, approaches based on multiple decoding attempts have attracted new interest.
- For example, consider multiple runs of errors-and-erasures (BM) decoding, each time with a different set of erasure patterns (the least reliable symbols are usually erased).



Multiple Errors-and-Erasures Decoding (2)

- Use the following definition to represent erasure patterns and error patterns.

Definition

- $\hat{x}^N \in \{0,1\}^N$ as an erasure pattern: at index i,

$$\hat{x}_i = \begin{cases} 0 & \text{if symbol is erased} \\ 1 & \text{if symbol is not erased} \end{cases}$$

- $x^N \in \{0,1\}^N$ as an error pattern: at index i,

$$x_i = \begin{cases} 0 & \text{if error occurs} \\ 1 & \text{if error does not occur} \end{cases}$$

Example Sets of Erasure Patterns

- Sort the codeword positions in increasing reliability order.

Error pattern
001011 . . .

- Generalized Minimum Distance (GMD) decoding [Forney]

Repeat error-and-erasure decoding while successively erasing an even number of the least reliable positions (LRPs).

Erasure patterns of GMD
111111 . . .
001111 . . .
000011 . . .

- Successive Error-and-erasure Decoding (SED) [Lee-Kumar]

SED(l, f) repeats error-and-erasure decoding with every combination of an even number $\leq f$ of erasures within the l LRPs.

Erasure patterns of SED(3,2)
111111 . . .
001111 . . .
010111 . . .
100111 . . .

- **Motivation:** How can one construct the "best" set of erasure patterns (in the sense of performance vs complexity)?

Example Sets of Erasure Patterns

- Sort the codeword positions in increasing reliability order.

Error pattern
001011...

- Generalized Minimum Distance (GMD) decoding [Forney]

Repeat error-and-erasure decoding while successively erasing an even number of the least reliable positions (LRPs).

Erasure patterns of GMD
111111...
001111...
000011...

- Successive Error-and-erasure Decoding (SED) [Lee-Kumar]

SED(l, f) repeats error-and-erasure decoding with every combination of an even number $\leq f$ of erasures within the l LRPs.

Erasure patterns of SED(3,2)
111111...
001111...
010111...
100111...

- **Motivation:** How can one construct the "best" set of erasure patterns (in the sense of performance vs complexity)?

Example Sets of Erasure Patterns

- Sort the codeword positions in increasing reliability order.

Error pattern

001011...

- Generalized Minimum Distance (GMD) decoding [Forney]

Repeat error-and-erasure decoding while successively erasing an even number of the least reliable positions (LRPs).

Erasure patterns of GMD

111111...
001111...
000011...

- Successive Error-and-erasure Decoding (SED) [Lee-Kumar]

SED(l, f) repeats error-and-erasure decoding with every combination of an even number $\leq f$ of erasures within the l LRPs.

Erasure patterns of SED$(3,2)$

111111...
001111...
010111...
100111...

- **Motivation:** How can one construct the "best" set of erasure patterns (in the sense of performance vs complexity)?

Example Sets of Erasure Patterns

- Sort the codeword positions in increasing reliability order.

Error pattern
001011 . . .

- Generalized Minimum Distance (GMD) decoding [Forney]

Repeat error-and-erasure decoding while successively erasing an even number of the least reliable positions (LRPs).

Erasure patterns of GMD
111111 . . .
001111 . . .
000011 . . .

- Successive Error-and-erasure Decoding (SED) [Lee-Kumar]

SED(l, f) repeats error-and-erasure decoding with every combination of an even number $\leq f$ of erasures within the l LRPs.

Erasure patterns of SED(3, 2)
111111 . . .
001111 . . .
010111 . . .
100111 . . .

- **Motivation:** How can one construct the "best" set of erasure patterns (in the sense of performance vs complexity)?

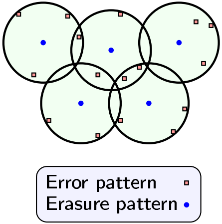
Connection to Rate-Distortion Theory

Single errors-and-erasures decoding threshold

Consider an (N, K) RS code. If e symbols are erased, the BM algorithm can correct v errors in unerased positions if $2v + e < N - K + 1$.

- A multiple errors-and-erasures decoding succeeds if the decoding threshold is satisfied during any decoding attempt.

- **Idea** [Nguyen et. al., Allerton '09]: Connect multiple decoding to a R-D (covering) problem where the decoder succeeds if the error pattern ■ is covered by any decoding ball centered at an erasure pattern •



Distortion Measure

Definition

Given a *letter-by-letter* distortion measure δ, the distortion between and error pattern x^N and an erasure pattern $\hat{x}^N$ is $d(x^N, \hat{x}^N) = \sum_{i=1}^{N} \delta(x_i, \hat{x}_i)$.

Proposition

If we choose $\delta : \{0,1\} \times \{0,1\} \to \mathbb{R}_{\geq 0}$ as follows

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

Error	Erasure Yes (0)	Erasure No (1)
Yes (0)	1	2
No (1)	1	0

then $2v + e < N - K + 1$ reduces to the form $d(x^N, \hat{x}^N) < N - K + 1$.

- **Extension:** Can be applied to other decoding schemes (e.g. ASD decoding, generalized errors-and-erasures decoding).

Top-ℓ Generalized Errors-and-Erasured Decoding.

Definition

Consider a positive integer ℓ.

- $x^N \in \{0,1,\dots,\ell\}^N$ is a generalized error pattern, and at index i,

$$x_i = \begin{cases} 0, & \text{none of the first } \ell \text{ most likely symbols is correct} \\ j, & \text{if the } j\text{-th most likely symbol is correct } (j = 1,2,\dots,\ell) \end{cases}$$

- $\hat{x}^N \in \{0,1,\dots,\ell\}^N$ is an generalized erasure pattern, and at index i,

$$\hat{x}_i = \begin{cases} 0, & \text{if an easure is used} \\ j, & \text{if the } j\text{-th most likely symbol is used as HD } (j = 1,2,\dots,\ell) \end{cases}$$

Theorem

We choose $\delta : \{0,1,\dots,\ell\} \times \{0,1,\dots,\ell\} \to \mathbb{R}_{\geq 0}$ defined by $\delta(x,\hat{x}) = [\Delta]_{x,\hat{x}}$ in terms of

$$\Delta = \begin{pmatrix} 1 & 2 & \dots & 2 & 2 \\ 1 & 0 & \dots & 2 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & \dots & 0 & 2 \\ 1 & 2 & \dots & 2 & 0 \end{pmatrix} \Longrightarrow$$

decoding threshold becomes $d(x^N, \hat{x}^N) < N - K + 1$.

Choosing the Erasure Patterns

- **Problem statement**: Build a set (codebook) $\mathcal{B}_R$ of no more than 2^R erasure patterns $\hat{x}^N$ in order to

$$\max_{\mathcal{B}_R : |\mathcal{B}_R| \le 2^R} \Pr\left(\min_{\hat{x}^N \in \mathcal{B}_R} d(x^N, \hat{x}^N) < \underbrace{N - K + 1}_{D} \right).$$

- **RD approach** [Nguyen et. al., Allerton '09] (for $R = N\overline{R}$ and $D = N\overline{D}$):
  - View the error pattern x^N as a source sequence and the erasure pattern $\hat{x}^N$ as a reproduction sequence
  - R-D theory: Trade-off between $\overline{R}$ and $\overline{D}$ such that sets $\mathcal{B}_{N\overline{R}}$ of $2^{N\overline{R}}$ reproduction (erasure) sequences can be generated randomly so that

$$\lim_{n\to\infty} \frac{1}{N} E\left[\min_{\hat{x}^N \in \mathcal{B}_{N\bar{R}}} d(x^N, \hat{x}^N) \right] < \bar{D}$$

  - Then, for N fixed but large enough, it is very likely that

$$\min_{\hat{x}^N \in \mathcal{B}_R} d(x^N, \hat{x}^N) < D$$

Choosing the Erasure Patterns

- **Problem statement**: Build a set (codebook) $\mathcal{B}_R$ of no more than 2^R erasure patterns $\hat{x}^N$ in order to

$$\max_{\mathcal{B}_R:|\mathcal{B}_R|\leq 2^R} \Pr\left(\min_{\hat{x}^N\in\mathcal{B}_R} d(x^N,\hat{x}^N) < \underbrace{N-K+1}_{D}\right).$$

- **RD approach** [Nguyen et. al., Allerton '09] (for $R = N\overline{R}$ and $D = N\overline{D}$):
  - View the error pattern x^N as a source sequence and the erasure pattern $\hat{x}^N$ as a reproduction sequence
  - R-D theory: Trade-off between $\overline{R}$ and $\overline{D}$ such that sets $\mathcal{B}_{N\overline{R}}$ of $2^{N\overline{R}}$ reproduction (erasure) sequences can be generated randomly so that

$$\lim_{n\to\infty} \frac{1}{N} E\left[\min_{\hat{x}^{\boldsymbol{N}}\in\mathcal{B}_{\boldsymbol{N}\bar{\boldsymbol{R}}}} d(x^N,\hat{x}^N)\right] < \bar{D}$$

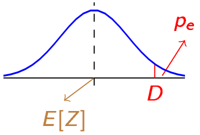
  - Then, for N fixed but large enough, it is very likely that

$$\min_{\hat{x}^{\boldsymbol{N}}\in\mathcal{B}_{\boldsymbol{R}}} d(x^N,\hat{x}^N) < D$$

Rate-Distortion Exponent (RDE) approach

For random codes, let $Z = \min_{\hat{x}^N \in \mathcal{B}_R} d(\hat{x}^N, x^N)$ be a random variable.

- RD approach: minimize the average minimum distortion $E[Z]$
- RDE approach: maximize the exponent $\overline{F}$ at which $p_e = \Pr(Z > D)$ decays with N.



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

RS decoding has a fixed threshold $D = N - K + 1$, so there is an $\overline{F}$ vs R trade-off.

Advantages of RDE approach

- Gives an estimate of $p_e \approx 2^{-N\overline{F}(R,D)}$ for each R where 2^R is the number of decoding attempts.
- This is the near-optimal choice of $\mathcal{B}_R$ (the set of erasure patterns) since
  - Upperbound: For the set $\mathcal{B}_R$ using RDE approach, for every $\varepsilon > 0$,

    $$p_e \leq 2^{-N[\overline{F}(R,D)-\varepsilon]}$$

    for N large enough.
  - Lowerbound: Consider an arbitrary set $\mathcal{B}_R$, for any $\varepsilon > 0$,

    $$p_e \geq 2^{-N[\overline{F}(R,D)+\varepsilon]}$$

    for N large enough

Proposed Multiple-Decoding Algorithm (1)

- Step 1: Transmit codeword c_1^N (where $c_i \in \mathbb{F}_m$) and receive r_1^N
- Step 2: Compute the $m \times N$ reliability matrix $[\Pi]_{j,i} = \Pr(c_i = \alpha_j | r_i)$.

Example: Consider the (7,5) RS code over $\mathbb{F}_8$ with $d_{min} = 3$.
Suppose the all-zero codeword was transmitted.

$$\Pi = \begin{array}{c} \begin{matrix} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 \end{matrix} \\ \begin{bmatrix} .95 & .37 & .88 & .79 & .95 & .24 & .97 \\ .01 & .01 & .09 & .01 & .01 & .19 & .02 \\ .01 & 0 & 0 & .09 & 0 & .10 & .01 \\ 0 & .62 & 0 & 0 & 0 & 0 & 0 \\ .03 & 0 & .03 & .11 & .04 & .47 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \alpha_1 = 0 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \\ \alpha_6 \\ \alpha_7 \\ \alpha_8 \end{matrix}$$

Proposed Multiple-Decoding Algorithm (2)

- Step 3: Process Π, keeping top-ℓ symbols and a "not top-ℓ" symbol to get **P** where $p_{i,j} = \Pr(x_i = j)$.

Example: Consider $\ell = 2$.

$$\mathbf{P} = \begin{bmatrix} .02 & .01 & .03 & .10 & .01 & .29 & .01 \\ .95 & .62 & .88 & .79 & .95 & .47 & .97 \\ .03 & .37 & .09 & .11 & .04 & .24 & .02 \end{bmatrix} \begin{matrix} \Pr(x_i = 0) \\ \Pr(x_i = 1) \\ \Pr(x_i = 2) \end{matrix}$$

with columns $x_1, x_2, x_3, x_4, x_5, x_6, x_7$.

Distortion measure matrix for *multiple BM top-2* (mBM-2) decoding is

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

Proposed Multiple-Decoding Algorithm (3)

- Step 4: Compute the RDE function for the processed reliabilities and the distortion measure for top-ℓ decoding. For the desired rate-distortion (R, D) pair, find the corresponding test-channel input-distribution $\mathbf{Q}$.

  Example: Consider $R = 2$ and $D = N - K + 1 = 3$, we have

$$F(R, D) = 1.415 > 0$$

$$Q = \begin{array}{c} \\ \end{array}\begin{bmatrix} \hat{x}_1 & \hat{x}_2 & \hat{x}_3 & \hat{x}_4 & \hat{x}_5 & \hat{x}_6 & \hat{x}_7 \\ 0 & 0 & 0 & .28 & 0 & .63 & 0 \\ .93 & .58 & .82 & .61 & .91 & .26 & .96 \\ .67 & .42 & .18 & .11 & .09 & .11 & .04 \end{bmatrix} \begin{array}{ll} \\ \Pr(\hat{x}_i = 0) & \text{erase} \\ \Pr(\hat{x}_i = 1) & \text{use 1st-ML symbol} \\ \Pr(\hat{x}_i = 2) & \text{use 2nd-ML symbol} \end{array}$$

Proposed Multiple-Decoding Algorithm (4)

- Step 5: Generate a set of 2^R candidate patterns by randomly sampling the test-channel input-probability distribution $\mathbf{Q}$.

  Example: $R = 2$ yields the set (codebook) of 4 erasure patterns

  $$\mathcal{B} = \{1111111, 1210111, 1221111, 1111121\}.$$

- Step 6: Run the corresponding decoding scheme for each candidate pattern to produce a list of candidate codewords. Use the ML rule to choose the best codeword on the list.

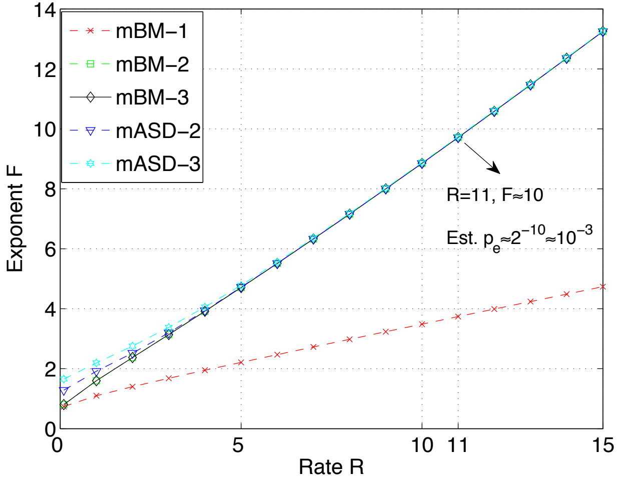
A Range of Algorithms

- Berlekamp-Massey (BM) Based Algorithms
  - mBM-ℓ = Multiple runs on Berlekamp-Massey using top-ℓ decoding
  - mBM-1 = Either use hard decision or erasure
  - mBM-2 = Either use 1st-ML symbol, 2nd ML-symbol, or erasure
    - Related to symbol-flipping Chase decoding [Kavcic-Bellorado]
- Algebraic Soft-Decision (ASD) Based Algorithms
  - Columns of the multiplicity matrix are chosen i.i.d. according to an RDE optimized distribution
  - mASD-μ = Sum of multiplicity in each column not greater than μ
  - mASD-1 = Set of multiplicity types = (1,0),(0,1),(0,0)
  - mASD-2 = Set of multiplicity types = (2,0),(1,1),(0,2),(0,0)

Some Special Cases of RDE Approach

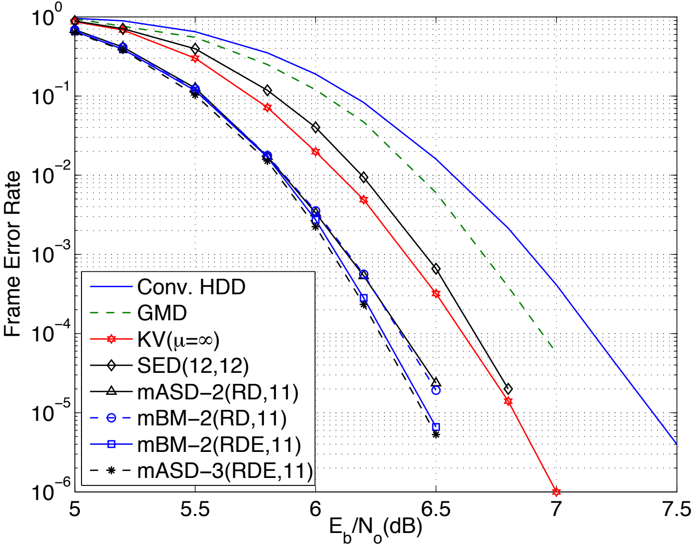
- When $F = 0$, the RDE approach becomes the RD approach.
- When $R = 0$, the codebook has only one entry.
  - The random codebook becomes deterministic (i.e., test-channel input-distribution $\mathbf{Q}$ consists of only entries of only 0 and 1).
  - Related to the following lines of work considered by other researchers.
    - Design an erasure pattern for a single BM decoding.
    - Design a multiplicity matrix for a single ASD decoding.

Rate-Distortion Exponent Function



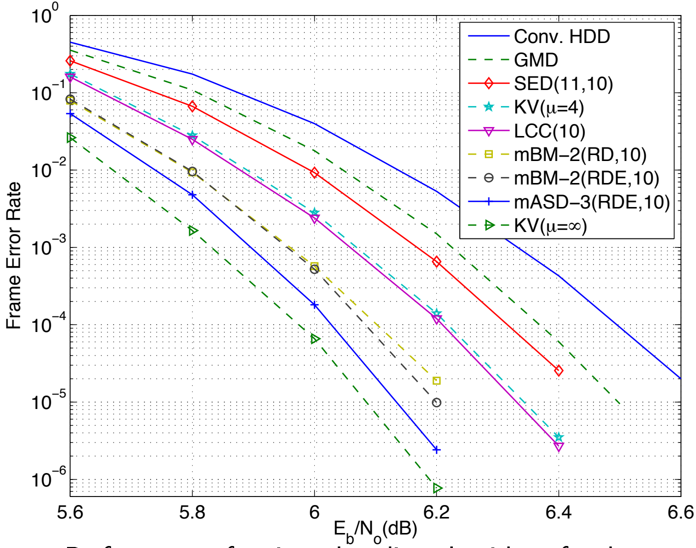
The RDE curves at $E_b/N_0 = 6$dB for the (255,239) RS code over $\mathbb{F}_{2^8}$ with BPSK over an AWGN channel.

Simulation Results for RS(255,239)



Performance of various decoding algorithms for the (255,239) RS code over $\mathbb{F}_{2^8}$ with BPSK over an AWGN channel.

Simulation Results for RS(458,410)



Performance of various decoding algorithms for the (458,410) RS code over $\mathbb{F}_{2^{10}}$ with BPSK over an AWGN channel.

Summary and Open Questions

- A RDE-based approach is proposed for multiple decoding attempts of RS codes.
- Can be applied to a wide range of decoders and error models.
- Open Questions:
  - Can we solve the decoding problem for each candidate more easily by using previous computations (c.f., Kavcic-Bellorado)?
  - Can we extend our framework to intersymbol interference channels?

THANK YOU!