

ADVANCED CODING TECHNIQUES  
WITH APPLICATIONS TO STORAGE SYSTEMS

A Dissertation

by

PHONG SY NGUYEN

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2012

Major Subject: Electrical Engineering

ADVANCED CODING TECHNIQUES  
WITH APPLICATIONS TO STORAGE SYSTEMS

A Dissertation

by

PHONG SY NGUYEN

Submitted to the Office of Graduate Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Henry D. Pfister Krishna R. Narayanan
Committee Members,	A. L. Narasimha Reddy Frank Sottile
Head of Department,	Costas N. Georghiades

May 2012

Major Subject: Electrical Engineering

## ABSTRACT

Advanced Coding Techniques with Applications to Storage Systems. (May 2012)

Phong S. Nguyen, B. Eng., Hanoi University of Technology

Co-Chairs of Advisory Committee: Dr. Henry D. Pfister  
Dr. Krishna R. Narayanan

This dissertation considers several coding techniques based on Reed-Solomon (RS) and low-density parity-check (LDPC) codes. These two prominent families of error-correcting codes have attracted a great amount of interest from both theorists and practitioners and have been applied in many communication scenarios. In particular, data storage systems have greatly benefited from these codes in improving the reliability of the storage media.

The first part of this dissertation presents a unified framework based on rate-distortion (RD) theory to analyze and optimize multiple decoding trials of RS codes. Finding the best set of candidate decoding patterns is shown to be equivalent to a covering problem which can be solved asymptotically by RD theory. The proposed approach helps understand the asymptotic performance-versus-complexity trade-off of these multiple-attempt decoding algorithms and can be applied to a wide range of decoders and error models.

In the second part, we consider spatially-coupled (SC) codes, or terminated LDPC convolutional codes, over intersymbol-interference (ISI) channels under joint iterative decoding. We empirically observe the phenomenon of threshold saturation whereby the belief-propagation (BP) threshold of the SC ensemble is improved to the maximum a posteriori (MAP) threshold of the underlying ensemble. More specifically, we derive a generalized extrinsic information transfer (GEXIT) curve for the

joint decoder that naturally obeys the area theorem and estimate the MAP and BP thresholds. We also conjecture that SC codes due to threshold saturation can universally approach the symmetric information rate of ISI channels.

In the third part, a similar analysis is used to analyze the MAP thresholds of LDPC codes for several multiuser systems, namely a noisy Slepian-Wolf problem and a multiple access channel with erasures. We provide rigorous analysis and derive upper bounds on the MAP thresholds which are shown to be tight in some cases. This analysis is a first step towards proving threshold saturation for these systems which would imply SC codes with joint BP decoding can universally approach the entire capacity region of the corresponding systems.

To my parents (Kính tặng ba mẹ)!

## ACKNOWLEDGMENTS

There is a saying in my mother language “Không thầy đố mày làm nên” (Without the teacher one cannot succeed) which I find perfectly applicable for my graduate studies. First and foremost, I would like to express my deepest respect and gratitude to my advisors, Prof. Henry Pfister and Prof. Krishna Narayanan, for their wonderful guidance and support over the years. Their deep insight, vast knowledge, and talent of seeing the big picture have been crucial for me in performing this research. More importantly, their great personalities set an admirable example which I hope I have absorbed part of.

I also want to take this opportunity to thank Prof. Narasimha Reddy and Prof. Frank Sottile for serving on my dissertation committee and being supportive throughout my doctoral studies. Their insightful comments and suggestions have helped me improve the quality of my research.

I am grateful to my labmates for the intellectual and stimulating environment that I have enjoyed. In particular, I thank Arvind for the fruitful collaboration we had on spatially-coupled codes and for introducing *TikZ* to our lab, using which most figures in this dissertation are made; Fan for keeping me updated with the trends in coding theory when I first joined the lab and for his generous job search advice; Hak for the fun and experience we shared, and for our future collaboration at Marvell; Yung-Yih for the mutual encouragement of each other when ideas were low, and for frequent cultural discussions.

Sincere thanks go to other colleagues in Dr. Pfister’s and Dr. Narayanan’s groups whom I had chances to interact with. Among them are Andrew, Aparna, Brett, Chia-Wen, Engin, Fatemeh, Jerry, Makesh, Santhosh, Sirish, and others. I would also like to thank all the professors and students in the TCSP group for organizing inspiring

weekly meetings, especially anh Hùng for our technical discussions in tiếng Việt. I send my warm thanks the ECE staff for their excellent administration support, particularly Tammy, Paula, Gayle, Claudia, Linda, and Anni.

I also want to thank my Vietnamese friends in Bryan and College Station for all the wonderful events we had. Thanks to the many pickup soccer players in town who shared the same hobby with me every Saturday and made my life outside research more enjoyable and fulfilled.

I gratefully acknowledge the financial support from Vietnam Education Foundation, the National Science Foundation under Grants 0747470 and 0802124, and Seagate through the NSF GOALI Program.

Finally, my warmest gratitude goes to my loving family. I would like to thank my parents and parents-in-law for their constant support and encouragement, my sister and brother-in-law for taking care of my parents while I am away. The ultimate words of recognition go to my dear wife for her unconditional love and sharing. My wife, Thu Anh, my son, Thành Vinh, and my daughter, Khánh An, have become my source of inspiration and happiness.

## TABLE OF CONTENTS

CHAPTER	Page
I	INTRODUCTION . . . . . 1
	A. Background on Reed-Solomon Codes . . . . . 3
	B. Background on LDPC Codes . . . . . 3
	1. Spatially-Coupled Codes . . . . . 5
	a. The $(d_l, d_r, L)$ Ensemble . . . . . 6
	b. The $(d_l, d_r, L, w)$ Ensemble . . . . . 7
	C. Dissertation Outline . . . . . 7
	D. Notation . . . . . 8
II	A RATE-DISTORTION APPROACH TO MULTIPLE DE- CODING ATTEMPTS OF REED-SOLOMON CODES . . . . . 9
	A. Introduction . . . . . 9
	B. A RD Framework for Multiple Errors-and-Erasures Decoding 12
	1. Conventional Error Patterns and Erasure Patterns. . . . 14
	a. RD Approach . . . . . 16
	b. RDE Approach . . . . . 17
	2. Generalized Error Patterns and Erasure Patterns . . . . 18
	3. Proposed General Multiple-Decoding Algorithm . . . . 20
	a. Algorithm A . . . . . 21
	b. Algorithm B . . . . . 22
	C. Computing the RD and RDE Functions . . . . . 23
	1. Computing the RD Function . . . . . 23
	2. Computing the RDE Function . . . . . 25
	3. Complexity of Computing RD/RDE Functions . . . . . 27
	a. Complexity of Computing RD Function. . . . . 27
	b. Complexity of Computing RDE Function. . . . . 28
	D. Multiple Algebraic Soft-Decision (ASD) Decoding . . . . . 29
	1. Bit-level ASD Case . . . . . 31
	2. Symbol-level ASD Case . . . . . 32
	a. High-Rate Reed-Solomon Codes . . . . . 41
	b. Lower-Rate Reed-Solomon Codes . . . . . 45
	E. Closed-Form Analysis of RD and RDE Functions for Some Distortion Measures . . . . . 51
	1. Closed-Form RD Function . . . . . 51
	2. Closed-form RDE function . . . . . 53

CHAPTER	Page
F. Some Extensions . . . . .	56
1. Erasure Patterns Using Covering Codes . . . . .	56
2. A Single Decoding Attempt . . . . .	59
G. Simulation Results . . . . .	62
H. Appendix . . . . .	71
1. Proof of Corollary 2 . . . . .	71
2. Proof of Lemma 6 . . . . .	73
3. Proof of Theorem 3 . . . . .	75
4. Analysis of RDE Computation . . . . .	77
5. Faster Algorithm to Compute RD Function for m-bASD . . . . .	79
 III	
SPATIALLY-COUPLED CODES AND THRESHOLD SAT- URATION ON INTERSYMBOL-INTERFERENCE CHANNELS . . . . .	84
A. Introduction . . . . .	84
B. Background . . . . .	86
1. ISI Channels and the SIR . . . . .	86
2. LDPC Ensembles and the Joint BP Decoder . . . . .	88
C. ISI Channels with Erasure Noise: The GECs . . . . .	90
1. BP and EBP Curves for the GEC . . . . .	90
2. Upper Bound on the MAP Threshold . . . . .	101
3. Tightness of the Upper Bound . . . . .	103
4. Spatially-Coupled Codes for the GEC . . . . .	110
D. General ISI Channels . . . . .	113
1. GEXIT Curves for the ISI channels . . . . .	113
a. BP-GEXIT Curve (with AWGN) . . . . .	118
2. Upper Bound on the MAP Threshold . . . . .	120
3. Spatially-Coupled Codes on General ISI Channels . . . . .	122
4. Simulation Results . . . . .	124
 IV	
ON THE MAP THRESHOLD OF MULTIUSER SYSTEMS WITH ERASURES . . . . .	126
A. Introduction . . . . .	126
1. Preliminaries . . . . .	127
B. Slepian-Wolf Problem with Erasures . . . . .	128
1. Channel Model . . . . .	128
2. EXIT Functions . . . . .	130
3. MAP Threshold . . . . .	132
a. Upper Bound on the MAP Threshold . . . . .	132

CHAPTER	Page
b. Tightness of the Upper Bound . . . . .	135
C. Multiple Access Channel with Erasures . . . . .	138
1. Channel Model . . . . .	138
2. EXIT Functions . . . . .	140
3. MAP Threshold . . . . .	142
a. Upper Bound on the MAP Threshold . . . . .	142
b. Tightness of the Upper Bound . . . . .	146
D. Threshold Saturation of Spatially-Coupled Codes . . . . .	148
V CONCLUSIONS AND FUTURE WORK . . . . .	150
A. A Rate-Distortion Framework to Analyze and Design Multiple Decoding Attempts of Reed-Solomon Codes . . . . .	150
1. Summary . . . . .	150
2. Future Work . . . . .	151
B. Applications of Spatially-Coupled Codes via Threshold Saturation . . . . .	152
1. Summary . . . . .	152
2. Future Work . . . . .	153
REFERENCES . . . . .	154
VITA . . . . .	166

## LIST OF TABLES

TABLE		Page
I	Example ranges of possible $a$ . . . . .	45
II	Example ranges of $a$ that gives the largest exponent . . . . .	47
III	Thresholds of $(d_l, d_r)$ -regular ensembles over the DEC, pDEC and PR2EC. . . . .	103
IV	Threshold estimates, measured in dB, of $(d_l, d_r)$ -regular ensembles over the dicode AWGN and PR2 AWGN channels. . . . .	122

## LIST OF FIGURES

FIGURE	Page
1	(Left) The protograph for the $(d_l, d_r, L)$ ensemble where $d_l = 3$ and $d_r = 6$ . (Right) The parity-check matrix associated with the protograph on the left before lifting. . . . . 6
2	Pictorial illustration of a covering problem . . . . . 16
3	Plot of exponent $F_a$ versus $a$ for $\mu = 2$ and $\mu = 3$ with a fixed rate $R = 6$ . Simulations are conducted for the (255,127) RS code using BPSK over an AWGN channel at $E_b/N_0 = 6.0$ dB and 6.5 dB. . . . . 48
4	Plot of exponent $F_a$ versus $a$ for $\mu = 10$ with a fixed rate $R = 6$ . The set of multiplicity types considered is the relaxed set $\mathcal{A}_0(10, 2)$ . Simulations are conducted for the (458,410) RS code over $\mathbb{F}_{2^{10}}$ using BPSK over an AWGN channel at $E_b/N_0 = 6.0$ dB and 6.5 dB. . . . . 50
5	Performance of mBM-1(RDE,11) and its approximation $2^{-F}$ where $F$ is given in (2.47) for the (255,239) RS code over an $m$ -SC( $p$ ) channel. . . . . 56
6	A realization of RD curves at $E_b/N_0 = 5.2$ dB for various decoding algorithms for the (255,239) RS code over an AWGN channel. . . . . 62
7	A realization of RDE curves at $E_b/N_0 = 6$ dB for various decoding algorithms for the (255,239) RS code over an AWGN channel. . . . . 63
8	Performance of various decoding algorithms for the (255,239) RS code using BPSK over an AWGN channel. . . . . 64
9	Performance of various decoding algorithms for the (255,239) RS code using 256-QAM over an AWGN channel. . . . . 66
10	Performance of various decoding algorithms for the (458,410) RS code over $\mathbb{F}_{2^{10}}$ using BPSK over an AWGN channel. . . . . 67
11	Performance of various decoding algorithms for the (458,410) RS code over $\mathbb{F}_{2^{10}}$ using BPSK over an AWGN channel. . . . . 68

FIGURE	Page
12	Performance of various decoding algorithms for the (255,127) RS code using BPSK over an AWGN channel. . . . . 69
13	Performance of various decoding algorithms for the (255,191) RS code using 256-QAM over an AWGN channel. . . . . 70
14	Performance of various decoding algorithms for the (255,223) RS code using BPSK over an AWGN channel. . . . . 71
15	Tanner graph of the joint BP decoder for ISI channels. The notations <b>a, b, c, d</b> denote the average densities of the messages traversing along the graph used in density evolution (DE). The quantities inside the brackets are erasure rates used in DE for the GEC case. The update schedule of the joint BP decoder is also implied by the arrows in this figure. . . . . 89
16	The joint graph for the $(d_l, d_r, L)$ ensemble over the ISI channels. Illustrated in this figure is the case where $d_l = 3$ and $d_r = 6$ . In the setup we consider, the bit transmission starts in the top left corner and proceeds row by row (e.g., see the green arrows). The (red) stars are to “connect” consecutive rows. . . . . 90
17	EBP-EXIT curves and BP thresholds for various LDPC ensembles over the DEC: (a) $(\lambda, \rho) = (x^2, x^5)$ , (b) $(\lambda, \rho) = (0.6x + 0.4x^9, x^5)$ , (c) $(\lambda, \rho) = (0.2x + 0.3x^2 + 0.5x^{15}, x^9)$ , (d) $(\lambda, \rho) = (0.4x + 0.6x^4, 0.2x^2 + 0.8x^7)$ . Also, by setting the area of the shaded region equal to the design rate of the corresponding ensemble, one obtains the upper bound $\bar{\epsilon}^{\text{MAP}}$ on the MAP threshold using the technique in Section 2 . . . . . 99
18	EBP-EXIT curves for (3,6) and (4,8) regular LDPC ensembles over the DEC. Projection of the left most point of the curves on to the $\epsilon$ -axis allows one to determine $\epsilon^{\text{BP}}$ . Setting the area under the EBP curves to be equal to the design rate $r$ allows one to find the upper bound $\bar{\epsilon}^{\text{MAP}}$ . . . . . 104
19	A trellis section in the residual graph for the DEC. The notation “?” denotes that an erasure is received at the channel output. One can form a larger bit node by merging all the bit nodes that attach to this trellis section. . . . . 107

FIGURE	Page
20	Function $\Psi(u)$ for the residual graph obtained after joint BP decoding of the $(3, 6)$ -regular LDPC ensemble over the DEC. This shows numerically that the MAP upper bound is tight in this case. . . . . 109
21	EBP-EXIT curves for $(3, 6, L, 5)$ over the DEC with $L = 2\hat{L} + 1$ where $\hat{L} = 2, 4, 8, 16, 32, 64, 128, 246$ . For small values of $L$ , the increase in threshold can be explained by the large rate-loss. As $L$ grows larger, the rate loss becomes negligible and the curves keep moving left, but they saturate at the MAP threshold of the underlying regular ensemble. . . . . 111
22	BP-EXIT curves for $(3, 6, L, 5)$ over the pDEC with $L = 2\hat{L} + 1$ where $\hat{L} = 2, 4, 8, 16, 32, 64, 128, 246$ . Threshold saturation can also be observed for this case. . . . . 111
23	BP-EXIT curves for $(3, 27, L, 5)$ over the PR2EC with $L = 2\hat{L} + 1$ where $\hat{L} = 2, 4, 8, 16, 32, 64, 128, 246$ . . . . . 112
24	The BP-GEXIT curve for $(3, 6)$ -regular and $(5, 10)$ -regular LDPC codes over an AWGN dicode channel with $a(D) = (1 - D)/\sqrt{2}$ . The upper bound $\bar{h}^{\text{MAP}}$ is obtained by setting the area under the BP-GEXIT curve (the shaded region) equal to the code rate. . . . . 122
25	A high-rate example: the BP-GEXIT curve for $(3, 27)$ -regular LDPC codes over an AWGN PR2 channel with $a(D) = (1 + 2D + D^2)/\sqrt{6}$ . The upper bound $\bar{h}^{\text{MAP}}$ is determined by the left border of the shaded region. . . . . 123
26	BER and BP thresholds for the $(3, 6)$ -regular, $(3, 6, 22)$ -SC and $(5, 10, 44)$ -SC LDPC codes over the AWGN dicode channel. . . . . 125
27	Tanner graph for an LDPC code and the SWE . . . . . 129
28	$\theta(e_1, e_2)$ of the residual graph for: (a) the SWE in Remark 22 and (b) the EMAC in Remark 23. . . . . 138
29	Tanner graph of the joint decoder for the EMAC . . . . . 140

FIGURE	Page
30	BP-EXIT curves and MAP threshold for: (a) the (4,6)-regular and (4,6, $L$ ,5) SC ensembles for the SWE where $A = 3/2$ and $L = 2, 4, 8, 16, 32, 64$ , (b) the (3,6,3,9) uncoupled and (3,6,3,9, $L$ ,5) SC ensembles for the EMAC for $L = 2, 4, 8, 16, 32, 64, 128$ . . . . . 148

## CHAPTER I

### INTRODUCTION

Due to unavoidable noise in many communication channels, it is important to use channel coding, a mechanism introduced by Shannon in his seminal paper in 1948 [1] to correct errors that may be introduced when the receiver tries to recover the transmitted data. This is done by carefully adding controlled redundancy to the original data that allows one to trade-off data rate for reliability. The sets of symbol vectors, over some input alphabet, that are to be transmitted are referred to as channel codes.

In his famous paper, Shannon also stated a channel coding theorem by showing that there exists a maximum rate, called the capacity of the channel, below which the fraction of errors can be made arbitrarily small and therefore reliable communication is possible. However, the original proof, based on random coding arguments, is elegant but only shows that good codes exist and does not provide practical constructions for such codes.

Since then, many coding theorists have searched for practical coding schemes that approach the Shannon capacity with affordable encoding and decoding complexity. Often, structure is introduced into the codes to facilitate the encoding and decoding processes. Linear codes are one example of this and they have been shown to achieve the capacity of symmetric channels under maximum-likelihood (ML) decoding (see [2, 3]). However, the complexity of ML decoding is still prohibitively large due to an enormous number of codewords. Over the years, researchers have been borrowing tools from diverse branches of mathematics to construct powerful codes based on a

---

This dissertation follows the style of *IEEE Trans. on Information Theory*.

variety of structures. For example, there are algebraic structure in algebraic codes such as Bose-Chaudhuri-Hocquenghem (BCH) codes and Reed-Solomon (RS) codes (see [4]), trellis structure in convolutional codes and turbo codes (see [5]), graph structure in low-density parity-check (LDPC) codes (see [6]) and the nested structure in polar codes [7]. In fact, turbo codes, LDPC codes, and polar codes can all be carefully designed to perform very close to or even achieve the capacity of binary-input memoryless symmetric channels.

In this dissertation, the main focus is on RS and LDPC codes. They are, perhaps, two of the most popular families of channel codes and have been widely used in various communications systems.

Magnetic recording systems for data storage are among the most critical applications of RS and LDPC codes. This is because, in this information age, there is an ever increasing demand for vast amounts of data. The data storage industry has reacted to this by pushing the limits on the density of recording on the physical medium. Extremely high areal densities require symbols to be physically recorded very close to each other, which causes significant inter-symbol interference (ISI) in the read-channel. In addition, there are many other important considerations that are specific to the recording application. For example, the redundancy from the channel coding must be kept very low to keep data density high and the frame error rates required are often around  $10^{-12}$  or smaller. These constraints make coding and signal processing for the read-channel a very challenging task. On the coding front, RS codes have been the answer to this problem for many years. However, recently, LDPC started to attract a lot of attention and have now appeared in many hard-disk drives [8].

The next two sections provide some background on RS and LDPC codes.

### A. Background on Reed-Solomon Codes

Many advances in coding theory during its first few decades of development involved algebraic codes. The structure of these codes can be exploited to yield practical encoding and decoding algorithms. The major design goal during this time was to maximize the minimum Hamming distance. One reason for this is that bounded distance decoding is guaranteed to correct any number of errors smaller than half the minimum distance.

RS codes are perhaps the most popular algebraic codes. They were introduced in 1960 by Irving Reed and Gustave Solomon [9] as reflected in the name. Because of their beautiful algebraic structure, RS codes possess many nice properties. An  $(n, k)$  RS code of length  $n$  and dimension  $k$  is a maximum distance separable (MDS) linear code with minimum distance  $d_{\min} = n - k + 1$ . With respect to minimum distance, RS codes are optimal because they achieve the Singleton bound [10], i.e., RS codes achieve the maximum  $d_{\min}$  given blocklength  $n$  and dimension  $k$ . RS codes also have efficient hard-decision decoding (HDD) algorithms, such as the Berlekamp-Massey (BM) algorithm, which guarantee to correct up to  $\lfloor \frac{d_{\min}-1}{2} \rfloor$  errors. They tend to perform very well in channels with mixture of both burst and random errors and can achieve very low error rates. However, a major drawback of RS codes is the lack of decoding algorithms that make good use of the soft information available at the output of the channel detector and simultaneously have an affordable complexity. More on RS codes and their decoding algorithms can be found in Chapter II.

### B. Background on LDPC Codes

Generally speaking, LDPC codes, aptly described by their names, are linear block codes with a very small fraction of non-zero entries in the parity-check matrices.

An LDPC code is called  $(d_l, d_r)$ -regular if there are  $d_l$  (and  $d_r$ ) non-zero entries per column (and row) in the parity-check matrix. Often,  $d_l$  and  $d_r$  are small compared to the blocklength of the code.

LDPC were introduced by Gallager in his doctoral dissertation [11] in the same year that RS codes was proposed. However, their value went unrecognized for decades until being rediscovered by MacKay [12], with Tanner's new way of graphically depicting LDPC codes [13] being a significant exception. Using Tanner's method, LDPC codes can be represented by bipartite graphs by using a set of variable nodes, corresponding to the codeword symbols, and a set of check nodes, corresponding to the parity-check constraints of the codes. For example, in the Tanner graph of the  $(d_l, d_r)$ -regular LDPC ensemble, all the bit nodes have degree  $d_l$  and all the check nodes have degree  $d_r$ . The sparse bipartite graph structure of LDPC codes turns out to work very well with belief propagation (BP), a low-complexity message-passing decoding, and in fact can achieve the capacity of several channels.

Since their renaissance, LDPC codes and related topics have attracted an enormous amount of research from the information theory society. As a result, researchers have developed many tools and analyses to improve understanding and the performance of these codes. A notable example is the work by Luby *et al.* [14, 15] where the idea of irregular LDPC codes was introduced. Another example is the work by Richardson, Shokrollahi, and Urbanke [16] where an important analysis termed *density evolution* was proposed to track the performance of the iterative BP decoder. Much of the history and progress associated with LDPC codes is captured well by the book of Richardson and Urbanke [6].

Throughout this dissertation, the standard degree distribution (d.d.) is used to characterize irregular LDPC ensembles. From the edge perspective, the d.d. pair consists of two polynomials  $\lambda(z) = \sum_{i \geq 1} \lambda_i z^{i-1}$  and  $\rho(z) = \sum_{i \geq 1} \rho_i z^{i-1}$  whose coefficients

$\lambda_i$  (or  $\rho_i$ ) give the fraction of edges that connect to bit (or check) nodes of degree  $i$ . Equivalently, the LDPC ensemble can also be viewed from the node perspective where its d.d. pair  $L(z) = \sum_{i \geq 1} L_i z^i$  and  $R(z) = \sum_{i \geq 1} R_i z^i$  have coefficients  $L_i$  (or  $R_i$ ) equal to the fraction of bit (or check) nodes of degree  $i$ . An LDPC ensemble of length  $n$  with d.d.  $(\lambda, \rho)$ , or equivalently  $(L, R)$ , is denoted as LDPC( $n, \lambda, \rho$ ), or equivalently LDPC( $n, L, R$ ). The design rate of an LDPC ensemble is given by

$$r = 1 - \frac{L'(1)}{R'(1)} = 1 - \frac{\int_0^1 \rho(z) dz}{\int_0^1 \lambda(z) dz}.$$

In the following subsection, we will briefly discuss a special class of LDPC codes, namely terminated LDPC convolutional codes which are also known as spatially-coupled (SC) codes. These SC codes will be the main subject of Chapters III and IV.

### 1. Spatially-Coupled Codes

The notion of LDPC convolutional codes were introduced by Feldstrom and Zigangirov in 1999 [17]. Later, it was shown that terminated LDPC convolutional codes have excellent BP thresholds which can get quite close to the capacity of many memoryless channels [18, 19]. Recently, the mechanism behind this impressive performance was explained by Kudekar, Richardson and Urbanke [20]. They describe a phenomenon, termed *threshold saturation via spatial coupling*, whereby the BP threshold of SC codes saturates to the MAP threshold of the underlying uncoupled ensemble.

The class of SC ensembles in general can be defined quite broadly. In this dissertation, we mainly consider two basic variants (see details in [20]) as discussed below.



According to [20], the design rate of the  $(d_l, d_r, L)$  ensemble is given by

$$r(d_l, d_r, L) = \left(1 - \frac{d_l}{d_r}\right) - \frac{d_l}{d_r} \cdot \frac{d_l - 1}{L}.$$

b. The  $(d_l, d_r, L, w)$  Ensemble

The  $(d_l, d_r, L, w)$  can be obtained with the introduction of a “smoothing” parameter  $w$ . One still places  $M$  variable nodes at each position in  $[1, L]$  but places  $\frac{d_l}{d_r}M$  check nodes at each position in  $[1, L + w - 1]$ . Each bit node at position  $i$  is connected uniformly and independently to a total of  $d_l$  check nodes at positions from the range  $[i, i+w-1]$ . By adding this randomization to the edge connections, the system behaves like a continuous one for large enough  $w$  and a proof of the threshold saturation effect becomes feasible [20]. The design rate of the  $(d_l, d_r, L, w)$  ensemble is given in [20] by

$$r(d_l, d_r, L, w) = \left(1 - \frac{d_l}{d_r}\right) - \frac{d_l}{d_r} \cdot \frac{w + 1 - 2 \sum_{i=0}^w \left(\frac{i}{w}\right)^{d_r}}{L}.$$

C. Dissertation Outline

This dissertation is organized as follows. In Chapter II, we propose a rate-distortion (RD) framework to analyze and design multiple decoding attempts of RS codes. In Chapter III, SC codes are considered over ISI channels and threshold saturation is also observed based on the construction of a generalized extrinsic information transfer (GEXIT) curve. In Chapter IV, a similar technique is extended to two multiuser channels where the MAP thresholds of LDPC codes over these channels are rigorously investigated. As a consequence, SC codes with threshold saturation are conjectured to universally achieve the entire capacity region of these three models in Chapters III and IV. Finally, conclusions and future directions of work are pointed out in Chapter V.

## D. Notation

Throughout this dissertation,  $n$  is used to denote the blocklength of the codes. The subvector  $(X_i, X_{i+1}, \dots, X_j)$  of the vector  $(X_1, X_2, \dots, X_n)$  is denoted by  $X_i^j$  for convenience. In Chapter IV, vectors of length  $n$  are also denoted by bold faced letters such as  $\mathbf{X}$ . For simplicity of notation, we write  $Y_{\sim i}$  to denote the vector  $Y_1^n \setminus Y_i$ . The standard symbols  $\mathbb{R}$ ,  $\mathbb{N}$ , and  $\mathbb{Z}$  are used to denote the set of real numbers, natural numbers, and integers, respectively. The set of non-negative real numbers is denoted by  $\mathbb{R}_{\geq 0}$  meanwhile  $\mathbb{E}$  is used to denote expectation. Finally,  $H_2(x)$  is used to denote the binary entropy function, which is defined by  $H_2(x) \triangleq -x \log_2(x) - (1-x) \log_2(1-x)$ .

## CHAPTER II

A RATE-DISTORTION APPROACH TO MULTIPLE DECODING ATTEMPTS  
OF REED-SOLOMON CODES\*

## A. Introduction

Since the discovery of RS codes [9], researchers have spent a considerable effort on improving the decoding performance at the expense of complexity. A breakthrough result of Guruswami and Sudan (GS) introduced an algebraic hard-decision list-decoding algorithm, based on bivariate interpolation and factorization, that can correct errors well beyond half the minimum distance of the code [22]. Nevertheless, hard-decision decoding (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 [23].

The GS and KV algorithms, however, have significant computational complexity. Therefore, multiple runs of errors-and-erasures and errors-only decoding with some low-complexity algorithm, such as the BM algorithm, has renewed the interest of researchers. These algorithms use the soft-information available at the channel output to construct a set of either erasure patterns [24, 25], test patterns [26], or patterns combining both [27, 28] and then attempt to decode using each pattern. Techniques have also been introduced to lower the complexity per decoding trial in [29, 30, 31].

---

\*Copyright 2011 IEEE. Reprinted, with permission, from P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "On multiple decoding attempts for Reed-Solomon codes: A rate-distortion approach," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 668-691, Feb. 2011. For more information, go to <http://thesis.tamu.edu/forms/IEEE%20permission%20note.pdf/view>.

Other soft-decision decoding algorithms for RS codes include [32, 33] that use the binary expansion of RS codes to work on the bit-level. In [32], belief propagation is run while the parity-check matrix is iteratively adapted on the least reliable basis. Meanwhile, [33] adapts the generator matrix on the most reliable basis and uses reprocessing techniques based on ordered statistics.

In the scope of multiple errors-and-erasures decoding, there have been several algorithms proposed that use different erasure codebooks (i.e., different sets of erasure patterns). After running the errors-and-erasures decoding algorithm multiple times, each time using one erasure pattern in the set, these algorithms produce a list of candidate codewords, whose size is usually small, and then pick the best codeword on this list. The common idea of constructing the set of erasure patterns in these multiple errors-and-erasures decoding algorithms 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) [24] and it repeats errors-and-erasures 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 [25] 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 errors-and-erasures 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 errors-and-erasures decoding. Inspired by this, we first develop a rate-distortion (RD) framework to analyze the asymptotic trade-off between performance and complexity of multiple errors-and-erasures decoding of RS codes. The main idea is to choose an appropriate distortion measure so that the decoding is suc-

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.

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 RDE approach essentially focuses on maximizing the exponent at which the error probability decays as  $n$  goes to infinity. The 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 rate-distortion exponent tends to zero. Using the RDE analysis, this approach also helps answer the following two questions: (i) What is the minimum error probability achievable for 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 certain error probability?

The RD and RDE approaches are also extended beyond conventional errors-and-erasures decoding to analyze multiple-decoding for decoding schemes such as ASD decoding. It is interesting to note that the RDE approach for ASD decoding schemes contains the special case where the codebook has exactly one entry (i.e., ASD decoding is run only once). 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 [34, 35, 36, 37] where various researchers solve for a multiplicity matrix that minimizes the error probability obtained by either using a

Gaussian approximation [34], applying a Chernoff bound [35, 36], or using Sanov's theorem [37].

Finally, we propose a family of multiple-decoding algorithms based on these two approaches that achieve better performance-versus-complexity trade-off than other algorithms.

The chapter is organized as follows. In Section B, we design an appropriate distortion measure and present a rate-distortion framework, for both the RD and RDE approaches, to analyze the performance-versus-complexity trade-off of multiple errors-and-erasures decoding of RS codes. Also in this section, we propose a general multiple-decoding algorithm that can be applied to errors-and-erasures decoding. Then, in Section C, we discuss numerical computations of RD and RDE functions together with their complexity analyses which are needed for the proposed algorithm. In Section D, we analyze both bit-level and symbol-level ASD decoding and design distortion measures compatible with the general algorithm. A closed-form analysis of some RD and RDE functions is presented in Section E. Next, in Section F, we offer some extensions that combine covering codes with random codes and also consider the case of a single decoding attempt. Finally, simulation results are presented in Section G. Part of the results in this chapter have appeared in [38, 39, 40].

## B. A RD Framework for Multiple Errors-and-Erasures Decoding

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

Let  $\mathbb{F}_m$  with  $m = 2^n$  be the Galois field with  $m$  elements denoted as  $\alpha_1, \alpha_2, \dots, \alpha_m$ . We consider an  $(n, k)$  RS code of length  $n$ , dimension  $k$  over  $\mathbb{F}_m$ . Assume that we transmit a codeword  $\mathbf{c} = (c_1, c_2, \dots, c_n) \in \mathbb{F}_m^n$  over some channel and receive a

vector  $\mathbf{r} = (r_1, r_2, \dots, r_n) \in \mathcal{Y}^n$  where  $\mathcal{Y}$  is the received alphabet for a single RS symbol. While our approach can be applied to much more general channels, our simulations focus on the Additive White Gaussian Noise (AWGN) channel and two common modulation formats, namely BPSK and  $m$ -QAM. Correspondingly, we use  $\mathcal{Y} = \mathbb{R}^n$  for BPSK and  $\mathcal{Y} = \mathbb{R}^2$  for  $m$ -QAM. For each codeword index  $i$ , let  $\varphi_i : \{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, m\}$  be the permutation given by sorting  $\pi_{i,j} = \Pr(c_i = \alpha_j | r_i)$  in decreasing order so that  $\pi_{i,\varphi_i(1)} \geq \pi_{i,\varphi_i(2)} \geq \dots \geq \pi_{i,\varphi_i(m)}$ . Then, we can specify  $y_{i,j} = \alpha_{\varphi_i(j)}$  as the  $j$ -th most reliable symbol for  $j = 1, \dots, m$  at codeword index  $i$ . To obtain the reliability of the codeword positions (indices), we construct the permutation  $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  given by sorting the probabilities  $\pi_{i,\varphi_i(1)}$  of the most likely symbols in increasing order.<sup>1</sup> Thus, codeword position  $\sigma(i)$  is the  $i$ -th LRP. These above notations will be used throughout this chapter.

**Example 1.** Consider  $n = 3$  and  $m = 4$ . Assume that we have the probability  $\pi_{i,j}$  written in a matrix form as follows:

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

then  $\varphi_1(1, 2, 3, 4) = (2, 3, 4, 1)$ ,  $\varphi_2(1, 2, 3, 4) = (3, 4, 2, 1)$ ,  $\varphi_3(1, 2, 3, 4) = (1, 2, 4, 3)$  and  $\sigma(1, 2, 3) = (2, 3, 1)$ .

**Condition 1.** (Classical decoding threshold, see [4, 5]): If  $e$  symbols are erased, a conventional hard-decision errors-and-erasures decoder such as the BM algorithm is

---

<sup>1</sup>Other measures such as entropy or the average number of guesses might improve Algorithm B in Section 3.

able to correct  $\nu$  errors in unerased positions if and only if

$$2\nu + e < n - k + 1. \quad (2.1)$$

1. Conventional Error Patterns and Erasure Patterns.

**Definition 1.** (*Conventional error patterns and erasure patterns*) We define  $x_1^n \in \mathbb{Z}_2^n \triangleq \{0, 1\}^n$  and  $\hat{x}_1^n \in \mathbb{Z}_2^n$  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 the symbol at index  $i$  is erased (i.e., an erasure is applied at index  $i$ ).  $X_1^n$  and  $\hat{X}_1^n$  will be used to denote the random vectors which generate the realizations  $x_1^n$  and  $\hat{x}_1^n$ , respectively.

**Example 2.** If  $d_{\min}$  is odd then the GMD algorithm corresponds to the set

$$\{111111\dots, 001111\dots, 000011\dots, \dots, \underbrace{00\dots 0}_{d_{\min}-1} 11\dots 1\}$$

of erasure patterns. Meanwhile, the SED(3, 2) uses the following set

$$\{\underline{111111}\dots, \underline{001111}\dots, \underline{010111}\dots, \underline{100111}\dots\}.$$

Here, in each erasure pattern, the letters are written in increasing reliability order of the codeword positions.

Let us revisit the question of how to construct the best set of erasure patterns for multiple errors-and-erasures decoding. First, it can be seen that a multiple errors-and-erasures decoding succeeds if the condition (2.1) is satisfied during at least one round of decoding. Thus, our approach is to design a distortion measure that converts the condition (2.1) into a form where the distortion between an error pattern  $x_1^n$  and an erasure pattern  $\hat{x}_1^n$ , denoted as  $d(x_1^n, \hat{x}_1^n)$ , is less than a fixed threshold.

**Definition 2.** Given a letter-by-letter distortion measure  $\delta$ , the distortion between an error pattern  $x_1^n$  and an erasure pattern  $\hat{x}_1^n$  is defined by

$$d(x_1^n, \hat{x}_1^n) = \sum_{i=1}^n \delta(x_i, \hat{x}_i).$$

**Proposition 1.** If we choose the letter-by-letter distortion measure  $\delta : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_{\geq 0}$ , where in this case  $\mathcal{X} = \hat{\mathcal{X}} = \mathbb{Z}_2$ , as follows:

$$\begin{aligned} \delta(0,0) &= 1, & \delta(0,1) &= 2, \\ \delta(1,0) &= 1, & \delta(1,1) &= 0, \end{aligned} \tag{2.2}$$

then the condition (2.1) for a successful errors-and-erasures decoding is equivalent to

$$d(x_1^n, \hat{x}_1^n) < n - k + 1 \tag{2.3}$$

where the distortion is less than a fixed threshold.

*Proof.* First, we define

$$\chi_{j,\hat{j}} \triangleq |\{i \in \{1, 2, \dots, n\} : x_i = j, \hat{x}_i = \hat{j}\}|$$

to count the number of  $(x_i, \hat{x}_i)$  pairs equal to  $(j, \hat{j})$  for every  $j \in \mathcal{X}$  and  $\hat{j} \in \hat{\mathcal{X}}$ . With the chosen distortion measure, we have

$$d(x_1^n, \hat{x}_1^n) = 2\chi_{0,1} + \chi_{0,0} + \chi_{1,0}.$$

Noticing that  $e = \chi_{0,0} + \chi_{1,0}$  and  $\nu = \chi_{0,1}$ , the condition (2.1) for one errors-and-erasures decoding attempt to succeed becomes  $2\chi_{0,1} + \chi_{0,0} + \chi_{1,0} < n - k + 1$  which is equivalent to  $d(x_1^n, \hat{x}_1^n) < n - k + 1$ .  $\square$

Next, we try to maximize the chance that this successful decoding condition is satisfied by at least one of the decoding attempts (i.e.,  $d(x_1^n, \hat{x}_1^n) < n - k + 1$  for at least one erasure pattern  $\hat{x}_1^n$ ). Mathematically, we want to build a set  $\mathcal{B}$  of no more than

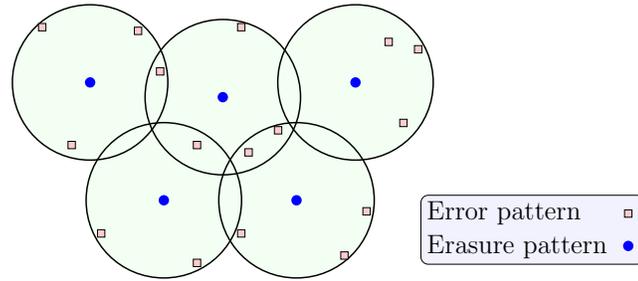


Fig. 2. Pictorial illustration of a covering problem

$2^R$  erasure patterns  $\hat{x}_1^n$  that achieves the maximum

$$\max_{\mathcal{B}:|\mathcal{B}|\leq 2^R} \Pr \left\{ \min_{\hat{x}_1^n \in \mathcal{B}} d(X_1^n, \hat{x}_1^n) < n - k + 1 \right\}.$$

Solving this problem exactly is very difficult. However, one can observe that it is a covering problem (see Fig. 2) where one tries to cover the most-likely error patterns using a fixed number of spheres centered at the chosen erasure patterns. This view leads to two asymptotic solutions of the problem based on rate-distortion theory. Taking this point of view, we view the error pattern  $x_1^n$  as a source sequence and the erasure pattern  $\hat{x}_1^n$  as a reproduction sequence.

#### a. RD Approach

Rate-distortion theory (see [41, Chapter 13]) characterizes the trade-off between  $\bar{R}$  and  $\bar{D}$  such that sets  $\mathcal{B}$  of  $2^{n\bar{R}}$  reproduction sequences exist (and can be generated randomly) so that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{X_1^n, \mathcal{B}} \left[ \min_{\hat{x}_1^n \in \mathcal{B}} d(X_1^n, \hat{x}_1^n) \right] < \bar{D}.$$

Under mild conditions, this implies that, for large enough  $n$ , we have

$$\min_{\hat{x}_1^n \in \mathcal{B}} d(X_1^n, \hat{x}_1^n) < n\bar{D}$$

with high probability. Here,  $\bar{R}$  and  $\bar{D}$  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  $\bar{R}$  and  $\bar{D}$ . In this chapter, we denote the rate and distortion by  $R$  and  $D$ , respectively, using unnormalized quantities, i.e.,  $R = n\bar{R}$  and  $D = n\bar{D}$ .

### b. RDE Approach

The above-mentioned RD approach focuses on minimizing the average minimum distortion with little knowledge of how the tail of the distribution behaves. In this RDE approach, we instead focus on directly minimizing the probability that the minimum distortion is not less than the predetermined threshold  $D = n - k + 1$  (due to the condition (2.3)) with the help of an error-exponent analysis. The exact probability of interest is

$$p_e = \Pr \left( X_1^n : \min_{\hat{x}_1^n \in \mathcal{B}} d(X_1^n, \hat{x}_1^n) > D \right)$$

that reflects how likely the decoding threshold (2.1) is going to fail. In other words, every error pattern  $x_1^n$  can be covered by a sphere centered at an erasure pattern  $\hat{x}_1^n$  except for a set of error patterns of probability  $p_e$ . The RDE analysis shows that  $p_e$  decays exponentially as  $n \rightarrow \infty$  and the maximum exponent attainable is the RDE function  $F(R, D)$ . Throughout this chapter, 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 [42, 43, 44] is often the normalized version  $\bar{F}(R, D) \triangleq \frac{F(R, D)}{n}$ .

RDE analysis is discussed extensively in [42, 43] and it is shown that a set  $\mathcal{B}$  of roughly  $2^{n\bar{R}}$  codewords, generated randomly using the test-channel input distribution, can be used to achieve  $\bar{F}(R, D)$ . An upper bound is also given that shows, for any

$\epsilon > 0$ , there is a sufficiently large  $n$  (see [45, p. 229]) such that

$$p_e \leq 2^{-n[\bar{F}(R,D)-\epsilon]}.$$

An exponentially tight lower bound for  $p_e$  can also be obtained (see [45, p. 236]) and it implies that the best sequence of codebooks satisfy

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log_2 p_e = \bar{F}(R, D).$$

**Remark 1.** *The RDE approach possesses several advantages. First, the converse of the RDE [45, p. 236] provides a lower bound for  $p_e$ . This implies that, given an arbitrary set  $\mathcal{B}$  of roughly  $2^{n\bar{R}}$  erasure patterns and any  $\epsilon > 0$ , the probability  $p_e$  cannot be made lower than  $2^{-n[\bar{F}(R,D)+\epsilon]}$  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. Second, 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.*

## 2. Generalized Error Patterns and Erasure Patterns

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

**Definition 3.** *(Generalized error patterns and erasure patterns) Consider a positive*

integer  $\ell$  smaller than the field size  $m$ . Let  $x_1^n \in \mathbb{Z}_{\ell+1}^n$  be a generalized error pattern where, at index  $i$ ,  $x_i = j$  implies that the  $j$ -th most likely symbol is correct for  $j \in \{1, 2, \dots, \ell\}$ , and  $x_i = 0$  implies none of the first  $\ell$  most likely symbols is correct. Let  $\hat{x}_1^n \in \mathbb{Z}_{\ell+1}^n$  be a generalized erasure pattern used for decoding where, at index  $i$ ,  $\hat{x}_i = \hat{j}$  implies that the  $\hat{j}$ -th most likely symbol is used as the hard-decision symbol for  $\hat{j} \in \{1, 2, \dots, \ell\}$ , and  $\hat{x}_i = 0$  implies that an erasure is used at that index.

For simplicity, we refer to  $x_1^n$  as the error pattern and  $\hat{x}_1^n$  as the erasure pattern like in the conventional case. Now, we need to convert the condition (2.1) to the form where  $d(x_1^n, \hat{x}_1^n)$  is less than a fixed threshold. Proposition 1 is thereby generalized into the following proposition.

**Proposition 2.** We choose the letter-by-letter distortion measure  $\delta : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_{\geq 0}$ , where in this case  $\mathcal{X} = \hat{\mathcal{X}} = \mathbb{Z}_{\ell+1}$ , defined by  $\delta(x, \hat{x}) = [\Delta]_{x, \hat{x}}$  in terms of the  $(\ell + 1) \times (\ell + 1)$  matrix

$$\Delta = \begin{pmatrix} 1 & 2 & \dots & 2 & 2 \\ 1 & 0 & \dots & 2 & 2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 2 & \dots & 0 & 2 \\ 1 & 2 & \dots & 2 & 0 \end{pmatrix}. \quad (2.4)$$

Using this, the condition (2.1) for a successful errors-and-erasures decoding is equivalent to

$$d(x_1^n, \hat{x}_1^n) < n - k + 1.$$

*Proof.* The reasoning is very similar to the proof of Proposition 1 using the fact that  $e = \sum_{j=0}^{\ell} \chi_{j,0}$  and  $\nu = \sum_{\hat{j}=1}^{\ell} \sum_{j=0, j \neq \hat{j}}^{\ell} \chi_{j, \hat{j}}$  where

$$\chi_{j, \hat{j}} \triangleq |\{i \in \{1, 2, \dots, n\} : x_i = j, \hat{x}_i = \hat{j}\}|$$

for every  $j, \hat{j} \in \mathbb{Z}_{\ell+1}$ . □

For each  $\ell = 1, 2, \dots, m$ , we will refer to this generalized case as mBM- $\ell$  decoding.

**Example 3.** Consider mBM-2 (or top- $\ell$  decoding with  $\ell = 2$ ). In this case, the distortion measure is given by following the matrix

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

**Remark 2.** The distortion measure matrix changes slightly if we use the errors-only decoding instead of errors-and-erasures decoding. In this case,  $\hat{\mathcal{X}} = \mathbb{Z}_{\ell+1} \setminus \{0\}$  and the chosen letter-by-letter distortion measure is given in terms of the  $(\ell + 1) \times \ell$  matrix obtained by deleting the first column of (2.4). When  $\ell = 2$ , we consider the first and second most likely symbols as the two hard-decision symbols at each codeword position. This is similar to the Chase-type decoding method proposed by Bellorado and Kavcic [29]. Das and Vardy also suggest this approach by considering only several highest entries in each column of the reliability matrix  $\Pi$  for single ASD decoding of RS codes [37].

### 3. Proposed General Multiple-Decoding Algorithm

In this section, we propose two general multiple-decoding algorithms for RS codes. In each algorithm, one can choose either Step 2a that corresponds to the RD approach or Step 2b that corresponds to the RDE approach. These general algorithms apply to not only multiple errors-and-erasures decoding but also multiple-decoding of other decoding schemes that we will discuss later. The common first step is designing a distortion measure  $\delta : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_{\geq 0}$  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.

a. Algorithm A

*Step 1:* Based on the received signal sequence, compute an  $m \times n$  reliability matrix  $\mathbf{\Pi}$  where  $[\mathbf{\Pi}]_{j,i} = \pi_{i,j}$ . From this, determine the probability matrix  $\mathbf{P}$  where  $p_{i,j} = \Pr(X_i = j)$  for  $i = 1, 2, \dots, n$  and  $j \in \mathcal{X}$ .

*Step 2a:* (RD approach) Compute the RD function of a source sequence (error pattern) with probability of source letters derived from  $\mathbf{P}$  and the chosen distortion measure (see Section C and Section E). Given the design rate  $R$ , determine the optimal input-probability distribution matrix  $\mathbf{Q}$ , for the test channel, with entries  $q_{i,\hat{j}} = \Pr(\hat{X}_i = \hat{j})$  for  $i = 1, 2, \dots, n$  and  $\hat{j} \in \hat{\mathcal{X}}$ .

*Step 2b:* (RDE approach) Given  $D$  (in most cases  $D = n-k+1$ ) and the design rate  $R$ , compute the RDE function of a source sequence (error pattern) with probability of source letters derived from  $\mathbf{P}$  and the chosen distortion measure (see Section C and Section E). Also determine the optimal input-probability distribution matrix  $\mathbf{Q}$ , for the test channel, with entries  $q_{i,\hat{j}} = \Pr(\hat{X}_i = \hat{j})$  for  $i = 1, 2, \dots, n$  and  $\hat{j} \in \hat{\mathcal{X}}$ .

*Step 3:* Randomly generate a set of  $2^R$  erasure patterns using the test-channel input-probability distribution matrix  $\mathbf{Q}$ .

*Step 4:* Run multiple attempts of the corresponding decoding scheme (e.g., errors-and-erasures decoding) using the set of erasure patterns in Step 3 to produce a list of candidate codewords.

*Step 5:* Use the maximum-likelihood (ML) rule to pick the best codeword on the list.

**Remark 3.** In Algorithm A, the RD (or RDE) function is computed on the fly, i.e.,

after every received signal sequence. In practice, it may be preferable to precompute the RD (or RDE) function based on the empirical distribution measured from the channel. We refer to this approach as Algorithm B, and simulation results show a negligible difference in the performance of these two algorithms.

b. Algorithm B

*Step 1:* Transmit  $\tau$  (e.g.,  $\tau = 10^3 - 10^6$ ) arbitrary test RS codewords, indexed by time  $t = 1, 2, \dots, \tau$ , over the channel and compute a set of  $\tau$   $m \times n$  matrices  $\mathbf{\Pi}_1^{(t)}$  where  $[\mathbf{\Pi}_1^{(t)}]_{j,i} = \pi_{i,\varphi_i^{(t)}(j)}^{(t)}$  is the probability of the  $j$ -th most likely symbol at position  $i$  during time  $t$ . For each time  $t$ , obtain the matrix  $\mathbf{\Pi}_2^{(t)}$  from  $\mathbf{\Pi}_1^{(t)}$  through a permutation  $\sigma^{(t)} : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  that sorts the probabilities  $\pi_{i,\varphi_i^{(t)}(1)}^{(t)}$  in increasing order to indicate the reliability order of codeword positions. Take the entry-wise average of all  $\tau$  matrices  $\mathbf{\Pi}_2^{(t)}$  to get an average matrix  $\bar{\mathbf{\Pi}}$ .<sup>2</sup> The matrix  $\bar{\mathbf{\Pi}}$  serves as  $\mathbf{\Pi}$  in Algorithm A and from this, determine the probability matrix  $\mathbf{P}$  where  $p_{i,j} = \Pr(X_i = j)$  for  $i = 1, 2, \dots, n$  and  $j \in \mathcal{X}$ .

*Step 2a:* (RD approach) Compute the RD function of a source sequence (error pattern) with probability of source letters derived from  $\mathbf{P}$  and the chosen distortion measure. Given a design rate  $R$ , determine the test-channel input-probability distribution matrix  $\mathbf{Q}$  where  $q_{i,\hat{j}} = \Pr(\hat{X}_i = \hat{j})$  for  $i = 1, 2, \dots, n$  and  $\hat{j} \in \hat{\mathcal{X}}$ .

*Step 2b:* (RDE approach) Given  $D$  (in most cases  $D = n - k + 1$ ) and the design rate  $R$ , compute the RDE function of a source sequence (error pattern) with probability of source letters derived from  $\mathbf{P}$  and the chosen distortion measure. Also determine the optimal test-channel input-probability distribution matrix  $\mathbf{Q}$  where  $q_{i,\hat{j}} = \Pr(\hat{X}_i = \hat{j})$  for  $i = 1, 2, \dots, n$  and  $\hat{j} \in \hat{\mathcal{X}}$ .

---

<sup>2</sup>In fact, one need not store separately each  $\mathbf{\Pi}_2^{(t)}$  matrix. The average  $\bar{\mathbf{\Pi}}$  can be computed on the fly.

*Step 3:* Based on the actual received signal sequence, compute  $\pi_{i,\varphi_i(1)}$  and determine the permutation  $\sigma$  that gives the reliability order of codeword positions by sorting  $\pi_{i,\varphi_i(1)}$  in increasing order.

*Step 4:* Randomly generate a set of  $2^R$  erasure patterns using the test-channel input-probability distribution matrix  $\mathbf{Q}$  and permute the indices of each erasure pattern by the permutation  $\sigma^{-1}$ .

*Step 5:* Run multiple attempts of the corresponding decoding scheme (e.g., errors-and-erasures decoding) using the set of erasure patterns in Step 4 to produce a list of candidate codewords.

*Step 6:* Use the ML rule to pick the best codeword on the list.

### C. Computing the RD and RDE Functions

In this section, we will discuss some numerical methods to compute the RD and RDE functions and the corresponding test-channel input-probability distribution matrix  $\mathbf{Q}$ , whose entries are  $q_{i,\hat{j}} = \Pr(\hat{X}_i = \hat{j})$  for  $i = 1, 2, \dots, n$  and  $\hat{j} \in \hat{\mathcal{X}}$ . These numerical methods allow us to efficiently compute the RD and RDE functions discussed in the previous section for arbitrary discrete distortion measures. For some simple distortion measures, closed-form solutions are given in Section E.

#### 1. Computing the RD Function

For an arbitrary discrete distortion measure, it can be difficult to compute the RD function analytically. Fortunately, for a single source  $X$ , the Blahut algorithm (see details in [46]) gives an alternating minimization technique that efficiently computes the RD function which is given by

$$R(D) = \min_{\mathbf{w} \in \mathcal{W}_D} \sum_j \sum_{\hat{j}} p_j w_{\hat{j}|j} \log_2 \frac{w_{\hat{j}|j}}{\sum_{j'} p_{j'} w_{\hat{j}|j'}}$$

where  $p_j \triangleq \Pr(X = j)$ ,  $q_{\hat{j}} \triangleq \Pr(\hat{X} = \hat{j})$ ,  $w_{\hat{j}|j} \triangleq \Pr(\hat{X} = \hat{j}|X = j)$ , and<sup>3</sup>

$$\mathcal{W}_D = \left\{ \mathbf{w} \left| \begin{array}{l} w_{\hat{j}|j} \geq 0, \sum_{\hat{j}} w_{\hat{j}|j} = 1 \\ \sum_j \sum_{\hat{j}} p_j w_{\hat{j}|j} \delta_{j\hat{j}} \leq D \end{array} \right. \right\}.$$

More precisely, given the Lagrange multiplier  $t \leq 0$  that represents the slope of the RD curve at a specific point (see [47, Thm 2.5.1]) and an arbitrary all-positive initial test-channel input-probability distribution vector  $\mathbf{q}^{(0)}$ , the Blahut algorithm shows us how to compute the rate-distortion pair  $(R_t, D_t)$ .

However, it is not straightforward to apply the Blahut algorithm to compute the RD for a discrete source sequence  $x_1^n$  (an error pattern in our context) of  $n$  independent but not necessarily identical (i.n.d.) source components  $x_i$ . In order to do that, we consider the group of source letters  $(j_1, j_2, \dots, j_n)$  where  $j_i \in \mathcal{X}$  as a super-source letter  $\mathcal{J} \in \mathcal{X}^n$ , the group of reproduction letters  $(\hat{j}_1, \hat{j}_2, \dots, \hat{j}_n)$  where  $\hat{j}_i \in \hat{\mathcal{X}}$  as a super-reproduction letter  $\hat{\mathcal{J}} \in \hat{\mathcal{X}}^n$ , and the source sequence  $x_1^n$  as a single source. For each super-source letter  $\mathcal{J}$ ,  $p_{\mathcal{J}} = \Pr(X_1^n = \mathcal{J}) = \prod_{i=1}^n \Pr(X_i = j_i) = \prod_{i=1}^n p_{j_i}$  follows from the independence of source components.<sup>4</sup>

While we could apply the Blahut algorithm to this source directly, the complexity is a problem because the alphabet sizes for  $\mathcal{J}$  and  $\hat{\mathcal{J}}$  become the super-alphabet sizes  $|\mathcal{X}|^n$  and  $|\hat{\mathcal{X}}|^n$  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.,  $q_{\hat{\mathcal{J}}}^{(0)} = \prod_{i=1}^n q_{\hat{j}_i}^{(0)}$ . One can verify that this factorization rule still applies after every step  $\tau$  of the iterative process, i.e.,  $q_{\hat{\mathcal{J}}}^{(\tau)} = \prod_{i=1}^n q_{\hat{j}_i}^{(\tau)}$ . Therefore, the convergence of the Blahut algorithm [48]

---

<sup>3</sup> $\delta(j, \hat{j})$  is sometimes written as  $\delta_{j\hat{j}}$  for convenience.

<sup>4</sup>In this chapter, the notations  $p_{j_i}$  and  $p_{i,j}$  are interchangeable. The notations  $q_{\hat{j}_i}$  and  $q_{i,\hat{j}}$  are also interchangeable.

implies that the optimal distribution is a product distribution, i.e.,  $q_{\hat{\mathcal{J}}}^* = \prod_{i=1}^n q_{j_i}^*$ .

One can also find that, for each parameter  $t$ , one only needs to compute the rate-distortion pair for each source component  $x_i$  separately and sum them together. This is captured into the following algorithm.

**Algorithm 1.** (*Factored Blahut algorithm for RD function*) Consider a discrete source sequence  $x_1^n$  of  $n$  i.n.d. source components  $x_i$ 's with probability  $p_{j_i} \triangleq \Pr(X_i = j_i)$ . Given a parameter  $t \leq 0$ , the rate and the distortion for this source sequence under a specified distortion measure are given by

$$R_t = \sum_{i=1}^n R_{i,t} \text{ and } D_t = \sum_{i=1}^n D_{i,t} \quad (2.5)$$

where the components  $R_{i,t}$  and  $D_{i,t}$  are computed by the Blahut algorithm with the Lagrange multiplier  $t$ . This rate-distortion pair can be achieved by the corresponding test-channel input-probability distribution  $q_{\hat{\mathcal{J}}} \triangleq \Pr(\hat{X}_1^n = \hat{\mathcal{J}}) = \prod_{i=1}^n q_{j_i}$  where the component probability distribution  $q_{j_i} \triangleq \Pr(\hat{X}_i = j_i)$ .

**Remark 4.** Equation (2.5) can also be derived from [47, Corollary 2.8.3] in a way that does not use the convergence property of the Blahut algorithm.

## 2. Computing the RDE Function

The original RDE function  $F(R, D)$ , defined in [42, Sec. VI] for a single source  $X$ , is given by

$$F(R, D) = \max_{\mathbf{w}} \min_{\tilde{\mathbf{p}} \in \mathcal{P}_{R,D}} \sum_j \tilde{p}_j \log_2 \frac{\tilde{p}_j}{p_j} \quad (2.6)$$

where  $p_j = \Pr(X = j)$ ,  $q_j = \Pr(\hat{X} = j)$ ,  $w_{j|j} = \Pr(\hat{X} = j|X = j)$ , and

$$\mathcal{P}_{R,D} = \left\{ \tilde{\mathbf{p}} \left| \begin{array}{l} \sum_j \sum_{\hat{j}} \tilde{p}_j w_{j|\hat{j}} \log_2 \frac{w_{j|\hat{j}}}{\sum_{j'} \tilde{p}_{j'} w_{j'|\hat{j}}} \geq R \\ \sum_j \sum_{\hat{j}} \tilde{p}_j w_{j|\hat{j}} \delta_{j\hat{j}} \geq D \end{array} \right. \right\}. \quad (2.7)$$

For a single source  $X$ , given two parameters  $s \geq 0$  and  $t \leq 0$  which are the Lagrange multipliers introduced in the optimization problem (see [42, p. 415]), the Arimoto algorithm given in [49, Sec. V] can be used to compute the exponent, rate, and distortion numerically.

In the context we consider, the source (error pattern)  $x_1^n$  comprises i.n.d. source components  $x_i$ 's. We follow the same method as in the RD function case, i.e., by choosing the initial distribution still arbitrarily but following a factorization rule  $q_{\mathcal{J}}^{(0)} = \prod_{i=1}^n q_{j_i}^{(0)}$ , and this gives the following algorithm.

**Algorithm 2.** (*Factored Arimoto algorithm for RDE function*) Consider a discrete source  $x_1^n$  of i.n.d. source components  $x_i$ 's with probability  $p_{j_i} \triangleq \Pr(X_i = j_i)$ . Given Lagrange multipliers  $s \geq 0$  and  $t \leq 0$ , the exponent, rate and distortion under a specified distortion measure are given by

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

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

**Remark 5.** Though it is standard practice to compute error-exponents using the implicit form given above, this approach may provide points that, while achievable, are strictly below the true RDE curve. The problem is that the true RDE curve may have a slope discontinuity that forces the implicit representation to have extra points. An example of this behavior for the channel coding error exponent is given by Gallager [3, p. 147]. For the i.n.d. source considered above, a cautious person could solve the problem as described and then check that the component RDE functions are differentiable at the optimum point. In this work, we largely neglect this subtlety.

### 3. Complexity of Computing RD/RDE Functions

#### a. Complexity of Computing RD Function.

For each parameter  $t < 0$ , if we directly apply of the original Blahut algorithm to compute the  $(R_t, D_t)$  pair, the complexity is  $O(\tau_{\max}|\mathcal{X}|^n|\hat{\mathcal{X}}|^n)$  where  $\tau_{\max}$  is the number of iterations in the Blahut algorithm. However, using the factored Blahut algorithm (Algorithm 1) greatly reduces this complexity to  $O(\tau_{\max}|\mathcal{X}||\hat{\mathcal{X}}|n)$ . In Section 3, one of the proposed algorithms needs to compute the RD function for a design rate  $R$ . To do this, we apply the bisection method on  $t$  to find the correct  $t$  that corresponds to the chosen rate  $R$ .

- *Step 0:* Set  $t_{\min} < 0$  (e.g.,  $t_{\min} = -10$ )
- *Step 1:* If  $R_{t_{\min}} > R$ , go to Step 3. Else go to Step 2.
- *Step 2:* If  $R_{t_{\min}} = R$  then stop. Else if  $R_{t_{\min}} < R$ , set  $t_{\min} \leftarrow 2t_{\min}$  and go to Step 1.
- *Step 3:* Find  $t$  using the bisection method to get the correct rate  $R$  within  $\epsilon_R$ .

The overall complexity of computing the RD function for a design rate  $R$  is

$$O\left(\tau_{\max} \log_2\left(\frac{-t_{\min}}{\epsilon_R}\right) |\mathcal{X}||\hat{\mathcal{X}}|n\right).$$

Now, we consider the dependence of  $\tau_{\max}$  on  $\epsilon_R$ . It follows from [48] that the error due to early termination of the Blahut algorithm is  $O\left(\frac{1}{\tau_{\max}}\right)$ . This implies that choosing  $\tau_{\max} = O\left(\frac{1}{\epsilon_R}\right)$  is sufficient. However, recent work has shown that a slight modification of the Blahut algorithm can drastically increase the convergence rate [50]. For this reason, we leave the number of iterations as the separate constant  $\tau_{\max}$  and do not consider its relationship to the error tolerance.

b. Complexity of Computing RDE Function.

Similarly, for each pair of parameters  $t < 0$  and  $s \geq 0$ , the complexity if we directly apply of the original Arimoto algorithm to compute the  $(R|_{s,t}, D|_{s,t})$  pair is  $O(\tau_{\max}|\mathcal{X}|^n|\hat{\mathcal{X}}|^n)$  where  $\tau_{\max}$  is the number of iterations. Instead, if the factored Arimoto algorithm (Algorithm 2) is employed, this complexity can also be reduced to  $O(\tau_{\max}|\mathcal{X}||\hat{\mathcal{X}}|^n)$ . In one of our proposed general algorithms in Section 3, we need to compute the RDE function for a pre-determined  $(R, D)$  pair. We use a nested bisection technique to find the Lagrange multipliers  $s, t$  that give the correct  $R$  and  $D$ .

- *Step 0:* Set  $t_{\min} < 0$  and  $s_{\max} > 0$  (e.g.,  $t_{\min} = -10$  and  $s_{\max} = 2$ )
- *Step 1:* If  $R|_{s_{\max}, t_{\min}} \leq R$ , set  $t_{\min} \leftarrow 2t_{\min}$  and repeat Step 1. Else go to Step 2.
- *Step 2:* Find  $t$  using the bisection method to obtain  $R|_{s_{\max}, t} = R$  within  $\epsilon_R$ . If  $D|_{s_{\max}, t} > D$ , go to Step 3. If  $D|_{s_{\max}, t} = D$  then stop. Else if  $D|_{s_{\max}, t} < D$ , set  $s_{\max} \leftarrow 2s_{\max}$  and go to Step 1.
- *Step 3:* Find  $s$  using the bisection method to get the correct distortion  $D$  within  $\epsilon_D$  while with each  $s$  doing the following steps
  - *Step 3a:* If  $R|_{s, t_{\min}} > R$ , go to Step 3c.
  - *Step 3b:* If  $R|_{s, t_{\min}} = R$ , then stop. Else if  $R|_{s, t_{\min}} < R$ , set  $t_{\min} \leftarrow 2t_{\min}$  and go to Step 1.
  - *Step 3c:* Find  $t$  using the bisection method to get the correct  $R$  within  $\epsilon_R$ .

The overall complexity of computing the RD function for a design rate  $R$  is therefore

$$O\left(\tau_{\max} \log_2\left(\frac{-t_{\min}}{\epsilon_R}\right) \log_2\left(\frac{s_{\max}}{\epsilon_D}\right) |\mathcal{X}||\hat{\mathcal{X}}|^n\right).$$

#### D. 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. Let  $\{\beta_1, \beta_2, \dots, \beta_n\}$  be a set of  $n$  distinct elements in  $\mathbb{F}_m$ . From each message polynomial  $f(X) = f_0 + f_1X + \dots + f_{k-1}X^{k-1}$  whose coefficients are in  $\mathbb{F}_m$ , we can obtain a codeword  $c = (c_1, c_2, \dots, c_n)$  by evaluating the message polynomial at  $\{\beta_i\}_{i=1}^n$ , i.e.,  $c_i = f(\beta_i)$  for  $i = 1, 2, \dots, n$ . Given a received vector  $\mathbf{r} = (r_1, r_2, \dots, r_n)$ , we can compute the *a posteriori* probability (APP) matrix  $\mathbf{\Pi}$  as follows:

$$[\mathbf{\Pi}]_{j,i} = \pi_{i,j} = \Pr(c_i = \alpha_j | r_i) \text{ for } 1 \leq i \leq n, 1 \leq j \leq m.$$

The ASD decoding as in [23] has the following main steps.

1. *Multiplicity Assignment*: Use a particular multiplicity assignment scheme (MAS) to derive an  $m \times n$  multiplicity matrix, denoted as  $\mathbf{M}$ , of non-negative integer entries  $\{M_{i,j}\}$  from the APP matrix  $\mathbf{\Pi}$ .
2. *Interpolation*: Construct a bivariate polynomial  $Q(X, Y)$  of minimum  $(1, k-1)$  weighted degree that passes through each of the point  $(\beta_j, \alpha_i)$  with multiplicity  $M_{i,j}$  for  $i = 1, 2, \dots, m$  and  $j = 1, 2, \dots, n$ .
3. *Factorization*: Find all polynomials  $f(X)$  of degree less than  $k$  such that  $Y - f(X)$  is a factor of  $Q(X, Y)$  and re-evaluate these polynomials to form a list of candidate codewords.

In this chapter, we denote  $\mu = \max_{i,j} M_{i,j}$  as the maximum multiplicity. Intuitively, higher multiplicity should be put on more likely symbols. A higher  $\mu$  generally allows

ASD decoding to achieve a better performance. However, one of the drawbacks of ASD decoding is that its decoding complexity is roughly  $O(n^2\mu^4)$  [51]. Even though there have been several reduced complexity variations and fast architectures as discussed in [52, 53, 54], the decoding complexity still increases rapidly with  $\mu$ . Thus, in this section we will mainly work with small  $\mu$  to keep the complexity affordable.

One of the main contributions of [23] 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

$$C_{\mathbf{M}} = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n M_{i,j} (M_{i,j} + 1).$$

**Condition 2.** *(ASD decoding threshold, see [23, 55, 51]). The transmitted codeword will be on the list if*

$$(a+1) \left[ S_{\mathbf{M}} - \frac{a}{2}(k-1) \right] > C_{\mathbf{M}} \quad (2.8)$$

for some  $a \in \mathbb{N}$  such that

$$a(k-1) < S_{\mathbf{M}} \leq (a+1)(k-1). \quad (2.9)$$

*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.*

### 1. Bit-level ASD Case

In this subsection, we consider multiple trials of ASD decoding using bit-level erasure patterns. A bit-level error pattern  $b_1^N \in \mathbb{Z}_2^N$  and a bit-level erasure pattern  $\hat{b}_1^N \in \mathbb{Z}_2^N$  have length  $N = n \times \eta$  since each symbol has  $\eta$  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 is applied. We also use  $B_1^N$  and  $\hat{B}_1^N$  to denote the random vectors which generate the realizations  $b_1^N$  and  $\hat{b}_1^N$ , respectively.

From each bit-level erasure pattern, we can specify entries of the multiplicity matrix  $\mathbf{M}$  using the bit-level MAS proposed in [55] 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  $\geq 2$  bits erased. All the other entries are zeros by default. This MAS has a larger decoding region compared to the conventional errors-and-erasures decoding scheme.

**Condition 3.** (*Bit-level ASD decoding threshold, see [55]*) For RS codes of rate  $\frac{k}{n} \geq \frac{2}{3} + \frac{1}{n}$ , ASD decoding using the bit-level MAS will succeed (i.e., the transmitted codeword is on the list) if

$$3\nu_b + e_b < \frac{3}{2}(n - k + 1) \quad (2.10)$$

where  $e_b$  is the number of bit-level erasures and  $\nu_b$  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 3.** *If we choose the bit-level letter-by-letter distortion measure  $\delta : \mathbb{Z}_2 \times$*

$\mathbb{Z}_2 \rightarrow \mathbb{R}_{\geq 0}$  as follows

$$\begin{aligned}\delta(0,0) &= 1, & \delta(0,1) &= 3, \\ \delta(1,0) &= 1, & \delta(1,1) &= 0,\end{aligned}$$

then the condition (2.10) becomes

$$d(b_1^N, \hat{b}_1^N) < \frac{3}{2}(n - k + 1). \quad (2.11)$$

*Proof.* The condition (2.10) can be seen to be equivalent to

$$\frac{2}{3}d(b_1^N, \hat{b}_1^N) < n - k + 1$$

using the same reasoning as in Proposition 1. The results then follows right away.  $\square$

**Remark 6.** We refer the multiple-decoding of bit-level ASD as *m-bASD*.

## 2. 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*) Consider a positive integer  $\ell \leq m$  where  $m$  is the number of elements in  $\mathbb{F}^m$ . For some codeword position, let us assign multiplicity  $m_j$  to the  $j$ -th most likely symbol for  $j = 1, 2, \dots, \ell$ . The remaining entries in the column are zeros by default. We call the sequence,  $(m_1, m_2, \dots, m_\ell)$ , the column multiplicity type for “top- $\ell$ ” 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 errors-and-erasures 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  $i$  is similar to erasing that position. Assigning the maximum multiplicity  $\mu$  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  $\hat{x}_1^n$  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 patterns and erasure patterns for ASD decoding*) Consider a MAS with  $T$  multiplicity types. Let  $\hat{x}_1^n \in \{1, 2, \dots, T\}^n$  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  $x_1^n \in \mathbb{Z}_{\ell+1}^n$  in Definition 3 applies unchanged here.

In our method, we generally choose an appropriate integer  $a$  in Condition 2 and design a distortion measure corresponding to the chosen  $a$  so that the condition for successful ASD decoding can be converted to the form where distortion is less than a fixed threshold. The following definition of allowable multiplicity types will lead us to the result of Lemma 1 and consequently,  $a \geq \mu$ , as stated in Corollary 1. Also, we want to find as many as possible multiplicity types since 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.

**Definition 7.** *The set of allowable multiplicity types for “top- $\ell$ ” decoding with max-*

imum multiplicity  $\mu$  is defined to be<sup>5</sup>

$$\mathcal{A}(\mu, \ell) \triangleq \left\{ (m_1, m_2, \dots, m_\ell) \left| \begin{array}{l} \sum_{j=1}^{\ell} m_j \leq \mu, \\ \sum_{j=1}^{\ell} m_j (\mu - m_j) \leq (\mu + 1) (|\{j : m_j \neq 0\}| - 1) \min_{j: m_j \neq 0} m_j \end{array} \right. \right\}. \quad (2.12)$$

We take the elements of this set in an arbitrary order and label them as  $1, 2, \dots, |\mathcal{A}(\mu, \ell)|$  with the convention that the multiplicity type 1 is always  $(\mu, 0, \dots, 0)$  which assigns the whole multiplicity  $\mu$  to the most likely symbol. The multiplicity type  $\hat{j}$  is denoted as  $(m_{1, \hat{j}}, m_{2, \hat{j}}, \dots, m_{\ell, \hat{j}})$ .

**Remark 7.** Multiplicity types  $(0, 0, \dots, 0), (1, 1, \dots, 1)$  as well as any permutations of  $(\mu, 0, \dots, 0)$  and  $(\lfloor \frac{\mu}{2} \rfloor, \lfloor \frac{\mu}{2} \rfloor, 0, \dots, 0)$  are always in the allowable set  $\mathcal{A}(\mu, \mu)$ . We use  $mASD-\mu$  to denote the proposed multiple ASD decoding using  $\mathcal{A}(\mu, \mu)$ .

**Example 4.** Consider  $mASD-2$ . In this case  $\mu = \ell = 2$  and we have  $\mathcal{A}(2, 2) = \{(2, 0), (1, 1), (0, 2), (0, 0)\}$  which comprises four allowable 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). We also consider a restricted set, called  $mASD-2a$ , that uses the set of multiplicity types  $\{(2, 0), (1, 1), (0, 0)\}$ .

**Example 5.** Consider  $mASD-3$ . In this case, the allowable set  $\mathcal{A}(3, 3)$  consists of all the permutations of  $(3, 0, 0), (0, 0, 0), (1, 1, 0), (2, 1, 0), (1, 1, 1)$ . We can see that the set  $\mathcal{A}(3, 2)$  consists of all permutations of  $(3, 0), (2, 1), (1, 1), (0, 0)$  and  $|\mathcal{A}(3, 2)| < |\mathcal{A}(3, 3)|$ .

---

<sup>5</sup>We use the convention that  $\min_{j: m_j \neq 0} m_j = 0$  if  $\{j : m_j \neq 0\} = \emptyset$ .

From now on, we assume that only allowable multiplicity types are considered throughout most of the chapter. With that setting in mind, we can obtain the following lemmas and theorems.

**Lemma 1.** *Consider a MAS( $\mu, \ell$ ) for “top- $\ell$ ” ASD decoding with multiplicity matrix  $\mathbf{M}$  that only uses multiplicity types in the allowable set  $\mathcal{A}(\mu, \ell)$ . Then, the score and the cost satisfy the following inequality*

$$2C_{\mathbf{M}} \geq (\mu + 1)S_{\mathbf{M}}.$$

*Proof.* Let us denote

$$e_{\hat{j}} = |\{i \in \{1, \dots, n\} : \hat{x}_i = \hat{j}\}|$$

to count the number of positions  $i$  that use multiplicity type  $\hat{j}$  for  $\hat{j} = 1, \dots, T$  and notice that  $\sum_{\hat{j}=1}^T e_{\hat{j}} = n$ . We also use

$$\nu_{j, \hat{j}} = |\{i \in \{1, \dots, n\} : x_i \neq j, \hat{x}_i = \hat{j}\}|$$

to count the number of positions  $i$  that use multiplicity type  $\hat{j}$  where the  $j$ -th most reliable symbol  $y_{i,j}$  is incorrect for  $j = 0, \dots, \ell$  and  $\hat{j} = 1, \dots, T$ . The notation

$$\chi_{j, \hat{j}} = |\{i \in \{1, \dots, n\} : x_i = j, \hat{x}_i = \hat{j}\}|$$

remains the same. Notice also that

$$e_{\hat{j}} = \sum_{j=0}^{\ell} \chi_{j, \hat{j}} \quad \text{and} \quad \chi_{j, \hat{j}} = e_{\hat{j}} - \nu_{j, \hat{j}}. \quad (2.13)$$

The score and the cost can therefore be written as

$$\begin{aligned} S_{\mathbf{M}}(\mathbf{c}) &= \sum_{j=1}^n M_{[c_j],j} \\ &= \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} \chi_{j,\hat{j}} \end{aligned} \quad (2.14)$$

$$= \mu \chi_{1,1} + \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} \chi_{j,\hat{j}} \quad (2.15)$$

$$= \mu \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} - \nu_{1,1} \right) + \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} (e_{\hat{j}} - \nu_{j,\hat{j}}) \quad (2.16)$$

and

$$\begin{aligned} C_{\mathbf{M}} &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n M_{i,j} (M_{i,j} + 1) \\ &= \frac{1}{2} \sum_{\hat{j}=1}^T e_{\hat{j}} \sum_{j=1}^{\ell} m_{j,\hat{j}} (m_{j,\hat{j}} + 1) \\ &= \frac{1}{2} \mu (\mu + 1) \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} \right) + \frac{1}{2} \sum_{\hat{j}=2}^T e_{\hat{j}} \sum_{j=1}^{\ell} m_{j,\hat{j}} (m_{j,\hat{j}} + 1) \end{aligned} \quad (2.17)$$

where (2.15) and (2.17) use the fact that the multiplicity type 1 is always assumed to be  $(\mu, 0, \dots, 0)$ .

Hence, we obtain

$$2C_{\mathbf{M}} - (\mu + 1)S_{\mathbf{M}} = \mu(\mu + 1)\nu_{1,1} + \sum_{\hat{j}=2}^T (\mu + 1) \sum_{j=1}^{\ell} m_{j,\hat{j}} \nu_{j,\hat{j}} - \sum_{\hat{j}=2}^T e_{\hat{j}} \sum_{j=1}^{\ell} m_{j,\hat{j}} (\mu - m_{j,\hat{j}}),$$

and therefore, since  $\mu$  and  $\nu_{1,1}$  are non-negative, Lemma 1 holds if we can show

$$(\mu + 1) \sum_{j=1}^{\ell} m_{j,\hat{j}} \nu_{j,\hat{j}} \geq e_{\hat{j}} \sum_{j=1}^{\ell} m_{j,\hat{j}} (\mu - m_{j,\hat{j}}) \quad (2.18)$$

for every  $\hat{j} = 2, \dots, T$ .

Next, we observe that

$$(\mu + 1) \sum_{\hat{j}=1}^{\ell} m_{j,\hat{j}} \nu_{j,\hat{j}} \geq (\mu + 1) \left( \sum_{j:m_{j,\hat{j}} \neq 0} \nu_{j,\hat{j}} \right) \min_{j:m_{j,\hat{j}} \neq 0} m_{j,\hat{j}} \quad (2.19)$$

and

$$\sum_{j:m_{j,\hat{j}} \neq 0} \nu_{j,\hat{j}} = \sum_{j:m_{j,\hat{j}} \neq 0} (e_{\hat{j}} - \chi_{j,\hat{j}}) \quad (2.20)$$

$$\begin{aligned} &= e_{\hat{j}} |\{j : m_{j,\hat{j}} \neq 0\}| - \sum_{j:m_{j,\hat{j}} \neq 0} \chi_{j,\hat{j}} \\ &\geq e_{\hat{j}} (|\{j : m_{j,\hat{j}} \neq 0\}| - 1) \end{aligned} \quad (2.21)$$

where (2.20) follows from (2.13) and (2.21) follows from

$$\sum_{j:m_{j,\hat{j}} \neq 0} \chi_{j,\hat{j}} \leq \sum_{j=0}^{\ell} \chi_{j,\hat{j}} = e_{\hat{j}}.$$

From (2.19) and (2.21), we have

$$(\mu + 1) \sum_{\hat{j}=1}^{\ell} m_{j,\hat{j}} \nu_{j,\hat{j}} \geq e_{\hat{j}} (\mu + 1) (|\{j : m_{j,\hat{j}} \neq 0\}| - 1) \min_{j:m_{j,\hat{j}} \neq 0} m_{j,\hat{j}} \quad (2.22)$$

and this motivates our definition of allowable multiplicity types.

Specifically, if we choose  $\{m_{1,\hat{j}}, m_{2,\hat{j}}, \dots, m_{\ell,\hat{j}}\}$  in the allowable set  $\mathcal{A}(\mu, \ell)$ , defined in (2.12), then by combining with (2.22), we obtain (2.18) and this completes the proof.  $\square$

**Corollary 1.** *With the setting as in Lemma 1, the integer  $a$  in Condition 2 must satisfy  $a \geq \mu$ .*

*Proof.* From  $(a + 1) \left[ S_{\mathbf{M}} - \frac{a}{2}(k - 1) \right] > C_{\mathbf{M}}$  and  $S_{\mathbf{M}} \leq (a + 1)(k - 1)$  in (2.8) and (2.9),

we know that

$$\begin{aligned} (a+1)S_{\mathbf{M}} - C_{\mathbf{M}} &> \frac{1}{2}a(a+1)(k-1) \\ &\geq \frac{1}{2}aS_{\mathbf{M}} \end{aligned}$$

and this implies that

$$2C_{\mathbf{M}} < (a+2)S_{\mathbf{M}}. \quad (2.23)$$

But, Lemma 1 states that  $2C_{\mathbf{M}} \geq (\mu+1)S_{\mathbf{M}}$ . Combining this with (2.23) gives a contradiction unless  $a > \mu - 1$ .  $\square$

In Condition 2, if we carefully design a distortion measure then for every  $a \geq \mu$ , the first constraint (2.8) can be equivalently converted to the form where distortion is smaller than a fixed threshold.

**Theorem 1.** *Consider an  $(n, k)$  RS code and a MAS $(\mu, \ell)$  for “top- $\ell$ ” decoding with multiplicity matrix  $\mathbf{M}$  that only uses  $T$  multiplicity types in the allowable set  $\mathcal{A}(\mu, \ell)$ . Consider an arbitrary integer  $a \geq \mu$ . Let  $\delta_a : \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathbb{R}_{\geq 0}$ , where in this case  $\mathcal{X} = \mathbb{Z}_{\ell+1}$  and  $\hat{\mathcal{X}} = \mathbb{Z}_{T+1} \setminus \{0\}$ , be a letter-by-letter distortion measure defined by  $\delta_a(x, \hat{x}) = [\Delta_a]_{x, \hat{x}}$ , where  $\Delta_a$  is the  $(\ell+1) \times T$  matrix<sup>6</sup>*

$$\Delta_a = \begin{pmatrix} \rho_{1,a} & \rho_{2,a} & \cdots & \rho_{T,a} \\ \rho_{1,a} - \frac{2m_{1,1}}{a} & \rho_{2,a} - \frac{2m_{1,2}}{a} & \cdots & \rho_{T,a} - \frac{2m_{1,T}}{a} \\ \rho_{1,a} - \frac{2m_{2,1}}{a} & \rho_{2,a} - \frac{2m_{2,2}}{a} & \cdots & \rho_{T,a} - \frac{2m_{2,T}}{a} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,a} - \frac{2m_{\ell,1}}{a} & \rho_{2,a} - \frac{2m_{\ell,2}}{a} & \cdots & \rho_{T,a} - \frac{2m_{\ell,T}}{a} \end{pmatrix} \quad (2.24)$$

---

<sup>6</sup>The first column of  $\Delta_a$  is  $[\frac{2\mu}{a}, 0, \frac{2\mu}{a}, \frac{2\mu}{a}, \dots, \frac{2\mu}{a}]^T$  since multiplicity type 1 is always chosen to be  $(\mu, 0, 0, \dots, 0)$ .

with

$$\rho_{\hat{j},a} = \frac{\mu(2a+1-\mu)}{a(a+1)} + \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}(m_{j,\hat{j}}+1)}{a(a+1)}$$

for  $\hat{j} = 1, \dots, T$ . Then, the equation (2.8) in Condition 2 is equivalent to

$$d(x_1^n, \hat{x}_1^n) < \frac{\mu(2a+1-\mu)}{a(a+1)}n - k + 1 \triangleq D_a,$$

and it is easy to verify that  $D_\mu = n - k + 1$ .

*Proof.* First, we show that  $\Delta_a$  consists of non-zero entries. It suffices to show that  $\rho_{\hat{j},a} \geq \frac{2m_{j,\hat{j}}}{a}$  for all  $j = 1, \dots, \ell$  and  $\hat{j} = 1, \dots, T$ , i.e.,

$$\mu(2a+1-\mu) + \sum_{j'=1}^{\ell} m_{j',\hat{j}}(m_{j',\hat{j}}+1) \geq 2m_{j,\hat{j}}(a+1)$$

which is equivalent to

$$2(a+1)(\mu - m_{j,\hat{j}}) + \sum_{j'=1}^{\ell} m_{j',\hat{j}}(m_{j',\hat{j}}+1) - \mu(\mu+1) \geq 0. \quad (2.25)$$

This is true since the left hand side of (2.25) is at least

$$2(\mu+1)(\mu - m_{j,\hat{j}}) + m_{j,\hat{j}}(m_{j,\hat{j}}+1) - \mu(\mu+1) = (\mu - m_{j,\hat{j}})(\mu+1 - m_{j,\hat{j}}) \geq 0.$$

With the same  $e_{\hat{j}}, \nu_{j,\hat{j}}, \chi_{j,\hat{j}}$  as defined in the proof of Lemma 1 and the chosen distortion matrix  $\Delta_a$ , we have

$$\begin{aligned} d(x_1^n, \hat{x}_1^n) &= \sum_{\hat{j}=1}^T \left( \sum_{j=1}^{\ell} \left( \rho_{j,a} - \frac{2m_{j,\hat{j}}}{a} \right) \chi_{j,\hat{j}} + \rho_{j,a} \chi_{0,\hat{j}} \right) \\ &= \sum_{\hat{j}=1}^T \left( \rho_{j,a} \sum_{j=0}^{\ell} \chi_{j,\hat{j}} - 2 \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{a} \chi_{j,\hat{j}} \right) \\ &= \sum_{\hat{j}=1}^T \left( \rho_{j,a} e_{\hat{j}} - 2 \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{a} \chi_{j,\hat{j}} \right). \end{aligned}$$

Noting that the first column of  $\Delta_a$  is always  $[\frac{2\mu}{a}, 0, \frac{2\mu}{a}, \frac{2\mu}{a}, \dots, \frac{2\mu}{a}]^T$  and  $\nu_{1,1} = e_1 - \chi_{1,1}$ ,

we obtain

$$d(x_1^n, \hat{x}_1^n) = \frac{2\mu}{a}\nu_{1,1} + \sum_{\hat{j}=2}^T \rho_{\hat{j},a} e_{\hat{j}} - 2 \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{a} \chi_{j,\hat{j}}. \quad (2.26)$$

Next, one can see that (2.8) can be rewritten as

$$\frac{2S_{\mathbf{M}}}{a} - k + 1 > \frac{2C_{\mathbf{M}}}{a(a+1)}$$

which, by substituting  $S_{\mathbf{M}}$  and  $C_{\mathbf{M}}$  in (2.16) and (2.17), is equivalent to

$$\begin{aligned} \frac{2\mu}{a} \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} - \nu_{1,1} \right) + 2 \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{a} \chi_{j,\hat{j}} - k + 1 \\ > \frac{\mu(\mu+1)}{a(a+1)} \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} \right) + \sum_{\hat{j}=2}^T e_{\hat{j}} \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}(m_{j,\hat{j}}+1)}{a(a+1)}. \end{aligned}$$

Equivalently, this gives

$$\begin{aligned} \left( \frac{2\mu}{a} - \frac{\mu(\mu+1)}{a(a+1)} \right) n - k + 1 \\ > \frac{2\mu}{a}\nu_{1,1} - 2 \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{a} \chi_{j,\hat{j}} + \sum_{\hat{j}=2}^T e_{\hat{j}} \left( \frac{2\mu}{a} - \frac{\mu(\mu+1)}{a(a+1)} + \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}(m_{j,\hat{j}}+1)}{\mu(\mu+1)} \right) \end{aligned}$$

which in turn is equivalent to

$$\frac{\mu(2a+1-\mu)}{a(a+1)} n - k + 1 > \frac{2\mu}{a}\nu_{1,1} + \sum_{\hat{j}=2}^T e_{\hat{j}} \rho_{\hat{j},a} - \frac{2}{a} \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} \chi_{j,\hat{j}}. \quad (2.27)$$

Finally, combining (2.26) and (2.27) gives the proof.  $\square$

**Example 6.** Consider *mASD-2* for  $a = \mu = 2$ . In this case, the distortion matrix is

$$\Delta = \begin{pmatrix} 2 & 5/3 & 2 & 1 \\ 0 & 2/3 & 2 & 1 \\ 2 & 2/3 & 0 & 1 \end{pmatrix}. \quad (2.28)$$

However, Condition 2 also requires the second constraint (2.9) to be satisfied. In addition, we need to choose an integer  $a \geq \mu$  in order to apply our proposed approach.

Therefore, we first consider the case of high-rate RS codes where if  $a = \mu$  then the satisfaction of (2.8) also implies the satisfaction of (2.9). For the case of lower-rate RS codes, we obtain a range of  $a$  and also propose a heuristic method to choose an appropriate  $a$ .

#### a. High-Rate Reed-Solomon Codes

In this subsection, we focus on high-rate RS codes which are usually seen in many practical applications. The high-rate constraint allows us to see that  $a = \mu$  is essentially the correct choice.

**Lemma 2.** *Consider an  $(n, k)$  RS code with rate  $\frac{k}{n} \geq \frac{1}{n} + \frac{\mu}{\mu+1}$ . If equation (2.8) is satisfied for  $a = \mu$ , or equivalently,  $d(x_1^n, \hat{x}_1^n) < n - k + 1$  under the distortion measure  $\Delta_\mu$ , then whole Condition 2 is satisfied and the transmitted codeword will be therefore on the list.*

*Proof.* Suppose (2.8) is satisfied for  $a = \mu$ , i.e.,

$$S_M > \frac{C_M}{\mu+1} + \frac{\mu}{2}(k-1). \quad (2.29)$$

We will show that

$$\mu(k-1) < S_M \quad (2.30)$$

$$\leq (\mu+1)(k-1) \quad (2.31)$$

and, therefore, both (2.8) and (2.9) in Condition 2 are satisfied for  $a = \mu$ .

Firstly, using Lemma 1 we have

$$\frac{S_M}{2} \geq S_M - \frac{C_M}{\mu+1}$$

and consequently, (2.30) is implied by (2.29) since

$$\frac{S_{\mathbf{M}}}{2} \geq S_{\mathbf{M}} - \frac{C_{\mathbf{M}}}{\mu + 1} > \frac{\mu}{2}(k - 1).$$

Secondly, note that (2.31) holds since

$$\begin{aligned} S_{\mathbf{M}} &= \mu \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} - \nu_{1,1} \right) + \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} (e_{\hat{j}} - \nu_{j,\hat{j}}) \\ &= \mu n - \mu \nu_{1,1} - \sum_{\hat{j}=2}^T \sum_{j=0}^{\ell} m_{j,\hat{j}} \nu_{j,\hat{j}} - \sum_{\hat{j}=2}^T e_{\hat{j}} \left( \mu - \sum_{j=1}^{\ell} m_{j,\hat{j}} \right) \\ &\leq \mu n \end{aligned} \tag{2.32}$$

$$\leq (\mu + 1)(k - 1) \tag{2.33}$$

where (2.32) is obtained by dropping non-negative terms and (2.33) follows from the high-rate constraint  $\frac{k-1}{n} \geq \frac{\mu}{\mu+1}$ .

Finally, by Theorem 1, one can verify that equation (2.8) with  $a = \mu$  is equivalent to

$$d(x_1^n, \hat{x}_1^n) < D_{\mu} = n - k + 1$$

under the distortion measure  $\Delta_{\mu}$ . □

However, there are possibly other integers  $a \neq \mu$  that can also satisfy Condition 2. If we consider higher-rate RS codes, as in the following theorem, then we can claim that  $a = \mu$  is the only such integer.

**Theorem 2.** *Consider an  $(n, k)$  RS code with rate  $\frac{k}{n} \geq \frac{1}{n} + \frac{\mu(\mu+3)}{(\mu+1)(\mu+2)}$ . The integer  $a$  in Condition 2 must satisfy  $a = \mu$  and, consequently, the set of constraints (2.8) and (2.9) in Condition 2 is equivalent to  $d(x_1^n, \hat{x}_1^n) < n - k + 1$  under the distortion measure  $\Delta_{\mu}$ .*

*Proof.* We first see that

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

in (2.8) implies

$$S_{\mathbf{M}} - \frac{a}{2}(k-1) > \frac{C_{\mathbf{M}}}{a+1}$$

and, with the score  $S_{\mathbf{M}}$  and the cost  $C_{\mathbf{M}}$  computed in (2.16) and (2.17), we obtain

$$\begin{aligned} \mu \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} - \nu_{1,1} \right) + \sum_{\hat{j}=2}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} (e_{\hat{j}} - \nu_{j,\hat{j}}) - \frac{a}{2}(k-1) \\ > \frac{\mu(\mu+1)}{2(a+1)} \left( n - \sum_{\hat{j}=2}^T e_{\hat{j}} \right) + \sum_{\hat{j}=2}^T e_{\hat{j}} \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}(m_{j,\hat{j}}+1)}{2(a+1)}. \end{aligned}$$

This gives

$$\begin{aligned} \left( \mu - \frac{\mu(\mu+1)}{2(a+1)} \right) n - \frac{a}{2}(k-1) > \mu \nu_{1,1} + \sum_{j=2}^T \sum_{j=1}^{\ell} \nu_{j,\hat{j}} \\ + \sum_{\hat{j}=2}^T e_{\hat{j}} \left( \mu - \sum_{j=1}^{\ell} m_{j,\hat{j}} + \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}(m_{j,\hat{j}}+1)}{2(a+1)} \right) \end{aligned} \quad (2.34)$$

$$\geq \sum_{\hat{j}=2}^T e_{\hat{j}} \left( \mu - \sum_{j=1}^{\ell} m_{j,\hat{j}} \right) \quad (2.35)$$

$$\geq 0 \quad (2.36)$$

where (2.35) is obtained by dropping non-negative terms.

Combining this inequality with the high-rate constraint implies that

$$\frac{\mu(2a+1-\mu)}{a(a+1)} > \frac{k-1}{n} \geq \frac{\mu(\mu+3)}{(\mu+1)(\mu+2)}$$

which leads to  $a < \mu + 1$ , i.e.  $a \leq \mu$ .

This, together with  $a \geq \mu$  according to Corollary 1, leave  $a = \mu$  as the only possible

choice. Finally, by seeing that

$$\frac{k}{n} \geq \frac{1}{n} + \frac{\mu(\mu+3)}{(\mu+1)(\mu+2)} > \frac{1}{n} + \frac{\mu}{\mu+1}$$

and applying Lemma 2 we conclude the proof.  $\square$

**Corollary 2.** *When the RD approach is used,  $R(D)$  is positive for  $D_{\min} \leq D < D_{\max}$  and is zero for  $D \geq D_{\max}$ . Computing  $D_{\max}$  reveals how good the distortion measure matrix is at rates close to zero (i.e., the erasure codebook has only one entry). For mASD- $\mu$ ,*

$$D_{\max}(\text{mASD-}\mu) = \sum_{i=1}^n \min_{\hat{j}=2, \dots, T} \left\{ 2(1 - p_{i,1}), \rho_{\hat{j}, \mu} - \sum_{j=1}^{\ell} \frac{m_{j, \hat{j}}}{\mu} p_{i,j} \right\}$$

while for mBM- $\ell$ ,

$$D_{\max}(\text{mBM-}\ell) = \sum_{i=1}^n \min\{1, 2(1 - p_{i,1})\}.$$

Moreover, if mASD- $\mu$  uses multiplicity type  $(0, 0, \dots, 0)$  then

$$D_{\max}(\text{mASD-}\mu) \leq D_{\max}(\text{mBM-}\ell)$$

for every  $\mu, \ell$ .

*Proof.* See Appendix 1.  $\square$

**Example 7.** *Consider mASD-2 with distortion matrix in (2.28). We have*

$$D_{\max}(\text{mASD-}2) = \sum_{i=1}^n \min \left\{ 1, 2(1 - p_{i,1}), \frac{5}{3} - \frac{2}{3}(p_{i,1} + p_{i,2}) \right\}$$

which is less than or equal to  $D_{\max}(\text{mBM-}\ell)$  for every  $\ell$ . This predicts that, as expected, ASD decoding will be superior when  $R$  is small.

Table I. Example ranges of possible  $a$ 

	RS(255,191)	RS(255,127)
$\mu = 2$	$2 \leq a \leq 3$	$2 \leq a \leq 6$
$\mu = 3$	$3 \leq a \leq 5$	$3 \leq a \leq 9$

b. Lower-Rate Reed-Solomon Codes

Without the high-rate constraint as in Theorem 2, we may not have  $a = \mu$ . However, we can obtain a range for  $a$  and heuristically choose the integer  $a$  that potentially give the highest rate-distortion exponent. After that, we can also apply the algorithms proposed in Section 3 with the corresponding distortion measure  $\Delta_a$  and distortion threshold  $D_a$  derived in Theorem 1.

The following lemma tells us the range of possible  $a$ .

**Lemma 3.** *Consider an  $(n,k)$  RS code. In order to satisfy (2.8), one must have*

$$\mu \leq a \leq \left\lceil \mu\theta - 1/2 + \sqrt{\mu^2\theta(\theta-1) + 1/4} \right\rceil - 1$$

where  $\theta \triangleq \frac{n}{k-1}$ .

*Proof.* First note that (2.36) holds for any  $(n,k)$ . Therefore, we have

$$\mu - \frac{\mu(\mu+1)}{2(a+1)} > \frac{a(k-1)}{2n}.$$

Combining this with  $a \geq \mu$  in Corollary 1, we obtain the stated result.  $\square$

**Example 8.** *Table I gives several example ranges of possible  $a$  for some choices of  $\mu$  and RS codes.*

Among possible choices of  $a$ , we are interested in choosing  $a$  that gives the largest rate-distortion exponent and therefore has a better chance to satisfy Condition 2. The

following lemma can give us an insight of how to choose such an integer  $a$ .

**Lemma 4.** *If*

$$a > \frac{1}{2} \left( \sqrt{1 + 4\theta\mu(\mu + 1)} - 3 \right) \quad (2.37)$$

where  $\theta = \frac{n}{k-1}$  then starting from  $a$ , the rate-distortion exponent  $F_a$  strictly decreases until reaching zero, i.e.,  $F_a > F_{a+1} > F_{a+2} > \dots \geq 0$  if rate  $R$  is fixed.

*Proof.* For a fixed rate  $R$ , the distortion measure  $\Delta_{a+1}$  and distortion  $D_{a+1}$  yield exponent  $F_{a+1}$ . Scaling both  $\Delta_{a+1}$  and  $D_{a+1}$  leaves  $F_{a+1}$  unchanged. Hence,  $\frac{a+1}{a}\Delta_{a+1}$  and  $\frac{a+1}{a}D_{a+1}$  also yield  $F_{a+1}$ . Next, we will show that

$$\frac{a+1}{a}\Delta_{a+1} \geq \Delta_a. \quad (2.38)$$

To prove (2.38), it suffices to show

$$\frac{a+1}{a}\rho_{\hat{j},a+1} \geq \rho_{\hat{j},a} \quad (2.39)$$

since

$$\frac{a+1}{a} \left( \rho_{\hat{j},a+1} - \frac{2m_{j,\hat{j}}}{a+1} \right) \geq \rho_{\hat{j},a} - \frac{2m_{j,\hat{j}}}{a}$$

is also equivalent to (2.39).

Equivalently, we need to show

$$\mu(\mu + 1) \geq \sum_{j=1}^{\ell} m_{j,\hat{j}}(m_{j,\hat{j}} + 1)$$

which is true because  $\mu \geq \sum_{i=1}^{\ell} m_{j,\hat{j}}$  by the definition of allowable multiplicity types.

Thus, (2.38) holds and, therefore, the exponent yielded by  $\Delta_a$  and  $\frac{a+1}{a}D_{a+1}$  is at

Table II. Example ranges of  $a$  that gives the largest exponent

	RS(255,191)	RS(255,127)
$\mu = 2$	$a = 2$	$a \in \{2, 3\}$
$\mu = 3$	$a = 3$	$a \in \{3, 4\}$
$\mu = 12$	$a \in \{12, 13\}$	$12 \leq a \leq 17$

least  $F_{a+1}$ . From (2.37) we have

$$\begin{aligned}
D_a &= \frac{\mu(2a+1-\mu)}{a(a+1)}n - k + 1 \\
&> \frac{\mu(2a+3-\mu)}{a(a+2)}N - \frac{a+1}{a}(k-1) \\
&= \frac{a+1}{a}D_{a+1}.
\end{aligned}$$

Since for a fixed  $R$ , exponent  $F$  is increasing in distortion  $D$  [45, Thm 6.6.2], we know that  $F_a > F_{a+1}$  where  $F_a$  is the exponent yielded by  $\Delta_a$  and  $D_a$ .  $\square$

**Corollary 3.** *The integer  $a$  that gives the largest exponent lies in the range*

$$\mu \leq a \leq \left\lfloor \frac{1}{2} \left( \sqrt{1 + 4\theta\mu(\mu+1)} - 3 \right) \right\rfloor + 1.$$

**Example 9.** *Table II presents several example ranges of  $a$  that gives the largest exponent for some choices of  $\mu$  and RS codes.*

**Remark 8.** *Simulation results also confirm our analysis. For example, in Fig. 3,  $a = 3$  and  $a = 4$  give roughly same and the largest exponents for  $\mu = 3$  while  $a = 2$  yields the largest exponent for  $\mu = 2$ . In fact, simulation results suggest that, typically, either  $a = \mu$  or  $a = \mu + 1$  gives the best exponent.*

In Condition 2, for lower-rate RS codes, so far we have only paid attention to

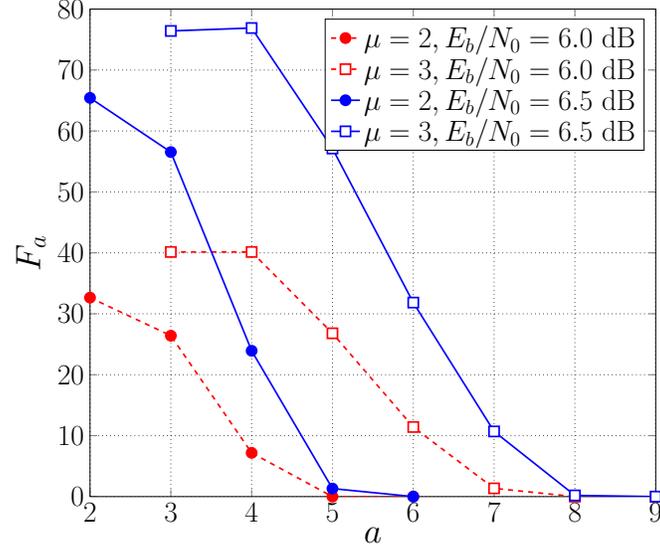


Fig. 3. Plot of exponent  $F_a$  versus  $a$  for  $\mu = 2$  and  $\mu = 3$  with a fixed rate  $R = 6$ . Simulations are conducted for the (255,127) RS code using BPSK over an AWGN channel at  $E_b/N_0 = 6.0$  dB and 6.5 dB.

(2.8). However, it is also required that

$$a(k-1) < S_M \leq (a+1)(k-1),$$

or equivalently

$$a+1 = \left\lceil \frac{S_M}{k-1} \right\rceil. \quad (2.40)$$

While it is hard to tell exactly which  $a$  will satisfy (2.40) with high probability right away, we can propose a heuristic method to choose the integer  $a$  that is likely to work.

We first need the following lemma.

**Lemma 5.** *Suppose we have obtained a test-channel input-probability distribution matrix  $\mathbf{Q}$  (e.g., during Step 2a or Step 2b in the proposed algorithms in Section 3) and the set of erasure patterns for mASD is generated independently and randomly*

according to  $\mathbf{Q}$ . Then, the expected score can be computed as follows:

$$\mathbb{E}[S_{\mathbf{M}}] = \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} \sum_{i=1}^n m_{j,\hat{j}} p_{i,j} q_{i,\hat{j}}. \quad (2.41)$$

*Proof.* The proof follows from the following equations:

$$\begin{aligned} \mathbb{E}[S_{\mathbf{M}}] &= \mathbb{E} \left[ \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} \chi_{j,\hat{j}} \right] \\ &= \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} \mathbb{E}[\chi_{j,\hat{j}}] \\ &= \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} m_{j,\hat{j}} \mathbb{E} \left[ \sum_{i=1}^n \mathbf{1}_{\{X_i=j, \hat{X}_i=\hat{j}\}} \right] \\ &= \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} \sum_{i=1}^n m_{j,\hat{j}} \Pr(X_i = j, \hat{X}_i = \hat{j}) \\ &= \sum_{\hat{j}=1}^T \sum_{j=1}^{\ell} \sum_{i=1}^n m_{j,\hat{j}} p_{i,j} q_{i,\hat{j}} \end{aligned} \quad (2.42)$$

where  $\mathbf{1}_{\mathcal{S}}$  denotes the indicator function of an event  $\mathcal{S}$  and (2.42) is implied by (2.14). □

Next, we propose a heuristic method to find the appropriate integer  $a$  to work with as follows.

**Algorithm 3.**

- *Step 1: Start with  $a = \mu$ , using distortion measure  $\Delta_a$  and distortion threshold  $D_a$  to get the corresponding distribution matrix  $\mathbf{Q}$  as discussed above.*
- *Step 2: Compute the expected score  $\mathbb{E}[S_{\mathbf{M}}]$  using (2.41). If  $\left\lceil \frac{\mathbb{E}[S_{\mathbf{M}}]}{k-1} \right\rceil = a + 1$  then output  $a$  and stop. If not set  $a \leftarrow a + 1$  and return to Step 1.*

**Remark 9.** *In simulations with small to moderate  $\mu$ , it is usually found that  $a$  is either  $\mu$  or  $\mu + 1$ . Typically,  $\frac{\mathbb{E}[S_{\mathbf{M}}]}{k-1} > \mu$  and a unit increase of  $a$  produces a small increase in  $\frac{\mathbb{E}[S_{\mathbf{M}}]}{k-1}$ .*

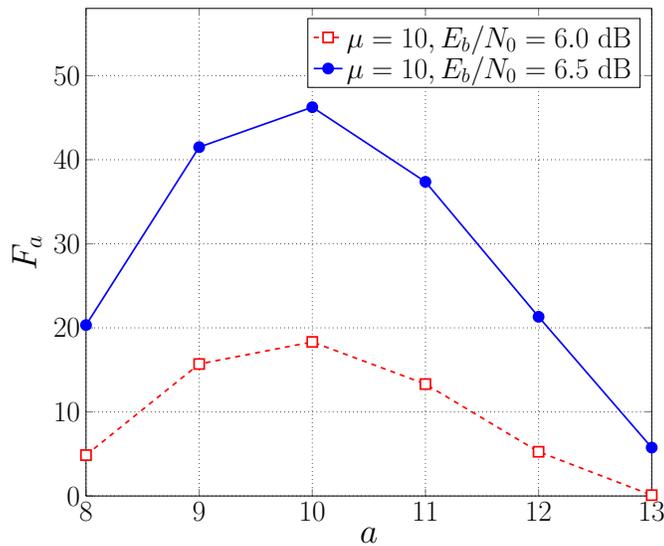


Fig. 4. Plot of exponent  $F_a$  versus  $a$  for  $\mu = 10$  with a fixed rate  $R = 6$ . The set of multiplicity types considered is the relaxed set  $\mathcal{A}_0(10, 2)$ . Simulations are conducted for the (458,410) RS code over  $\mathbb{F}_{2^{10}}$  using BPSK over an AWGN channel at  $E_b/N_0 = 6.0$  dB and 6.5 dB.

So far, we have considered only the allowable multiplicity types in Definition 7. It is possible to obtain better performance if we relax some constraints and allow multiplicity types to be in the relaxed set

$$\mathcal{A}_0(\mu, \ell) \triangleq \left\{ (m_1, m_2, \dots, m_\ell) \mid \sum_{j=1}^{\ell} m_j \leq \mu \right\}.$$

In this case, some theoretical results, e.g., results in Lemma 1 and Theorem 2, do not hold. However, this modification combined with the heuristic method above can improve the decoding performance, especially with large  $\mu$ . Specifically, we consider  $\text{mASD}_{0-\mu}$  which denotes our proposed multiple ASD decoding algorithm that only uses multiplicity types  $(0, 0)$  and  $(m_1, m_2)$  of the form  $m_1 + m_2 = \mu$ . These multiplicity types form a subset of  $\mathcal{A}_0(\mu, 2)$ . The choice of  $\ell = 2$  is suggested by observations that top-2 decoding performs almost as good as top- $\ell$  decoding for  $\ell > 2$ . The integer  $a$  used in  $\text{mASD}_{0-\mu}$  is found through the heuristic method. In Fig. 4, simulations are

conducted for the (458,410) RS code using BPSK over an AWGN channel. For  $\mu = 10$ , it can again be observed that  $a = \mu$  gives the best exponent. More simulation results of this heuristic method can be seen in Section G.

## E. Closed-Form Analysis of RD and RDE Functions for Some Distortion Measures

### 1. Closed-Form RD Function

For some simple distortion measures, we can compute the RD functions analytically in closed form. First, we observe an error pattern as a sequence of i.n.d. random source components. Then, we compute the component RD functions at each index of the sequence and use convex optimization techniques to allocate the total rate and distortion to various components. This method converges to the solution faster than the numerical method in Section C. The following two theorems describe how to compute the RD functions for the simple distortion measures of Proposition 1 and 3.

**Lemma 6.** *Consider a binary source  $X$  where  $\Pr(X = 1) = p$  and  $\Pr(X = 0) = 1 - p$ . With the distortion measure in (2.2), the rate-distortion function for this source is<sup>7</sup>*

$$R(D) = [H_2(p) - H_2(D + p - 1)]^+.$$

*Proof.* See Appendix 2. □

**Theorem 3.** *(Conventional errors-and-erasures “mBM-1” decoding) Let  $p_{i,1} \triangleq \Pr(X_i = 1)$  for  $i = 1, \dots, n$ . The overall rate-distortion function is given by*

$$R(D) = \sum_{i=1}^n [H_2(p_{i,1}) - H_2(\tilde{D}_i)]^+$$

---

<sup>7</sup>Here  $[x]^+$  denotes the non-negative part of  $x$ , i.e.,  $[x]^+ = \begin{cases} x & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$

where  $\tilde{D}_i \triangleq D_i + p_{i,1} - 1$  and  $\tilde{D}_i$  can be found by a reverse water-filling procedure (see [41, Theorem 13.3.3]):

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

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

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

*Proof.* See Appendix 3. □

**Theorem 4.** (*Bit-level ASD “m-bASD” decoding*) Let  $r_{i,1} \triangleq \Pr(B_i = 1)$  and  $r_{i,0} \triangleq \Pr(B_i = 0)$  for  $i = 1, 2, \dots, N$ . The overall rate-distortion function in m-bASD scheme is given by

$$R(D) = \sum_{i=1}^N [R_i(\lambda)]^+ \quad (2.43)$$

where

$$R_i(\lambda) = H_2(r_{i,1}) - H_2\left(\frac{1 + \lambda}{1 + \lambda + \lambda^2}\right) + \left(r_{i,1} - \frac{1 + \lambda}{1 + \lambda + \lambda^2}\right) H_2\left(\frac{\lambda}{1 + \lambda}\right) \quad (2.44)$$

and the distortion component  $D_i$  is given by

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

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 following test-channel input-probability distribution

$$s_{i,0} \triangleq \Pr(\hat{B}_i = 0) = \frac{(1 + \lambda) - r_{i,1}(1 + \lambda + \lambda^2)}{1 - \lambda^2} \quad (2.45)$$

and

$$s_{i,1} \triangleq \Pr(\hat{B}_i = 1) = \frac{r_{i,1}(1 + \lambda + \lambda^2) - \lambda(1 + \lambda)}{1 - \lambda^2}. \quad (2.46)$$

*Sketch of proof.* With the distortion measure in (3), using the method in [47, Chapter 2] we can compute the rate-distortion function components

$$R_i(\lambda_i) = H_2(r_{i,1}) - H_2\left(\frac{1 + \lambda_i}{1 + \lambda_i + \lambda_i^2}\right) + \left(r_{i,1} - \frac{1 + \lambda_i}{1 + \lambda_i + \lambda_i^2}\right) H_2\left(\frac{\lambda_i}{1 + \lambda_i}\right)$$

where  $\lambda_i$  is a Lagrange multiplier such that

$$D_i = \frac{1 + 2\lambda_i + 3\lambda_i^2}{1 + \lambda_i + \lambda_i^2} - r_{i,1} \frac{1 + 2\lambda_i}{1 + \lambda_i}$$

for each bit index  $i$ . Then, the Kuhn-Tucker conditions define the overall rate allocation using the similar argument as in the proof of Theorem 3.  $\square$

**Remark 10.** While the RD function for mBM-1 as in Theorem 3 can be computed by strictly following a water-filling schedule, the RD function for m-bASD in Theorem 4 can also be found by a similar algorithm that converges to the true solution in a finite number of steps. The detail of this algorithm and related discussions are left to Appendix 5.

## 2. Closed-form RDE function

In this subsection, we consider the case mBM-1 whose distortion measure is given in (2.2). 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  $r, c \in \mathbb{F}_m$ . For this channel model, we restrict our attention to the range of  $p$  where the received symbol is the most-likely (i.e.,  $p > (1-p)/(m-1)$ ). 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  $p_{i,1} = \Pr(X_i = 1) = p$  for every index  $i$  of the error pattern  $x_1^n$ . That means, in this context we have a source  $x_1^n$  with i.i.d. binary components  $x_i$ . Since the components  $x_i$ 's are i.i.d, we can treat each  $x_i$  as a binary source  $X$  with  $\Pr(X = 1) = p$  and first compute the RDE function for this source  $X$  as given by an analysis in Appendix 4. Based on this analysis, we obtain the following lemmas and theorems for the mBM-1 decoding algorithm of RS codes over an  $m$ -SC channel.

**Lemma 7.** *Let  $h(u) = H_2(u) - H_2(u + D - 1)$  map  $u \in [1 - D, 1 - \frac{D}{2})$  to  $R$ . Then, the inverse mapping of  $h$ ,*

$$h^{-1} : (0, H_2(1 - D)] \rightarrow \left[1 - D, 1 - \frac{D}{2}\right),$$

*is well-defined and maps  $R$  to  $u$ .*

*Proof.* We first notice that  $h(u)$  is strictly decreasing since the derivative is negative over  $[1 - D, 1 - \frac{D}{2})$ , hence the mapping  $h : [1 - D, 1 - \frac{D}{2}) \rightarrow (0, H_2(1 - D)]$  is one-to-one. From the analysis in Appendix 4, one can also see that  $h$  is onto.  $\square$

**Theorem 5.** *Using mBM-1 with  $2^R$  decoding attempts where  $R \in (0, nH_2(1 - \frac{D}{n})]$ , the maximum rate-distortion exponent that can be achieved is<sup>8</sup>*

$$F = n D_{\text{KL}} \left( h^{-1} \left( \frac{R}{n} \right) \parallel p \right). \quad (2.47)$$

*Proof.* First, note that in our context where we have a source sequence  $x_1^n$  of  $n$  i.i.d. source components, the rate and exponent for each source component are now  $\frac{R}{n}$  and

---

<sup>8</sup>The Kullback-Leibler divergence is  $D_{\text{KL}}(u||p) \triangleq u \log_2 \frac{u}{p} + (1 - u) \log_2 \frac{1-u}{1-p}$ .

$\frac{F}{n}$ . From Case 3 in Appendix 4 and from Lemma 7, we have

$$\frac{F}{n} = D_{\text{KL}}(u||p) = D_{\text{KL}}\left(h^{-1}\left(\frac{R}{n}\right) \parallel p\right)$$

and the theorem follows.  $\square$

**Lemma 8.** *Let  $g(u) = D_{\text{KL}}(u||p)$  map  $u \in [1 - D, p]$  to  $F$ . Then, the inverse mapping of  $g$ ,*

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

*is well-defined and maps  $F$  to  $u$ .*

*Proof.* We first see that  $g(u)$  is a strictly convex function and achieves minimum value at  $u = p$  and therefore  $g(u)$  is strictly decreasing over  $[1 - D, p]$ . Thus, the mapping  $g : [1 - D, p] \rightarrow [0, D_{\text{KL}}(1 - D || p)]$  is one-to-one. From the analysis in Appendix 4, one can also see that  $g$  is onto.  $\square$

**Theorem 6.** *To achieve a rate-distortion exponent of  $F \in [0, n D_{\text{KL}}(1 - D || p)]$ , the minimum number of decoding attempts required for mBM-1 is  $2^R$  where*

$$R = n \left[ H_2\left(g^{-1}\left(\frac{F}{n}\right)\right) - H_2\left(g^{-1}\left(\frac{F}{n}\right) + \frac{D}{n} - 1\right) \right]^+$$

*Proof.* We also note that the rate, distortion and exponent for each source component are  $\frac{R}{n}$ ,  $\frac{D}{n}$  and  $\frac{F}{n}$  respectively. Combining all the cases in Appendix 4, we have

$$\frac{R}{n} = \left[ H_2\left(g^{-1}\left(\frac{F}{n}\right)\right) - H_2\left(g^{-1}\left(\frac{F}{n}\right) + \frac{D}{n} - 1\right) \right]^+$$

and the theorem follows.  $\square$

**Remark 11.** *In Fig. 5, we simulate the performance of mBM-1(RDE,11) for the (255,239) RS code over an m-SC channel. One curve reflects the simulated frame-*

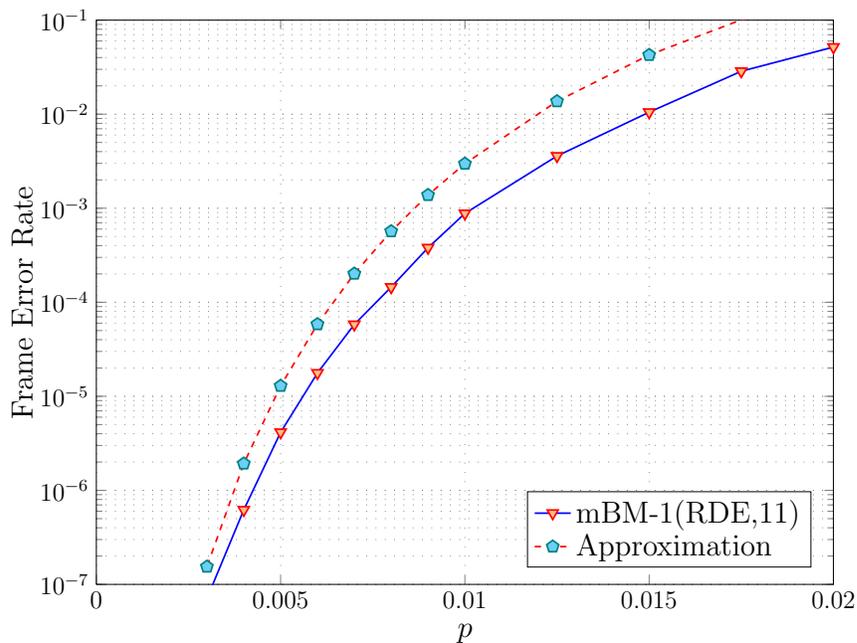


Fig. 5. Performance of mBM-1(RDE,11) and its approximation  $2^{-F}$  where  $F$  is given in (2.47) for the (255,239) RS code over an  $m$ -SC( $p$ ) channel.

error rate (FER) and the other is the approximation derived from  $2^{-F}$  where  $F$  is given in (2.47) with  $R = 11$ .

## F. Some Extensions

### 1. Erasure Patterns Using Covering Codes

The RD framework we use is most suitable when  $n \rightarrow \infty$ . For a finite  $n$ , choosing random codes for only a few LRPs can be risky. We can instead use good covering codes to handle these LRPs. In the scope of covering problems, one can use an  $\ell$ -ary  $t_c$ -covering code (e.g., a perfect Hamming or Golay code) with covering radius  $t_c$  to cover the whole space of  $\ell$ -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. The method of using covering codes in the LRPs was proposed earlier in

[56] to choose the test patterns in iterative bounded distance decoding algorithms for binary linear block codes.

In order take care of up to the  $\ell$  most likely symbols at each of the  $n_c$  LRPs of an  $(n, k)$  RS, we consider an  $(n_c, k_c)$   $\ell$ -ary  $t_c$ -covering code whose codeword alphabet is  $\mathbb{Z}_{\ell+1} \setminus \{0\} = \{1, 2, \dots, \ell\}$ . 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 errors-only decoding is used. For errors-only decoding, Condition 1 for successful decoding becomes

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

**Definition 8.** (*Error patterns and erasure patterns for errors-only decoding*) Let us define  $x_1^n \in \mathbb{Z}_{\ell+1}^n$  as an error pattern where, at index  $i$ ,  $x_i = j$  implies that the  $j$ -th most likely symbol is correct for  $j \in \{1, 2, \dots, \ell\}$ , and  $x_i = 0$  implies none of the first  $\ell$  most likely symbols is correct. Let  $\hat{x}_1^n \in \{1, 2, \dots, \ell\}^n$  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  $j \in \{1, 2, \dots, \ell\}$ .

**Proposition 4.** If we choose the letter-by-letter distortion measure  $\delta : \mathbb{Z}_{\ell+1} \times \mathbb{Z}_{\ell+1} \setminus \{0\} \rightarrow \mathbb{R}_{\geq 0}$  defined by  $\delta(x, \hat{x}) = [\Delta]_{x, \hat{x}}$  in terms of the  $(\ell + 1) \times \ell$  matrix

$$\Delta = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 0 \end{pmatrix} \quad (2.48)$$

then the condition for successful errors-only decoding then becomes

$$d(x_1^n, \hat{x}_1^n) < \frac{1}{2}(n - k + 1). \quad (2.49)$$

*Proof.* It follows directly from

$$d(x_1^n, \hat{x}_1^n) = \sum_{\hat{j}=1}^{\ell} \sum_{j=0, j \neq \hat{j}}^{\ell} \chi_{j, \hat{j}} = \nu.$$

□

**Remark 12.** *If we delete the first row which corresponds to the case where none of the first  $\ell$  most likely symbols is correct then the distortion measure is exactly the Hamming distortion.*

**Split covering approach:** In this approach, one breaks an error pattern  $x_1^n$  into two sub-error patterns  $x^{\text{LRPs}} \triangleq x_{\sigma(1)}x_{\sigma(2)}\dots x_{\sigma(n_c)}$  of  $n_c$  least reliable positions and  $x^{\text{MRPs}} \triangleq x_{\sigma(n_c+1)}\dots x_{\sigma(n)}$  of  $n - n_c$  most reliable positions. Similarly, one can break an erasure pattern  $\hat{x}_1^n$  into two sub-erasure patterns  $\hat{x}^{\text{LRPs}} \triangleq \hat{x}_{\sigma(1)}\hat{x}_{\sigma(2)}\dots \hat{x}_{\sigma(n_c)}$  and  $\hat{x}^{\text{MRPs}} \triangleq \hat{x}_{\sigma(n_c+1)}\dots \hat{x}_{\sigma(n)}$ . Let  $z_{n_c}$  be the number of positions in the  $n_c$  LRPs where none of the first  $\ell$  most likely symbols is correct, or

$$z_{n_c} = \left| \left\{ i = 1, 2, \dots, n_c : x_{\sigma(i)} = 0 \right\} \right|.$$

If we assign the set of all sub-error patterns  $\hat{x}^{\text{LRPs}}$  to be an  $(n_c, k_c)$   $t_c$ -covering code then

$$d(x^{\text{LRPs}}, \hat{x}^{\text{LRPs}}) \leq t_c + z_{n_c}$$

because this covering code has covering radius  $t_c$ . Since

$$d(x_1^n, \hat{x}_1^n) = d(x^{\text{LRPs}}, \hat{x}^{\text{LRPs}}) + d(x^{\text{MRPs}}, \hat{x}^{\text{MRPs}}),$$

in order to increase the probability that the condition (2.49) is satisfied we want to make  $d(x^{\text{MRPs}}, \hat{x}^{\text{MRPs}})$  as small as possible by the use of the RD approach. The following proposition summarizes how to generate a set of  $2^R$  erasure patterns for multiple runs of errors-only decoding.

**Proposition 5.** *In each erasure pattern, the letter sequence at  $n_c$  LRPs is set to be a codeword of an  $(n_c, k_c)$   $\ell$ -ary  $t_c$ -covering code. The letter sequence of the remaining  $n - n_c$  MRPs is generated randomly by the RD method (see Section 3) with rate  $R_{\text{MRPs}} = R - k_c \log_2 \ell$  and the distortion measure in (2.48). Since this covering code has  $\ell^{k_c}$  codewords, the total rate is  $R_{\text{MRPs}} + \log_2 \ell^{k_c} = R$ .*

**Example 10.** *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  $\hat{x}_8\hat{x}_9 \dots \hat{x}_n$  are generated randomly using the RD approach with rate  $(R - 4)$ .*

**Remark 13.** *While it also makes sense to use a covering codes for the  $n_c$  LRPs of the erasure patterns and set 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 usually be improved by using a combination of a covering code and a random (i.e., generated by the RD approach) code. More discussions are presented in Section G.*

## 2. A Single Decoding Attempt

In this subsection, we investigate a special case of our proposed RDE framework when  $R = 0$  (i.e., the set of erasure patterns consists of one pattern). In this case, our proposed approach is related to another line of work where one tries to design

a good erasure pattern for a single BM decoding or a good multiplicity matrix for a single ASD decoding [34, 36, 35, 37]. We will see that the RDE approach for  $R = 0$  is quite similar to optimizing a Chernoff bound [36, 35] or using the method of types [37]. The main difference is that this approach starts from Condition 2 rather than its large multiplicity approximation.

**Lemma 9.** *When rate  $R = 0$ , the distribution matrix  $\mathbf{Q}$  that optimizes the RDE/RD function consists of only binary entries. Consequently, the random codebook using the proposed RDE approach (the set of erasure patterns) becomes a single deterministic pattern.*

*Sketch of proof.* For each  $(s, t)$  pair, the total rate is the sum of  $n$  individual components as seen in Proposition 2. Therefore, the zero total rate implies all components are zero. Thus, it suffices to show that if an arbitrary rate component (denoted as  $R$  in the proof) is zero then the corresponding column of  $\mathbf{Q}$  has all entries equal to 0 or 1.

For the RD case, it is well known [47, p. 27] that if  $R = 0$  then the distortion is given by  $D_{\max} = \min_{\hat{j}} \sum_j p_j \delta_{j\hat{j}}$  where  $\hat{j}^*$  is the argument that achieves this minimum and the test-channel input distribution is

$$q_{\hat{j}}^* = \begin{cases} 1 & \text{if } \hat{j} = \hat{j}^* \\ 0 & \text{otherwise} \end{cases}.$$

Computing the RDE for the source distribution  $p_j$  is equivalent to solving the RD problem for an appropriately tilted source distribution  $\tilde{p}_j^*$ . Therefore, the above property is inherited by the RDE as well. In particular, the distortion at  $R = 0$  is given by  $\min_{\hat{j}} \sum_j \tilde{p}_j^* \delta_{j\hat{j}}$  and the test-channel input distribution is supported on the singleton element that achieves this minimum.

This result can also be shown directly by solving (2.6) while dropping the rate constraint from (2.7).  $\square$

Let  $G_{\hat{j}}(D)$  be the large deviation rate-function for the distortion when the reconstruction symbol is fixed to  $\hat{j}$ . It is well-known that this can be computed using either a Chernoff bound or the method of types [41]. Both techniques result in the same function; for  $\alpha \geq 0$ , it is described implicitly by

$$D(\alpha) = \frac{\sum_j p_j 2^{\alpha \delta_{j,\hat{j}}} \delta_{j,\hat{j}}}{\sum_{j'} p_{j'} 2^{\alpha \delta_{j',\hat{j}}}},$$

$$G_{\hat{j}}(\alpha) = \sum_j \frac{p_j 2^{\alpha \delta_{j,\hat{j}}}}{\sum_{j'} p_{j'} 2^{\alpha \delta_{j',\hat{j}}}} \log_2 \frac{2^{\alpha \delta_{j,\hat{j}}}}{\sum_{j'} p_{j'} 2^{\alpha \delta_{j',\hat{j}}}}.$$

**Theorem 7.** *The RDE function for  $R = 0$  is equal to*

$$F(0, D) = \max_{\hat{j}} G_{\hat{j}}(D).$$

*Proof.* Lemma 9 shows that the reconstruction distribution must be supported on a single element. Since the exponential failure probability for any fixed reconstruction symbol follows from a standard large-deviations analysis, the only remaining degree of freedom is which symbol to use. Choosing the best symbol maximizes the RDE.  $\square$

**Remark 14.** *This means that the single decoding attempt with the best error-exponent can be computed as a special case of the RDE approach. Simplifying our proposed algorithm to use the single Lagrange multiplier  $\alpha$  leads to an algorithm that is very similar to the one proposed in [37]. It also seems unlikely that this new algorithm will provide any significant performance gains either in performance or complexity.*

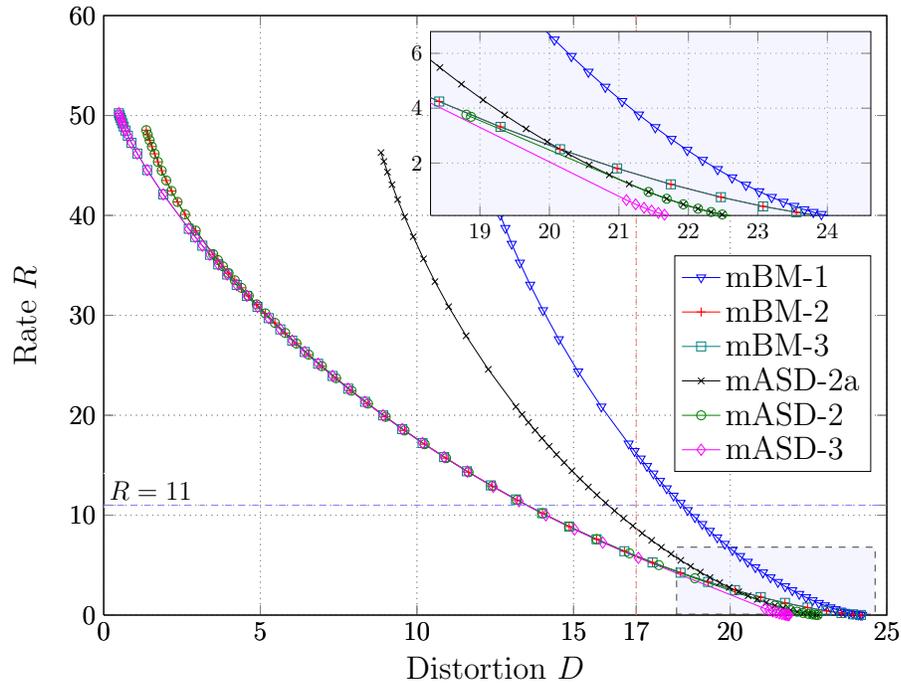


Fig. 6. A realization of RD curves at  $E_b/N_0 = 5.2$  dB for various decoding algorithms for the (255,239) RS code over an AWGN channel.

### G. Simulation Results

In this section, we present simulation results on the performance of RS codes over an AWGN channel with either BPSK or 256-QAM as the modulation format. In all the figures, the curve labeled mBM-1 corresponds to standard errors-and-erasures BM decoding with multiple erasure patterns. For  $\ell > 1$ , the curves labeled mBM- $\ell$  correspond to errors-and-erasures BM decoding with multiple decoding trials using both erasures and the top- $\ell$  symbols. The curves labeled mASD- $\mu$  correspond to multiple ASD decoding trials with maximum multiplicity  $\mu$ . The number of decoding attempts is  $2^R$  where  $R$  is denoted in parentheses in each algorithm's acronym (e.g., mBM-2(RD,11) uses the RD approach with  $R = 11$  while mBM-2(RDE,10) uses the RDE approach with  $R = 10$ ). Please note that not all the algorithms listed in this

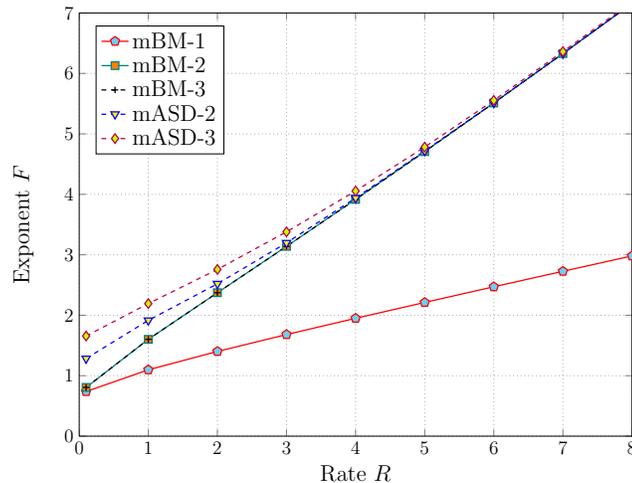


Fig. 7. A realization of RDE curves at  $E_b/N_0 = 6$  dB for various decoding algorithms for the (255,239) RS code over an AWGN channel.

section are of the same complexity unless stated explicitly.

In Fig. 6, the RD curves are shown for various algorithms using the RD approach at  $E_b/N_0 = 5.2$  dB where BPSK is used. For the (255,239) RS 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 RD curve allows one to estimate the number of decoding patterns required to achieve this FER. 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 for conventional decoding is close to 1. The RD 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 algorithm using rate  $R = 11$  has an expected distortion of less than 14. So we expect (and our simulations confirm) that the FER should be less than  $\frac{1}{2}$ . Fig. 6 also depicts the fact obtained in Example 7.

One weakness of this RD approach is that RD describes only the average distor-

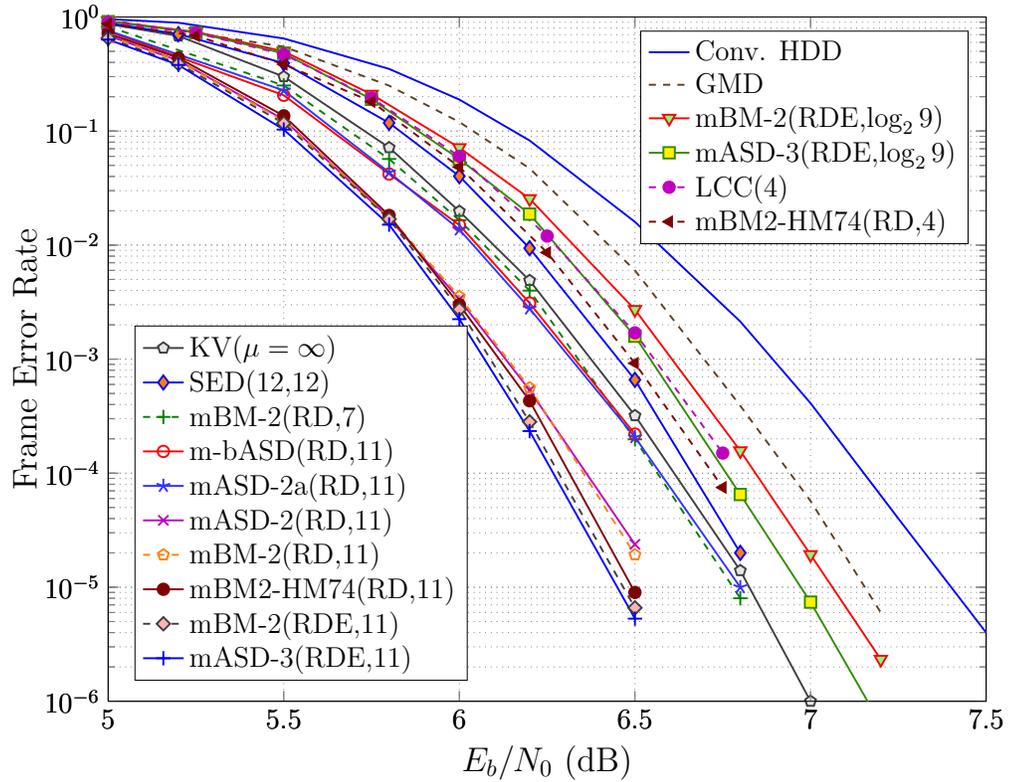


Fig. 8. Performance of various decoding algorithms for the (255,239) RS code using BPSK over an AWGN channel.

tion and does not directly consider the probability that the distortion is greater than 17. Still, we can make the following observations from the RD curve. Even at high rates (e.g.,  $R \geq 5$ ), 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 (see Example 4) and mBM-1. This implies that, for this RS code, mBM-2 using the RD approach is no worse than the more complicated ASD based approaches for a wide range of rates (i.e.,  $5 \leq R \leq 35$ ). This is also true if the RDE approach is used as can be seen in Fig. 7 which depicts the trade-off between rate  $R$  and exponent  $F$  for various algorithms at  $E_b/N_0 = 6$  dB. For this RS code, ASD based approaches have a better exponent than mBM-2 at low rates (i.e., small number of decoding trials) and have roughly the

same exponent for rates  $R \geq 5$ .

In Fig. 8, a plot of the FER versus  $E_b/N_0$  is shown for the (255,239) RS code over an AWGN channel with BPSK as the modulation format. The conventional HDD and the GMD algorithms have modest performance since they use only one or a few decoding attempts. Choosing  $R = 11$  allows us to make fair comparisons with SED(12,12). With the same number of decoding trials, mBM-2(RD,11) outperforms SED(12,12) by 0.3 dB at FER=  $10^{-4}$ . Even mBM-2(RD,7), with many fewer decoding trials, outperforms both SED(12,12) and the KV algorithm with  $\mu = \infty$ . Among all our proposed algorithms using the RD approach with rate  $R = 11$ , the mBM2-HM74(RD,11) achieves the best performance. This algorithm uses the Hamming (7,4) covering code for the 7 LRPs and the RD approach for the remaining codeword positions. Meanwhile, small differences in the performance among mBM-2(RD,11), mBM-3(RD,11), mASD-2(RD,11), and mASD-3(RD,11) suggest that: (i) taking care of the 2 most likely symbols at each codeword position is good enough for multiple decoding of this RS code and (ii) multiple runs of errors-and-erasures decoding is generally almost as good as multiple runs of ASD decoding. Recall that this result is also correctly predicted by the RD analysis. When the RDE approach is used, mBM-2(RDE,11) still has roughly the same performance as a more complex mASD-3(RDE,11). One can also observe that these two algorithms using the RDE approach achieve better performance than mBM-2(RD,11) and mBM2-HM74(RD,11) that use the RD approach. We also simulate our proposed algorithm at  $R = \log_2 9$  to compare with the GMD algorithm. While both mBM-2(RDE,  $\log_2 9$ ) and the GMD algorithm use the same number of 9 errors-and-erasures decoding attempts, mBM-2(RDE,  $\log_2 9$ ) yields roughly a 0.1 dB gain. The simulation results show that, at this low rate  $R = \log_2 9$ , mASD-3 has a larger gain over mBM-2 than at a higher rate  $R = 11$ . This phenomenon can be predicted in Fig. 7 where mASD-3 starts to achieve a larger

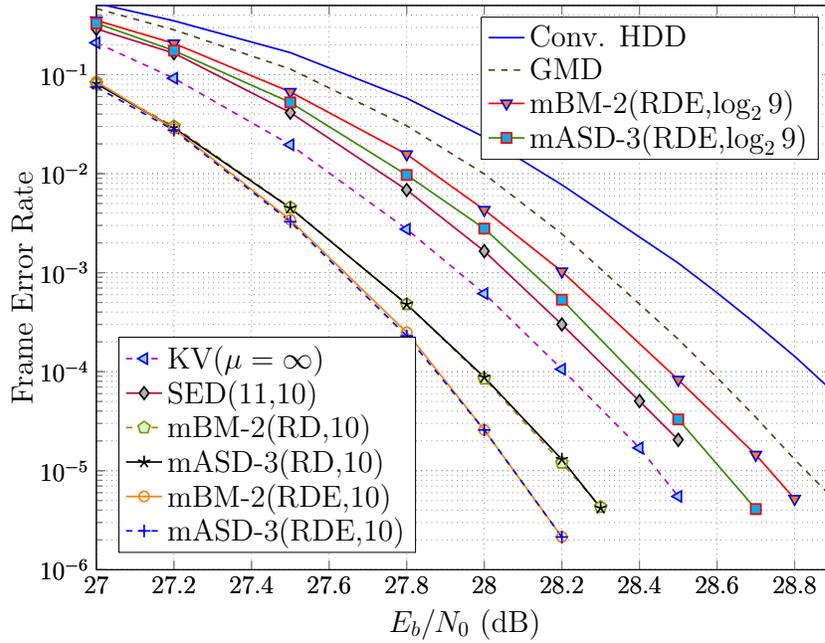


Fig. 9. Performance of various decoding algorithms for the (255,239) RS code using 256-QAM over an AWGN channel.

exponent  $F$  at small values of  $R$ .

To compare with the Chase-type approach (LCC) used in [29], in Fig. 8 we also consider the mBM2-HM74(4) algorithm that uses the Hamming (7,4) covering code for the 7 LRPs and the hard decision pattern for the remaining codeword positions. This shows that, for the (255,239) RS code, the mBM2-HM74 achieves better performance than the LCC(4) with the same number ( $2^4$ ) of decoding attempts. For the (458,410) RS code considered in Fig. 10, one can also observe that the group of algorithms that we propose have better performance than LCC(10) with the same number ( $2^{10}$ ) of decoding attempts. However, the implementation complexity of LCC(10) may be lower than the algorithms proposed here due to their clever techniques that reduce the decoding complexity per trial. It is also interesting to note that the method proposed here, based on covering codes and random codebook generation, is also

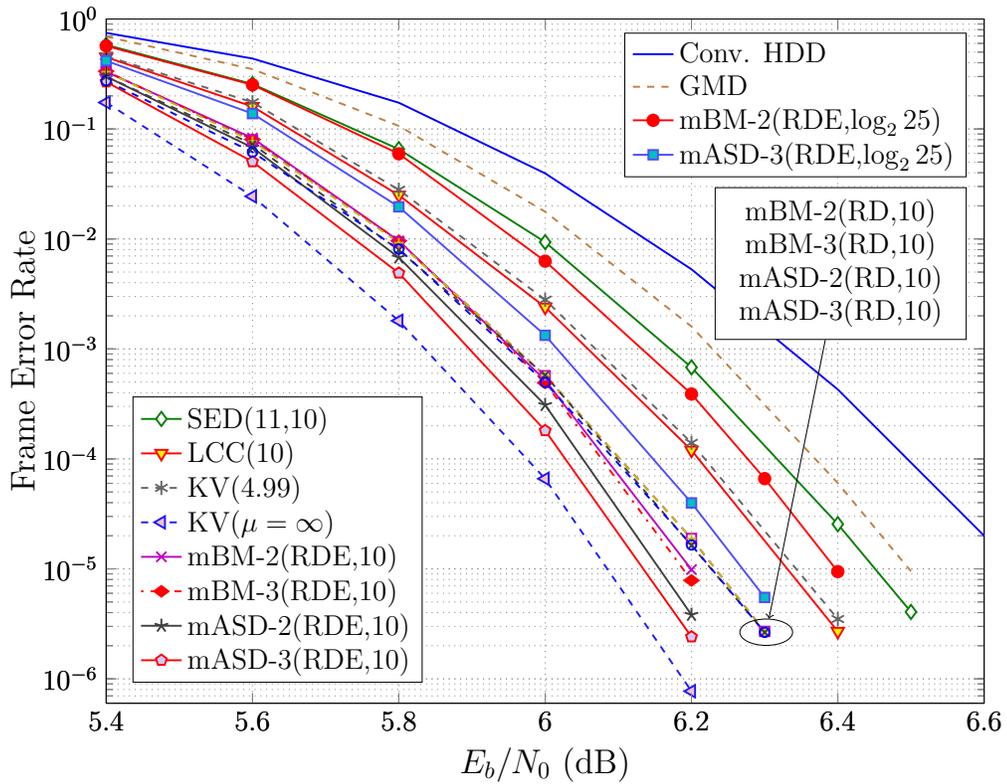


Fig. 10. Performance of various decoding algorithms for the (458,410) RS code over  $\mathbb{F}_{2^{10}}$  using BPSK over an AWGN channel.

compatible with some of the fast techniques used by the LCC decoding.

We also performed simulations using QAM and Fig. 9 shows FER versus  $E_b/N_0$  performance of the same (255,239) RS code transmitted over an AWGN channel with 256-QAM modulation. At FER= $10^{-4}$ , our proposed algorithms mBM-2(RD,10) and mBM-2(RDE,10) achieve 0.3–0.4 dB gain over SED(11,10) (with the same complexity) and also outperform KV( $\mu = \infty$ ). At  $R = 10$ , mBM-2 still achieves roughly the same performance as mASD-3.

In Fig. 10, a plot of the FER versus  $E_b/N_0$  is shown for the (458,410) RS code that has a longer block length. In this plot, BPSK is used as the modulation format and we also focus on rate  $R = 10$ . With algorithms that use the RD approach,

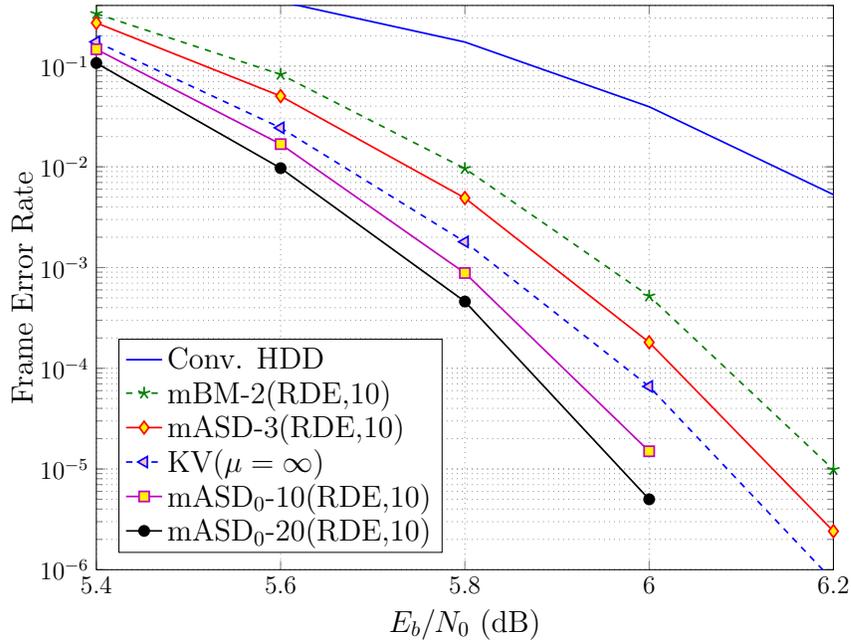


Fig. 11. Performance of various decoding algorithms for the (458,410) RS code over  $\mathbb{F}_{2^{10}}$  using BPSK over an AWGN channel.

mBM-2(RD,10) still has approximately the same performance as mBM-3(RD,10), mASD-2(RD,10), mASD-3(RD,10). However, when the RDE approach is employed, algorithms that run multiple ASD decoding attempts have a recognizable gain over algorithms that use multiple runs of BM decoding. The performance gain of the RDE approach (over the RD approach) is small, but can be seen easily by comparing mASD-3(RDE,10) to mASD-3(RD,10). As a reference, we also plot the performance of KV(4.99) which corresponds to the proportional KV algorithm [52] with the scaling factor 4.99.

In Fig. 11, the same setting is used as in Fig. 10. As can be seen in the figure, KV( $\mu = \infty$ ) achieve better performance than mASD-3(RDE,10) and mBM-2(RDE,10). However, by considering higher  $\mu$ , our algorithms using the heuristic method mASD<sub>0</sub>-10(RDE,10) and mASD<sub>0</sub>-20(RDE,10) can outperform KV( $\mu = \infty$ ).

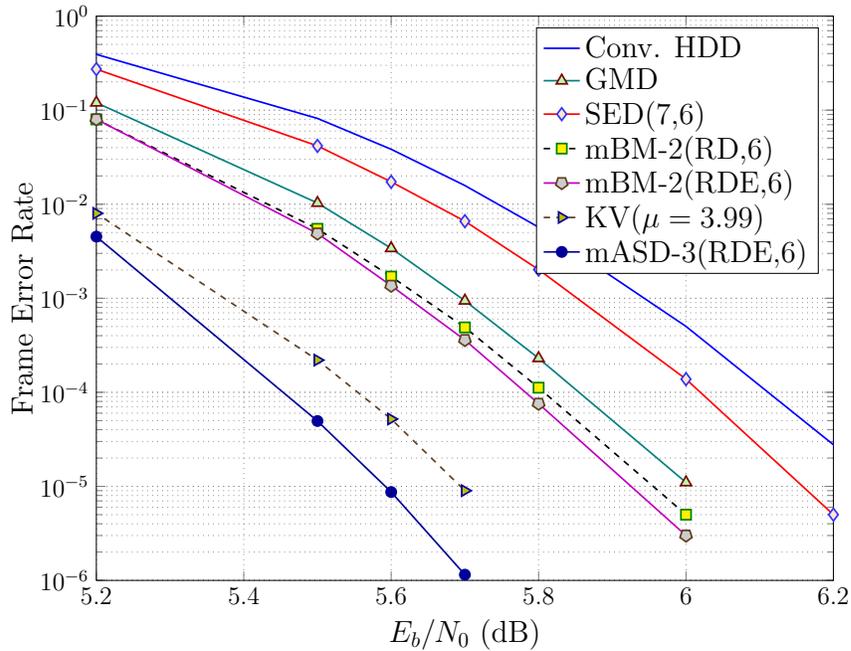


Fig. 12. Performance of various decoding algorithms for the (255,127) RS code using BPSK over an AWGN channel.

To target RS codes of lower rate, we also ran simulations of the (255,127) RS code over an AWGN channel with BPSK modulation and the results can be seen in Fig. 12. While mBM-2(RDE,6), mBM-2(RD,6), SED(7,6) and GMD all use the same number of about 64 errors-and-erasures decoding attempts, our proposed mBM-2 algorithms outperforms the other two algorithms. As seen in the plot, mASD-3(RDE,6) has quite a large gain over mBM-2(RD,6) which is reasonable since ASD decoding is known to perform very well compared to BM decoding with low-rate RS codes. In this figure, KV(3.99) denotes the proportional KV algorithm [52] with the scaling factor 3.99 and therefore with maximum multiplicity  $\mu = 3$ . While mASD-3(RDE,6) with 64 decoding attempts outperforms KV(3.99) as expected, the small gain of roughly 0.5 dB at FER= $10^{-4}$  suggests that with low-rate RS codes, one might prefer increasing  $\mu$  in a single ASD decoding attempt to running multiple ASD decoding attempts of

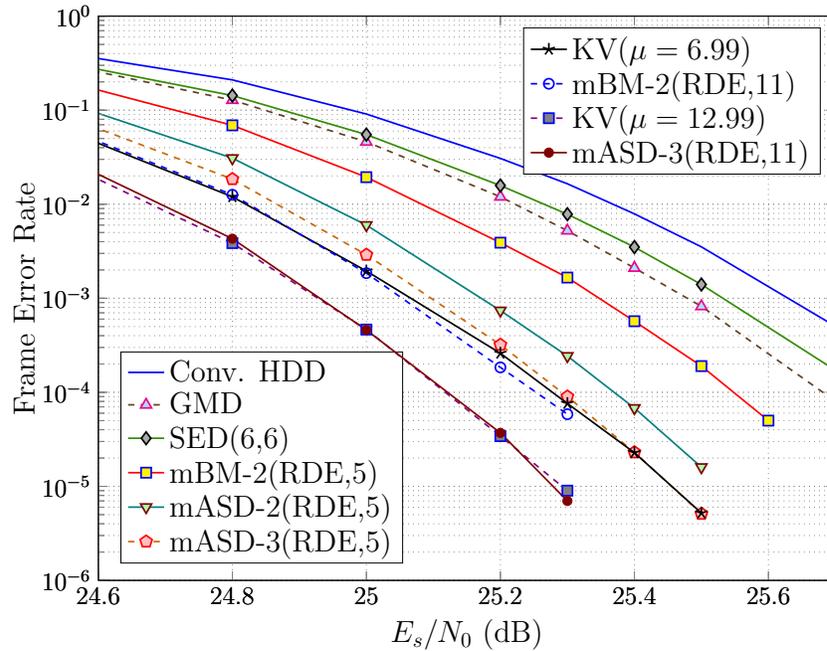


Fig. 13. Performance of various decoding algorithms for the (255,191) RS code using 256-QAM over an AWGN channel.

a lower  $\mu$ .

In Fig. 13, we show the FER versus  $E_s/N_0$  performance for the (255,191) RS codes using 256-QAM. Again, our proposed algorithm mBM-2(RDE,5) performs favorably compared to SED(6,6) and GMD with the same number of about 32 errors-and-erasures decoding attempts. Under this setup, mASD-2(RDE,5) and mASD-3(RDE,5) achieve significant gains over mBM-2(RDE,5). Our proposed mASD-3(RDE,11) and mASD-3(RDE,5) algorithms have fairly the same performance as the proportional KV algorithm with the scaling factor 12.99 and 6.99, respectively.

To compare with the iterative erasure and error decoding (IEED) algorithm proposed in [28], we also conducted simulations of the (255,223) RS code over an AWGN channel using BPSK and the results are shown in Fig. 14. With the same number of about 17 errors-and-erasures decoding attempts, our proposed mBM-2(RDE, $\log_2 17$ )

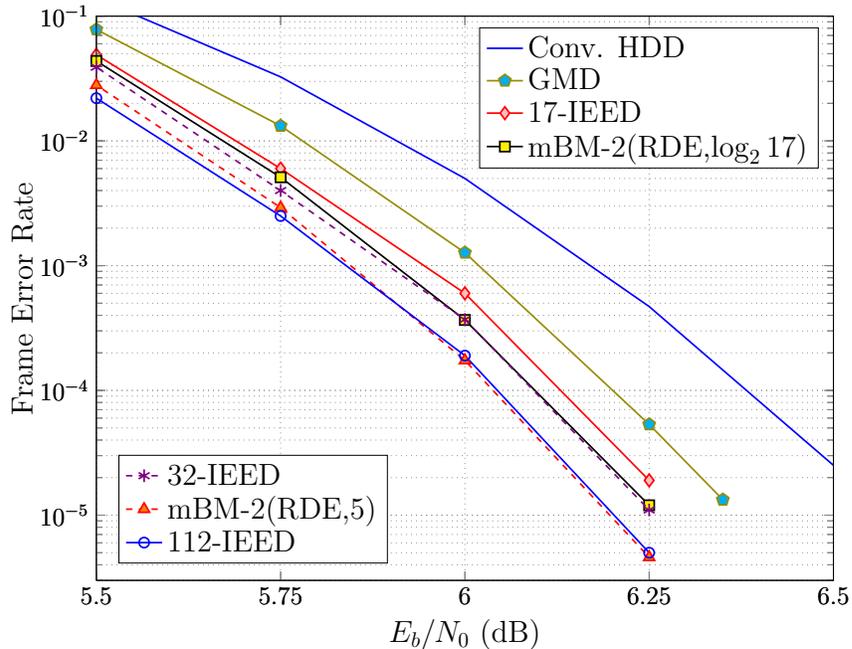


Fig. 14. Performance of various decoding algorithms for the (255,223) RS code using BPSK over an AWGN channel.

algorithm outperforms both the GMD and 17-IEED algorithms. In fact, at FER smaller than  $10^{-3}$ , mBM-2(RDE,  $\log_2 17$ ) has roughly the same performance as 32-IEED which needs to use 32 decoding attempts. Meanwhile, mBM-2(RDE, 5) that uses 32 decoding attempts performs as good as 112-IEED where 112 decoding attempts are required.

## H. Appendix

### 1. Proof of Corollary 2

*Proof.* Using the formula in [47, p. 27], we have

$$D_{\max} = \sum_{i=1}^n \min_{\hat{j}} \sum_{j=0}^{\ell} p_{i,j} \delta_{j\hat{j}}.$$

For mBM- $\ell$  with distortion matrix in (2.4), we have  $\sum_{j=0}^{\ell} p_{i,j} \delta_{j\hat{j}} = \sum_{j \neq \hat{j}} 2p_{i,j} =$

$2(1 - p_{i,\hat{j}})$  for  $\hat{j} \geq 1$  and  $\sum_{j=0}^{\ell} p_{i,j} \delta_{j0} = \sum_{j=0}^{\ell} p_{i,j} = 1$ . Therefore,

$$\begin{aligned} D_{\max}(\text{mBM-}\ell) &= \sum_{i=1}^n \min_{\hat{j}=1,\dots,\ell} \{1, 2(1 - p_{i,\hat{j}})\} \\ &= \sum_{i=1}^n \min\{1, 2(1 - p_{i,1})\} \end{aligned}$$

since  $p_{i,1} = \max_{\hat{j} \geq 1} \{p_{i,\hat{j}}\}$ .

Similarly, for mASD- $\mu$  with distortion matrix  $\Delta_{\mu}$  in (2.24), we have

$$\begin{aligned} \sum_{j=0}^{\ell} p_{i,j} \delta_{j\hat{j}} &= p_{i,0} \rho_{\hat{j},\mu} + \sum_{j=1}^{\ell} p_{i,j} \left( \rho_{\hat{j},\mu} - \frac{2m_{j,\hat{j}}}{\mu} \right) \\ &= \rho_{\hat{j},\mu} - \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{\mu} p_{i,j} \end{aligned}$$

for  $\hat{j} = 1, \dots, T$ . Since multiplicity type 1 is always defined to be  $(\mu, 0, \dots, 0)$ , we have

$\rho_{1,\mu} = 2$  and consequently,

$$\sum_{j=0}^{\ell} p_{i,j} \delta_{j1} = 2(1 - p_{i,1}).$$

Therefore, we obtain

$$D_{\max}(\text{mASD-}\mu) = \sum_{i=1}^n \min_{\hat{j}=2,\dots,T} \left\{ 2(1 - p_{i,1}), \rho_{\hat{j},\mu} - \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{\mu} p_{i,j} \right\}.$$

If mASD- $\mu$  uses multiplicity type  $(0, 0, \dots, 0)$  which is, for example, labeled as type  $T$  then we have

$$\rho_{T,\mu} - \sum_{j=1}^{\ell} \frac{m_{j,T}}{\mu} p_{i,j} = \rho_{T,\mu} = 1.$$

Consequently,

$$\begin{aligned} D_{\max}(\text{mASD-}\mu) &= \sum_{i=1}^n \min_{\hat{j}=2,\dots,T-1} \left\{ 1, 2(1 - p_{i,1}), \rho_{\hat{j},\mu} - \sum_{j=1}^{\ell} \frac{m_{j,\hat{j}}}{\mu} p_{i,j} \right\} \\ &\leq \sum_{i=1}^n \min\{1, 2(1 - p_{i,1})\} \\ &= D_{\max}(\text{mBM-}\ell) \end{aligned}$$

and this completes the proof.  $\square$

## 2. Proof of Lemma 6

*Proof.* With the notation  $\bar{p} = 1 - p$ , according to [47, p. 27] we have

$$D_{\min} = \bar{p} \min_{\hat{j}} \delta_{0\hat{j}} + p \min_{\hat{j}} \delta_{1\hat{j}} = 1 - p$$

$$D_{\max} = \min_{\hat{j}} (\bar{p} \delta_{0\hat{j}} + p \delta_{1\hat{j}}) = \min\{1, 2(1 - p)\}.$$

The function  $R(D)$  is not defined for  $D < D_{\min}$  and  $R(D) = 0$  for  $D \geq D_{\max}$ . For the case  $D_{\min} \leq D < D_{\max}$ , the rate-distortion function  $R(D)$  is given by solving the following convex optimization problem

$$\begin{aligned} \min_{\mathbf{w}} \quad & I(X; \hat{X}) \\ \text{subject to} \quad & w_{\hat{j}|j} \triangleq \Pr(\hat{X} = \hat{j} | X = j) \geq 0 \quad \forall j, \hat{j} \in \{0, 1\} \\ & w_{0|0} + w_{1|0} = 1 \\ & w_{0|1} + w_{1|1} = 1 \\ & \bar{p}w_{0|0} + pw_{0|1} + 2\bar{p}w_{1|0} = D \end{aligned}$$

where the mutual information

$$I(X; \hat{X}) = \bar{p} \sum_{\hat{j}} w_{\hat{j}|0} \log_2 \frac{w_{\hat{j}|0}}{q_{\hat{j}}} + p \sum_{\hat{j}} w_{\hat{j}|1} \log_2 \frac{w_{\hat{j}|1}}{q_{\hat{j}}}$$

and the test-channel input probability-distribution

$$q_{\hat{j}} = \Pr(\hat{X} = \hat{j}) = \bar{p}w_{\hat{j}|0} + pw_{\hat{j}|1}.$$

We then form the Lagrangian

$$J(W) = I(X; \hat{X}) + \sum_j \gamma_j (w_{0|j} + w_{1|j} - 1) + \gamma (\bar{p}w_{0|0} + pw_{0|1} + 2\bar{p}w_{1|0} - D) - \sum_{\hat{j}, j} \lambda_{\hat{j}j} w_{\hat{j}|j}$$

and the Karush-Kuhn-Tucker (KKT) conditions become<sup>9</sup>

$$\begin{cases} \frac{\partial J}{\partial w_{\hat{j}|j}} = 0 & \forall j, \hat{j} \in \{0, 1\} \\ w_{0|j} + w_{1|j} - 1 = 0 & \forall j \in \{0, 1\} \\ w_{\hat{j}|j}, \lambda_{j\hat{j}} \geq 0 & \forall j, \hat{j} \in \{0, 1\} \\ \lambda_{j\hat{j}} w_{\hat{j}|j} = 0 & \forall j, \hat{j} \in \{0, 1\} \end{cases}.$$

By [47, Lemma 1, p. 32], we only need to consider the following cases.

- Case 1:  $w_{0|0} = w_{0|1} = 0$ . In this case, we further have  $w_{1|0} = w_{1|1} = 1$ . This leads to  $R = 0$  and  $D = 2(1 - p) \geq D_{\max}$  which is a contradiction as we only consider  $D \in [D_{\min}, D_{\max})$ .

- Case 2:  $w_{1|0} = w_{1|1} = 0$ . In this case, we have  $w_{0|0} = w_{0|1} = 1$ . This leads to  $R = 0$  and  $D = 1 \geq D_{\max}$  which is also a contradiction.

- Case 3:  $w_{\hat{j}|j} > 0 \forall j, \hat{j} \in \{0, 1\}$ . In this case, we know  $\lambda_{j\hat{j}} = 0$  and then, from  $\frac{\partial J}{\partial w_{\hat{j}|j}} = 0$ , we obtain

$$\begin{aligned} \bar{p}(\log_2 \frac{w_{\hat{j}|0}}{q_{\hat{j}}} + \delta_{0\hat{j}}\gamma) + \gamma_0 &= 0 \quad \forall k \in \{0, 1\}, \\ p(\log_2 \frac{w_{\hat{j}|1}}{q_{\hat{j}}} + \delta_{1\hat{j}}\gamma) + \gamma_1 &= 0 \quad \forall k \in \{0, 1\}. \end{aligned}$$

Equivalently, we have

$$\begin{aligned} w_{\hat{j}|0} &= q_{\hat{j}} 2^{-\delta_{0\hat{j}}\gamma} 2^{\frac{-\gamma_0}{\bar{p}}} \quad \forall k \in \{0, 1\}, \\ w_{\hat{j}|1} &= q_{\hat{j}} 2^{-\delta_{1\hat{j}}\gamma} 2^{\frac{-\gamma_1}{p}} \quad \forall k \in \{0, 1\}. \end{aligned}$$

---

<sup>9</sup>Here we use some abuse of notation and still write the optimizing values in their old forms without a \* notation.

Letting  $\alpha \triangleq 2^{-\mu}$  and noticing that  $w_{0|j} + w_{1|j} = 1 \ \forall j \in \{0, 1\}$ , we get

$$\begin{aligned} w_{0|0} &= \frac{q_0}{q_0 + q_1\alpha}, & w_{0|1} &= \frac{q_0\alpha}{q_0\alpha + q_1}, \\ w_{1|0} &= \frac{q_1\alpha}{q_0 + q_1\alpha}, & w_{1|1} &= \frac{q_1}{q_0\alpha + q_1}. \end{aligned}$$

Putting this into the constraints

$$\begin{cases} \bar{p}w_{0|0} + pw_{0|1} + 2\bar{p}w_{1|0} = D \\ q_0 = \bar{p}w_{0|0} + pw_{0|1} \\ q_1 = \bar{p}w_{1|0} + pw_{1|1} \end{cases}$$

we have a set of 3 equations involving 3 variables  $\alpha, q_0, q_1$ . Solving this gives us

$$\begin{aligned} \alpha &= \frac{D + p - 1}{2 - (D + p)}, \\ q_0 &= \frac{2(1 - p) - D}{3 - 2(D + p)}, \\ q_1 &= \frac{1 - D}{3 - 2(D + p)}. \end{aligned}$$

Therefore, we can obtain the optimizing  $w_{j|j}$  and have

$$\begin{aligned} R &= H_2(p) - H_2\left(\frac{1}{1 + \alpha}\right) \\ &= H_2(p) - H_2(D + p - 1). \end{aligned}$$

Hence, in all cases  $R = [H_2(p) - H_2(D + p - 1)]^+$  and we conclude the proof.  $\square$

### 3. Proof of Theorem 3

*Proof.* The objective here is to compute the RD function for a discrete source sequence  $x_1^n$  of i.n.d. source components  $x_i$ . First, with the notations  $p_{i,j} \triangleq \Pr(X_i = j)$  and  $q_{i,j} \triangleq \Pr(\hat{X}_i = j)$  for  $j \in \{0, 1\}$  and  $i \in \{1, 2, \dots, n\}$ , Lemma 6 gives us the rate-distortion

components

$$R_i(D_i) = [H_2(p_i) - H_2(D_i + p_{i,1} - 1)]^+$$

along with the test-channel input-probability distributions

$$q_{i,0} = \frac{2(1 - p_{i,1}) - D_i}{3 - 2(p_{i,1} + D_i)} \quad \text{and} \quad q_{i,1} = \frac{1 - D_i}{3 - 2(p_{i,1} + D_i)}$$

for each index  $i$  of the codeword. The overall rate-distortion function is given by

$$\begin{aligned} R(D) &= \min_{\sum_{i=1}^n D_i = D} R_i(D_i) \\ &= \min_{\sum_{i=1}^n D_i = D} \sum_{i=1}^n [H_2(p_i) - H_2(D_i + p_{i,1} - 1)]^+ \end{aligned}$$

which is a convex optimization problem.

Using Lagrange multipliers, we form the functional

$$J(D) = \sum_{i=1}^n (H_2(p_{i,1}) - H_2(D_i + p_{i,1} - 1)) + \gamma \left( \sum_{i=1}^n D_i - D \right)$$

and compute the derivatives

$$\frac{\partial J}{\partial D_i} = \log_2 \left( \frac{D_i + p_{i,1} - 1}{2 - D_i - p_{i,1}} \right) + \gamma.$$

The Kuhn-Tucker condition (see the restated version in [3], page 86) then tells us that there is  $\gamma$  such that

$$\frac{\partial J}{\partial D_i} \begin{cases} = 0 & \text{if } R_i(D_i) > 0 \\ \leq 0 & \text{if } R_i(D_i) = 0 \end{cases}$$

which is equivalent to

$$\frac{D_i + p_{i,1} - 1}{2 - D_i - p_{i,1}} \begin{cases} = 2^{-\gamma} & \text{if } H_2(p_{i,1}) - H_2(D_i + p_{i,1} - 1) > 0 \\ \leq 2^{-\gamma} & \text{if } H_2(p_{i,1}) - H_2(D_i + p_{i,1} - 1) \leq 0 \end{cases}.$$

With the notations  $\tilde{D}_i \triangleq D_i + p_{i,1} - 1$  and  $\lambda \triangleq \frac{2^{-\gamma}}{1+2^{-\gamma}}$ , it is equivalent to

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

Finally, it becomes

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

where

$$\begin{aligned} \sum_{i=1}^n \tilde{D}_i &= \sum_{i=1}^n (D_i + p_{i,1} - 1) \\ &= D + \sum_{i=1}^n p_{i,1} - n \end{aligned}$$

and we conclude the proof.  $\square$

#### 4. Analysis of RDE Computation

Consider a binary single source  $X$  with  $\Pr(X = 1) = p$  and  $\Pr(X = 0) = 1 - p \triangleq \bar{p}$ . According to [42], 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

$$\begin{aligned} F(R, D) &= sR - stD + \max_{q_1} (-\log_2 f(q_1)) \\ &= sR - stD - \log_2 \min_{q_1} f(q_1) \end{aligned}$$

where

$$f(q_1) = \bar{p} \left( \sum_{\hat{j}} q_{\hat{j}} 2^{t\delta_{0\hat{j}}} \right)^{-s} + p \left( \sum_{\hat{j}} q_{\hat{j}} 2^{t\delta_{1\hat{j}}} \right)^{-s}$$

and  $R, D$  are given in terms of optimizing  $\underline{q}^*$ .

For the distortion measure in (2.2) and with  $q_0 = 1 - q_1$ , 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  $q_1$ . Taking the derivative  $\frac{\partial f}{\partial q_1} = 0$  gives us

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

In order to minimize  $f(q_1)$  over  $q_1 \in [0, 1]$ , we consider three following cases where the optimal  $q_1^*$  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  $q_1^* = 0$  and we can also compute

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

- Case 2:  $1 \geq p \geq \frac{1}{1+2^t(2^{s+1})}$  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  $q_1^* = 1$  and we get

$$D = \frac{2\bar{p}}{p2^{2ts} + \bar{p}}; \quad R = 0; \quad F = D_{\text{KL}}(u||p)$$

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:  $\frac{2^t}{1+2^t} < p < \frac{1}{1+2^t(2^{s+1})}$  then  $\beta \in (0, 1)$ . In this case, the optimal  $q_1^* = \beta$ . We can find  $w_{j|j}^* = \frac{q_j^* 2^{t\delta_{jj}}}{\sum_j q_j^* 2^{t\delta_{jj}}}$  according to [42] and then obtain

$$D = \frac{2^t}{1 + 2^t} + 1 - u,$$

$$R = H_2(u) - H_2(u + D - 1),$$

$$F = D_{\text{KL}}(u||p)$$

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

$$q_1^* = \frac{1-D}{3-2(u+D)} \quad \text{and} \quad q_0^* = \frac{2(1-u)-D}{3-2(u+D)}.$$

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  $(1-D, 1 - \frac{D}{2})$  and  $R$  spans  $(0, H_2(1-D))$ .

### 5. Faster Algorithm to Compute RD Function for m-bASD

Given a total distortion  $D$  in the range  $[D_{\min}, D_{\max}]$  where  $D_{\min} = \sum_{i=1}^N (1 - r_{i,1})$  and  $D_{\max} = \sum_{i=1}^N \min\{1, 3(1 - r_{i,1})\}$ , the following algorithm gives the corresponding total rate  $R$  and the test-channel input-probability distribution  $s_{i,1}$ . We assume that  $D \in (D_{\min}, D_{\max})$  because the solution is trivial if  $D$  is an endpoint. Let us denote

$$\omega(x, y) \triangleq \frac{1 + 2x + 3x^2}{1 + x + x^2} - y \frac{1 + 2x}{1 + x}.$$

The algorithm proceeds as follows.

- Step 1: Start with initial values  $\bar{D}^{(0)} = \frac{D}{N}$ ,  $\bar{r}^{(0)} = \frac{1}{N} \sum_{i=1}^N r_{i,1}$ ,  $\mathcal{I}^{(0)} = \{1, 2, \dots, N\}$  and set  $t \leftarrow 0$ .
- Step 2: Find the unique  $\lambda_{(t)} \in (0, 1)$  such that

$$\bar{D}^{(t)} = \omega(\lambda_{(t)}, \bar{r}^{(t)}). \tag{2.50}$$

- Step 3: For convenience, let  $\mathcal{I}_+^{(t)}$  and  $\mathcal{I}_-^{(t)}$  denote  $\{i \in \mathcal{I}^{(t)} : R_i(\lambda_{(t)}) > 0\}$  and

$\{i \in \mathcal{I}^{(t)} : R_i(\lambda_{(t)}) \leq 0\}$ , respectively where  $R_i(\cdot)$  is given in (). Update

$$D_i^{(t)} = \begin{cases} \omega(\lambda_{(t)}, r_{i,1}) & \text{if } i \in \mathcal{I}_+^{(t)}, \\ \min\{1, 3(1 - r_{i,1})\} & \text{if } i \in \mathcal{I}_-^{(t)}, \\ D_i^{(t-1)} & \text{if } i \notin \mathcal{I}^{(t)}. \end{cases}$$

- Step 4: If  $\mathcal{I}_-^{(t)} \neq \emptyset$ , update new values

$$\begin{aligned} \mathcal{I}^{(t+1)} &\leftarrow \mathcal{I}_+^{(t)}, \\ \bar{D}^{(t+1)} &\leftarrow \frac{1}{|\mathcal{I}^{(t+1)}|} \left( |\mathcal{I}^{(t)}| \bar{D}^{(t)} - \sum_{i \in \mathcal{I}_-^{(t)}} D_i^{(t)} \right), \end{aligned} \quad (2.51)$$

$$\bar{r}^{(t+1)} \leftarrow \frac{1}{|\mathcal{I}^{(t+1)}|} \left( \sum_{i \in \mathcal{I}_+^{(t)}} r_{i,1} \right), \quad (2.52)$$

set  $t \leftarrow t + 1$  and go back to Step 2. Otherwise, if  $\mathcal{I}_-^{(t)} = \emptyset$  then output  $\lambda \leftarrow \lambda_{(t)}$  and stop. The final rate is given by (2.43) and (2.44). The corresponding test-channel input-probability distribution is given by (2.45) and (2.46).

To analyze the above algorithm, we first have the following lemma.

**Lemma 10.** *For  $\lambda \in (0, 1)$ , one has*

$$\omega(\lambda, r_{i,1}) < \min\{1, 3(1 - r_{i,1})\} \text{ if and only if } R_i(\lambda) > 0, \quad (2.53)$$

$$\omega(\lambda, r_{i,1}) > 1 - r_{1,i}, \quad (2.54)$$

$$\min\{1, 3(1 - r_{i,1})\} \geq 1 - r_{1,i}, \quad (2.55)$$

$$\min\{1, 3(1 - r_{i,1})\} \leq \frac{1}{2}(4 - 3r_{i,1}). \quad (2.56)$$

*Proof.* From the formula for  $R_i(\lambda)$  in (2.44), the RHS is a concave function in  $r_{1,i}$  and equal to zero at  $\frac{\lambda + \lambda^2}{1 + \lambda + \lambda^2}$  and  $\frac{1 + \lambda}{1 + \lambda + \lambda^2}$ . Thus,  $R_i(\lambda) > 0$  if and only if  $\frac{\lambda + \lambda^2}{1 + \lambda + \lambda^2} < r_{i,1} < \frac{1 + \lambda}{1 + \lambda + \lambda^2}$  which can be shown to be equivalent to  $\omega(x, r_{i,1}) < \min\{1, 3(1 - r_{i,1})\}$ . Thus, (2.53)

holds.

Meanwhile, (2.54) can be seen from

$$\omega(\lambda, r_{i,1}) = 1 - r_{i,1} + \frac{\lambda}{1 + \lambda} \left( \frac{\lambda^2 + 2\lambda}{1 + \lambda + \lambda^2} + 1 - r_{i,1} \right) > 1 - r_{i,1}.$$

Furthermore, we have (2.56) because  $2 \min\{1, 3(1 - r_{i,1})\} \leq \min\{1, 3(1 - r_{i,1})\} + \max\{1, 3(1 - r_{i,1})\} = 4 - 3r_{i,1}$  while (2.55) holds trivially.  $\square$

Now, we show that the proposed algorithm produces the desired solution.

One can see from the construction of the algorithm that  $\mathcal{I}^{(t)} = \mathcal{I}^{(t+1)} \cup \mathcal{I}_-^{(t)}$  and  $\mathcal{I}^{(0)} = \left( \cup_{j=0}^t \mathcal{I}_-^{(j)} \right) \cup \mathcal{I}^{(t+1)}$  for every  $t$ . The algorithm must stop after a finite number of steps because  $\mathcal{I}^{(0)}$  has a finite number of elements. Also, one has  $D_i^{(t)} = \min\{1, 3(1 - r_{i,1})\}$  for every  $i \in \cup_{j=0}^t \mathcal{I}_-^{(j)}$ . From (2.51), it can be shown by induction that

$$\begin{aligned} |\mathcal{I}^{(t)}| \bar{D}^{(t)} &= |\mathcal{I}^{(0)}| \bar{D}^{(0)} - \sum_{i \in \cup_{j=0}^{t-1} \mathcal{I}_-^{(j)}} D_i^{(j)} \\ &= D - \sum_{i \in \cup_{j=0}^{t-1} \mathcal{I}_-^{(j)}} \min\{1, 3(1 - r_{i,1})\}. \end{aligned} \quad (2.57)$$

Suppose the algorithm stops after  $t = \tau$ , we have  $\mathcal{I}_-^{(\tau)} = \emptyset$  and therefore  $\mathcal{I}^{(\tau)} = \mathcal{I}_+^{(\tau)}$ .

This implies

$$\sum_{i \in \mathcal{I}^{(\tau)}} D_i^{(\tau)} = |\mathcal{I}^{(\tau)}| \omega(\lambda_{(\tau)}, \bar{r}^{(\tau)}) = |\mathcal{I}^{(\tau)}| \bar{D}^{(\tau)} \quad (2.58)$$

where the last equation follows from (2.50).

Therefore, one has

$$\sum_{i=1}^N D_i^{(\tau)} = \sum_{i \in \cup_{j=0}^{\tau-1} \mathcal{I}_-^{(j)}} \min\{1, 3(1 - r_{i,1})\} + \sum_{i \in \mathcal{I}^{(\tau)}} D_i^{(\tau)} = D$$

by combining (2.57) and (2.58).

Thus, at this point, this algorithm produces the solution to the procedure in

Theorem 4.

However, there is another technical detail regarding (2.50) in the algorithm that also needs to be addressed.

**Lemma 11.** *The above algorithm can proceed because (2.50) has one and only one solution of  $\lambda_{(t)}$  in  $(0, 1)$ .*

*Proof.* To show that, one can prove, for all  $t$ , that the cubic equation  $\Gamma^{(t)}(\lambda) = a_0\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0$ , which is equivalent to (2.50), has one and only one root in  $(0, 1)$  where  $a_0 = \bar{D}^{(t)} + 2\bar{r}^{(t)} - 3$ ,  $a_1 = 2\bar{D}^{(t)} + 3\bar{r}^{(t)} - 5$ ,  $a_2 = 2\bar{D}^{(t)} + 3\bar{r}^{(t)} - 3$ , and  $a_3 = \bar{D}^{(t)} + \bar{r}^{(t)} - 1$ .

We claim that

$$\Gamma^{(t)}(0) = \bar{D}^{(t)} + \bar{r}^{(t)} - 1 > 0 \text{ and } \Gamma^{(t)}(1) = 3(2\bar{D}^{(t)} + 3\bar{r}^{(t)} - 4) < 0. \quad (2.59)$$

Thus,  $\Gamma^{(t)}$  has at least one root in  $(0, 1)$ . Because of (2.59), we further have  $a_1 = \frac{\Gamma^{(t)}(1)}{3} - 1 < 0$  and  $a_0 = \frac{\Gamma^{(t)}(1)}{3} - \Gamma^{(t)}(0) < 0$ . Vieta's formulas tell us that sum of all roots (may not be real) equals  $-\frac{a_1}{a_0} < 0$  and product of the roots equals  $-\frac{a_3}{a_1} > 0$ . Hence  $\Gamma^{(t)}$  has only one positive root. Thus, we conclude that  $\Gamma^{(t)}$  must have one and only one real root in  $(0, 1)$ .

Now we will prove the claim (2.59).

We start by seeing that  $\Gamma^{(0)}(0) = \frac{D}{N} + \frac{1}{N} \sum_{i=1}^N r_{i,1} - 1 = \frac{1}{N}(D - D_{\min}) > 0$  and  $\Gamma^{(0)}(1) = \frac{3}{N} (2D - \sum_{i=1}^N (1 + 3(1 - r_{i,1}))) \leq \frac{3}{N} (2D - 2D_{\max}) < 0$ .

For  $t \geq 0$ , we have

$$|\mathcal{I}^{(t+1)}|\Gamma^{(t+1)}(0) = |\mathcal{I}^{(t)}|\bar{D}^{(t)} - \sum_{i \in \mathcal{I}_-^{(t)}} \min\{1, 3(1 - r_{i,1})\} - \sum_{i \in \mathcal{I}_+^{(t)}} (1 - r_{i,1}) \quad (2.60)$$

$$> |\mathcal{I}^{(t)}|\bar{D}^{(t)} - \sum_{i \in \mathcal{I}_-^{(t)}} \min\{1, 3(1 - r_{i,1})\} - \sum_{i \in \mathcal{I}_+^{(t)}} \omega(\lambda_{(t)}, r_{i,1}) \quad (2.61)$$

$$\geq 0 \quad (2.62)$$

where (2.60) follows from the update rule (2.52) and (2.52), (2.61) follows from (2.54), and (2.62) follows from

$$\begin{aligned} |\mathcal{I}^{(t)}|\bar{D}^{(t)} &= |\mathcal{I}^{(t)}|\omega(\lambda_{(t)}, \bar{r}^{(t)}) \\ &= \sum_{i \in \mathcal{I}^{(t)}} \omega(\lambda_{(t)}, r_{i,1}) \\ &\geq \sum_{i \in \mathcal{I}_-^{(j)}} \min\{1, 3(1 - r_{i,1})\} + \sum_{i \in \mathcal{I}_+^{(j)}} \omega(\lambda_{(t)}, r_{i,1}). \end{aligned}$$

Meanwhile, we also have

$$\begin{aligned} \frac{N}{3} |\mathcal{I}^{(t+1)}|\Gamma^{(t+1)}(1) &= 2 \left( |\mathcal{I}^{(j)}|\bar{D}^{(j)} - \sum_{i \in \mathcal{I}_-^{(t)}} \min\{1, 3(1 - p_i)\} \right) + \sum_{i \in \mathcal{I}_+^{(t)}} (3r_{i,1} - 4) \\ &< 2 \sum_{i \in \mathcal{I}_+^{(t)}} \min\{1, 3(1 - p_i)\} + \sum_{i \in \mathcal{I}_+^{(t)}} (3r_{i,1} - 4) \\ &\leq 0 \end{aligned} \quad (2.63)$$

where (2.63) follows from (2.56).  $\square$

## CHAPTER III

SPATIALLY-COUPLED CODES AND THRESHOLD SATURATION ON  
INTERSYMBOL-INTERFERENCE CHANNELS

## A. Introduction

Irregular low-density parity-check (LDPC) codes can be carefully designed to achieve the capacity of the binary erasure channel (BEC) [15] and closely approach the capacity of general binary-input symmetric-output memoryless (BMS) channels [57] under belief-propagation (BP) decoding. LDPC convolutional codes, which were introduced in [17] and shown to have excellent BP thresholds in [18, 19], have recently been observed to *universally* approach the capacity of various channels. The fundamental mechanism behind this is explained well in [20], where it is proven analytically for the BEC that the BP threshold of a particular spatially-coupled ensemble converges to the maximum a-posteriori (MAP) threshold of the underlying ensemble. A similar result was also observed independently in [58] and stated as a conjecture. Such a phenomenon is now called “threshold saturation via spatial coupling” and has also been empirically observed for general BMS channels [59]. In fact, threshold saturation seems to be quite general and has now been observed in a wide range of problems, e.g., see [60, 61, 62, 63, 64, 65]<sup>1</sup>.

In the realm of channels with memory and particularly intersymbol interference (ISI) channels, the capacity may not be achievable via equiprobable signaling. For linear codes, a popular practice is to compare instead with the symmetric information rate (SIR), which is also known as  $C_{\text{i.u.d.}}$  [66], because this rate is achievable by

---

<sup>1</sup>To be precise, the papers [60, 62, 63] only observe the threshold saturation effect indirectly because the considered EXIT-like curves provide no direct information about the MAP threshold of the underlying ensemble.

random linear codes with maximum-likelihood (ML) decoding. A numerical method for tightly estimating the SIR of general finite-state channels was first proposed in [67, 68]. For LDPC codes over ISI channels, a joint iterative BP decoder that operates on a large graph representing both the channel and the code constraints [66, 69] can perform quite well and even approach the SIR [70, 71]. Progress has also been made on the design of SIR-approaching irregular LDPC codes for some specific ISI channels [70, 71, 72, 73, 74]. However, channel parameters must be known at the transmitter for such designs and therefore universality across ISI channels appears difficult to achieve.

Now that the threshold saturation effect of spatially-coupled codes has shown benefits in a number of communication problems, it is quite natural to consider them as a potential candidate to *universally* approach the SIR of ISI channels with low decoding complexity. In fact, the combination of spatially-coupled codes and ISI channels was recently considered by Kudekar and Kasai [62] for the simple diode erasure channel (DEC) from [71, 75]. They provided a numerical evidence that the joint BP threshold of the spatially coupled codes can approach the SIR over the DEC (by increasing the degrees while keeping the rate fixed). Also, they outlined a tentative proof approach for the threshold saturation following the ideas in [20]. However, the EXIT-like curves they considered were not equipped with an area theorem and therefore could not be directly connected with the MAP threshold of the underlying ensemble. Thus, the threshold saturation effect was only indirectly observed. In a more recent work, Sekido *et al.* observed that spatially-coupled codes under joint iterative decoding can also approach the SIR over the class-II Partial Response (PR2) channel with erasure noise (PR2EC) [76].

In this chapter, we consider the transmission of the spatially-coupled codes over the family of generalized erasure channels (GECs) of which the BEC, DEC, and

PR2EC are three particular examples. For these GECs, we provide a rigorous analysis of the upper bound on the MAP threshold of LDPC codes based on the extension of [77]. Note that for the DEC, this extension was first described in an earlier paper by one of the authors [78]. We then employ a counting argument and present a numerical evidence that this bound is indeed tight for the regular LDPC ensembles and the DEC. With the MAP threshold estimated, the threshold saturation phenomenon can be numerically observed to occur exactly for several channels from the family of GECs. Next, we also consider the case of more general ISI channels where, by deriving the appropriate GEXIT curve and associated area theorem, the MAP threshold upper bound can be computed and threshold saturation can be seen. If the threshold saturation conjecture holds for these systems, then it is possible for spatially-coupled codes to universally approach the SIR of ISI channels under joint iterative BP decoding because regular LDPC codes can achieve the SIR under MAP decoding [79]. Recently, progress has been made in constructing a general proof for threshold saturation of spatially-coupled systems over various models among which are ISI channels [80]. Part of the results reported in this chapter have appeared in [81, 82].

## B. Background

In this section, we briefly describe our notation for ISI channels, LDPC ensembles, the joint iterative decoder and spatially-coupled codes.

### 1. ISI Channels and the SIR

For a finite input alphabet  $\mathcal{X}$  and an output alphabet  $\mathcal{Y}$ , let  $\{X_i\}_{i \in \mathbb{Z}}$  be the discrete-time input sequence and  $\{Y_i\}_{i \in \mathbb{Z}}$  be the discrete-time output sequence, i.e.,  $X_i \in \mathcal{X}$

and  $Y_i \in \mathcal{Y}$ . Many ISI channels of interest admit linear models of the form

$$Y_i = \sum_{t=0}^{\nu} a_t X_{i-t} + N_i, \quad (3.1)$$

where  $\mathcal{Y} = \mathbb{R}$ , the channel memory is  $\nu$ ,  $\{a_t\}_{t=0}^{\nu}$  is the set of tap coefficients and  $\{N_i\}_{i \in \mathbb{Z}}$  is a sequence of independent noise random variables. One can also write the above as  $Y_i = Z_i + N_i$  where  $Z_i = \sum_{t=0}^{\nu} a_t X_{i-t}$  is the ISI channel output without noise. In this chapter, we restrict our attention to the class of binary-input ISI channels. Often, the tap coefficients are represented through a transform domain polynomial  $a(D) = \sum_{t=0}^{\nu} a_t D^t$ . For example, when  $a(D) = 1 - D$ , the channel is known as the dicode channel.

The main subject of Section C is the family of GECs in [71, 75]. For a GEC, one can evaluate its SIR (see [71, 75] for details) as

$$I_s(\epsilon) = 1 - \int_0^1 f(t, \epsilon) dt \quad (3.2)$$

where  $f(t, \epsilon)$  is the function which maps the *a priori* erasure rate  $t$  from the code and the channel erasure rate  $\epsilon$  to the erasure rate at the output of the channel detector [71]. Strictly speaking, in this chapter we mainly consider a subclass of the GECs where the channel output sequence can be modeled as a deterministic mapping of the input sequence plus erasure noise.

The simplest example is the dicode erasure channel (DEC), which is basically a discrete-time 1st-order differentiator (i.e.,  $a(D) = 1 - D$ ), whose output is erased with probability  $\epsilon$  and transmitted perfectly with probability  $1 - \epsilon$ . Furthermore, if the input bits are differentially encoded prior to transmission, the resulting channel is called the precoded dicode erasure channel (pDEC). The simplicity of the channel models allows one to analyze the recursions used by the Bahl-Cocke-Jelinek-Raviv

(BCJR) algorithm [83] to compute

$$f_{\text{DEC}}(t, \epsilon) = \frac{4\epsilon^2}{(2 - t(1 - \epsilon))^2} \quad (3.3)$$

for the DEC and

$$f_{\text{pDEC}}(t, \epsilon) = \frac{4\epsilon^2 t(1 - \epsilon(1 - t))}{(1 - \epsilon(1 - 2t))^2} \quad (3.4)$$

for the pDEC. For both cases, explicit calculations give  $I_s(\epsilon) = 1 - \frac{2\epsilon^2}{1+\epsilon}$  [71]. For the PR2 channel with erasure noise (PR2EC) where  $a(D) = 1 + 2D + D^2$ , one has

$$f_{\text{PR2EC}}(t, \epsilon) = \frac{4\epsilon^3(4 - 4(1 - \epsilon)t + (1 - \epsilon)t^2)}{(4 - 2(1 - \epsilon^2)t - (1 - \epsilon)\epsilon^2 t^2)^2} \quad (3.5)$$

which gives  $I_s(\epsilon) = 1 - \frac{2\epsilon^3(1+\epsilon)}{2+\epsilon^2+\epsilon^3}$  [76]. Note that this formula also applies to the standard BEC, where one has  $f_{\text{BEC}}(t, \epsilon) = \epsilon$  and  $I_s(\epsilon) = 1 - \epsilon$ .

Section D considers more general ISI channels among which the most common is linear ISI channels with additive white Gaussian noise (AWGN). For this class of ISI channels, the SIR is given by

$$C_{\text{i.u.d.}} = \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, Y_1^n) \Big|_{p_{X_1^n}(x_1^n) = 2^{-n}}.$$

Unfortunately, no closed-form solutions for the SIR are known in this case. Instead, the numerical method described in [67, 68, 84] is typically used to give tight estimates of the SIR.

## 2. LDPC Ensembles and the Joint BP Decoder

When an LDPC code is transmitted over an ISI channel defined by (3.1), one can construct a large graph by joining the code graph and the channel graph together as depicted in Fig. 15. Working on this joint graph, a joint iterative decoder typically passes the information back and forth between the channel detector and the LDPC

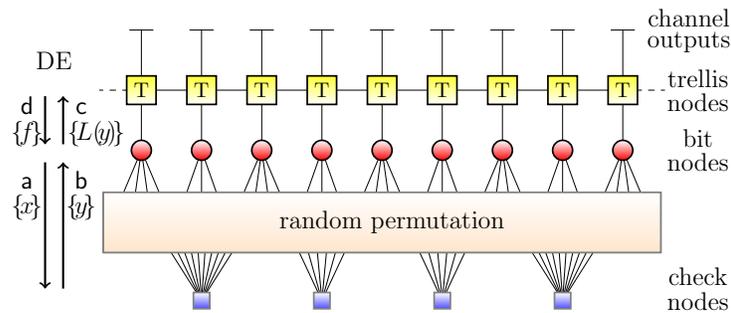


Fig. 15. Tanner graph of the joint BP decoder for ISI channels. The notations  $a$ ,  $b$ ,  $c$ ,  $d$  denote the average densities of the messages traversing along the graph used in density evolution (DE). The quantities inside the brackets are erasure rates used in DE for the GEC case. The update schedule of the joint BP decoder is also implied by the arrows in this figure.

decoder. This technique is termed as turbo equalization and was first considered by Douillard *et al.* as a new application of the turbo principle [85]. For analysis, one typically considers a windowed BCJR detector so that the computation graph becomes tree-like as  $n \rightarrow \infty$  (see [6, Ch. 6.4], [66]) and the addition of a random scrambling vector to symmetrize the effective channel [86]. The latter is very similar to using a random coset of the LDPC code to allow a general analysis of the decoder using the all-zero codeword, which was also used in [66] to derive the density evolution (DE) equation and prove a concentration theorem for the ISI channels. Throughout this chapter, a superscript  $W$  is used to imply that a windowed BCJR detector of size  $W$  is employed.

In this dissertation, we also consider the transmission of SC codes, a special class of LDPC codes, over the ISI channels. In this case, for example, a joint code/channel graph for the  $(3, 6, L)$  SC ensemble and the ISI channels is shown in Fig. 16.

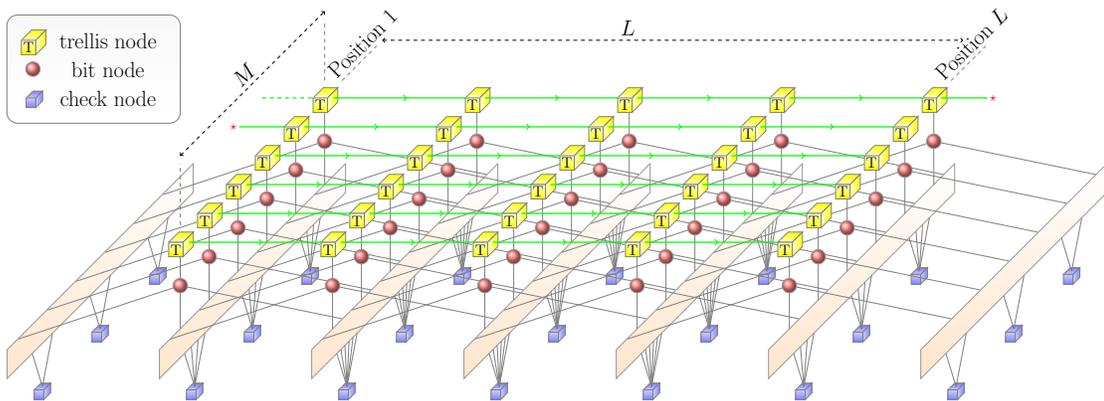


Fig. 16. The joint graph for the  $(d_l, d_r, L)$  ensemble over the ISI channels. Illustrated in this figure is the case where  $d_l = 3$  and  $d_r = 6$ . In the setup we consider, the bit transmission starts in the top left corner and proceeds row by row (e.g., see the green arrows). The (red) stars are to “connect” consecutive rows.

### C. ISI Channels with Erasure Noise: The GECs

In this section, we focus on the family of GECs. A closed-form analysis of the (E)BP-EXIT curves is presented for some systems. This analysis allows us to obtain an upper bound on the MAP threshold of the underlying ensemble. Then, DE is used to compute the BP thresholds of the corresponding SC ensembles and demonstrate the threshold saturation effect.

#### 1. BP and EBP Curves for the GEC

For the family of GECs, the DE update equation of the joint BP decoder is given by

$$x^{(\ell+1)} = f(L(1 - \rho(1 - x^{(\ell)}), \epsilon)\lambda(1 - \rho(1 - x^{(\ell)})) \quad (3.6)$$

where  $x^{(\ell)}$  is the average erasure rate emitted from bit nodes to check nodes during the  $\ell$ th iteration [71].

Since  $f(\cdot, \epsilon)$  is an increasing function for a fixed channel erasure rate  $\epsilon$ , the RHS

of (3.6) is increasing in  $x^{(\ell)}$  because both  $L(1 - \rho(1 - x^{(\ell)}))$  and  $\lambda(1 - \rho(1 - x^{(\ell)}))$  are increasing in  $x^{(\ell)}$ . Thus, by induction,  $x^{(\ell)}$  is a monotone sequence which is bounded within  $[0, 1]$  and therefore convergent. If one lets  $x$  denote the limit of  $x^{(\ell)}$  when  $\ell \rightarrow \infty$ , then the fixed point (FP) equation is given by

$$x = f(L(y(x)), \epsilon)\lambda(y(x)), \quad (3.7)$$

where, for simplicity of notation, we use  $y(x) \triangleq 1 - \rho(1 - x)$  (or the shorthand  $y$ ).

For most GECs,  $f(t, \epsilon)$  is strictly increasing in  $\epsilon$  for fixed  $t$ . In this case, there exists a unique function  $\xi(t, s)$  such that  $f(t, \xi(t, s)) = s$  and one can obtain

$$\epsilon(x) = \xi\left(L(y(x)), \frac{x}{\lambda(y(x))}\right). \quad (3.8)$$

**Example 11.** For the DEC case, one has  $f(t, \epsilon) = \frac{4\epsilon^2}{(2-t(1-\epsilon))^2}$  and this gives the FP equation  $x = \frac{4\epsilon^2\lambda(y)}{(2-L(y)(1-\epsilon))^2}$ . One can also solve for  $\xi(t, s) = (2-t)/\left(\frac{2}{\sqrt{s}} - t\right)$  and gets

$$\epsilon(x) = \frac{2 - L(y(x))}{2\sqrt{\frac{\lambda(y(x))}{x}} - L(y(x))}. \quad (3.9)$$

**Definition 9.** Consider a d.d.  $(\lambda, \rho)$  pair and the sequence of LDPC ensembles  $LDPC(n, \lambda, \rho)$ . For each  $\mathcal{C}_n$  picked uniformly at random from  $LDPC(n, \lambda, \rho)$ , let  $X_1^n$  be chosen randomly and uniformly from  $\mathcal{C}_n$  and  $Y_1^n$  be the received sequence after transmission over a GEC with erasure rate  $\epsilon$  and initial state  $S_0$ . The EXIT function associated with  $\mathcal{C}_n$  is defined by

$$h(\mathcal{C}_n, \epsilon) = \frac{dH(X_1^n|Y_1^n(\epsilon), S_0)}{n d\epsilon}$$

and the (asymptotic) EXIT function is given by

$$h(\epsilon) \triangleq \limsup_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} [h(\mathcal{C}_n, \epsilon)].$$

**Theorem 8.** *The above definition of the EXIT function naturally obeys the area theorem*

$$\frac{1}{n}H(X_1^n|Y_1^n(\epsilon^*), S_0) = \int_0^{\epsilon^*} h(\mathcal{C}_n, \epsilon) d\epsilon.$$

*In particular,  $\epsilon^* = 1$  implies  $\int_0^1 h(\mathcal{C}_n, \epsilon) d\epsilon = \frac{1}{n}H(X_1^n|S_0)$ , which equals the code rate  $r$  if there is a uniform prior on the set of codewords.*

When the BP estimator is used at each bit instead of the optimal MAP estimator, one also has the BP-EXIT function which is given by the following definition.

**Definition 10.** *Consider the same setting as in Definition 9, the (asymptotic) BP-EXIT function is defined to be*

$$h^{\text{BP}}(\epsilon) \triangleq \lim_{W \rightarrow \infty} \lim_{\ell \rightarrow \infty} h^{\text{BP}, \ell, W}(\epsilon) \quad (3.10)$$

where

$$h^{\text{BP}, \ell, W}(\epsilon) = \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{1}{n} \sum_{i=1}^n h_i^{\text{BP}, \ell, W}(\epsilon) \right] \quad (3.11)$$

and

$$h_i^{\text{BP}, \ell, W}(\epsilon) = \left. \frac{\partial}{\partial \epsilon_i} H(X_1^n | Y_i(\epsilon_i), \mathcal{E}_i^{\text{BP}, \ell, W}(Y_{\sim i}), S_0) \right|_{\epsilon_i = \epsilon}$$

and  $\mathcal{E}_i^{\text{BP}, \ell, W}(Y_{\sim i})$  is the extrinsic BP estimate of the  $i$ -th bit after iteration  $\ell$  that operates on the computation graph of depth  $\ell$  associated with the windowed BCJR channel detector of a fixed window size  $W$  (for the computation graph and the windowed BCJR, please refer to [6, Ch. 6.4] and [66]). Here, we imagine that  $\epsilon_i$  is the erasure rate of the channel from  $Z_i$  to  $Y_i$  which is characterized by a common parameter  $\epsilon$  where  $\epsilon_i = \epsilon$  for all  $i$ .

**Remark 15.** *By considering the windowed BCJR detector with a fixed window size  $W$ , the associated depth- $\ell$  computation graph becomes tree-like, for any fixed  $\ell$  as  $n \rightarrow \infty$ , and one can employ the concentration theorem for joint iterative decoding.*

This implies the limit in (3.11) exists. Also, the existence of limit in (3.10) is implied by the fact that  $h^{\text{BP},\ell,W}(\epsilon)$  is non-increasing in  $\ell$  and in  $W$ .

**Lemma 12.** *The EXIT function and BP-EXIT function (after iteration  $\ell$ ) can be written as*

$$h(\epsilon) = \limsup_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{1}{n} \sum_{i=1}^n H(Z_i | Y_{\sim i}(\epsilon), S_0) \right], \quad (3.12)$$

$$h^{\text{BP},\ell,W}(\epsilon) = \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{1}{n} \sum_{i=1}^n H(Z_i | \mathcal{E}_i^{\text{BP},\ell,W}(Y_{\sim i}(\epsilon)), S_0) \right], \quad (3.13)$$

where  $Z_i$  is the  $i$ -th output without noise. From this, one can see that  $h(\epsilon) \leq h^{\text{BP}}(\epsilon)$ .

*Proof.* Let  $\epsilon_i$  be the erasure rate of the channel from  $Z_i$  to  $Y_i$  where  $\epsilon_i = \epsilon$  for all  $i$ .

For the case of the optimal MAP estimator  $\mathcal{E}^{\text{MAP}}$ , one has

$$\begin{aligned} \frac{d}{d\epsilon} H(X_1^n | Y_1^n(\epsilon), S_0) &= \sum_{i=1}^n \frac{\partial}{\partial \epsilon_i} H(X_1^n | Y_1^n(\epsilon), S_0) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \epsilon_i} H(X_1^n | Y_i(\epsilon_i), \mathcal{E}_i^{\text{MAP}}(Y_{\sim i}), S_0). \end{aligned}$$

Furthermore, for any extrinsic estimator  $\mathcal{E}$ , the following holds

$$\begin{aligned} H(X_1^n | Y_i(\epsilon_i), \mathcal{E}_i(Y_{\sim i}), S_0) &= H(Z_1^n | Y_i(\epsilon_i), \mathcal{E}_i(Y_{\sim i}), S_0) \\ &= H(Z_i | Y_i(\epsilon_i), \mathcal{E}_i(Y_{\sim i}), S_0) + H(Z_{\sim i} | Y_i(\epsilon_i), \mathcal{E}_i(Y_{\sim i}), Z_i, S_0) \\ &= \epsilon_i H(Z_i | \mathcal{E}_i(Y_{\sim i}), S_0) + H(Z_{\sim i} | \mathcal{E}_i(Y_{\sim i}), Z_i, S_0) \end{aligned} \quad (3.14)$$

and this gives

$$\begin{aligned} \sum_{i=1}^n \frac{\partial}{\partial \epsilon_i} H(X_1^n | Y_i(\epsilon_i), \mathcal{E}_i(Y_{\sim i}), S_0) &= \sum_{i=1}^n \frac{\partial}{\partial \epsilon_i} H(Z_1^n | Y_i(\epsilon_i), \mathcal{E}_i(Y_{\sim i}), S_0) \\ &= \sum_{i=1}^n H(Z_i | \mathcal{E}_i(Y_{\sim i}), S_0) \end{aligned} \quad (3.15)$$

where (3.15) follows from the fact that second term on the RHS of (3.14) does not depend on  $\epsilon_i$ .

Thus, by considering two specific cases of  $\mathcal{E}$ , i.e.,  $\mathcal{E}^{\text{BP},\ell,W}$  and the optimal  $\mathcal{E}^{\text{MAP}}$ , one obtains (3.13) and (3.12).

Furthermore, the data processing inequality [41] implies that has

$$H(Z_i|Y_{\sim i}(\epsilon), S_0) \leq H(Z_i|\mathcal{E}_i^{\text{BP},\ell,W}(Y_{\sim i}(\epsilon)), S_0),$$

which can be combined with (3.13) and (3.12) to imply that  $h(\epsilon) \leq h^{\text{BP},\ell,W}(\epsilon)$  and hence  $h(\epsilon) \leq h^{\text{BP}}(\epsilon)$ .  $\square$

**Remark 16.** *To simplify the notation, from now on we will largely drop  $S_0$  in related expressions even though the dependency on  $S_0$  is always assumed throughout the chapter.*

While computing the (MAP) EXIT function in general is hard, it is relatively easy to compute the BP-EXIT function.

**Lemma 13.** *The BP-EXIT function for the GEC is given by*

$$h^{\text{BP}}(\epsilon) = \frac{d}{d\tilde{\epsilon}} \int_0^{L(y)} f(t, \tilde{\epsilon}) dt \Big|_{\tilde{\epsilon}=\epsilon}. \quad (3.16)$$

where  $L(y)$  is the extrinsic erasure rate given by the FP equation at channel erasure rate  $\epsilon$ .

*Proof.* Let  $Y_1^n(\tilde{\epsilon})$  be the result of passing  $X_1^n$  through the communication channel, e.g., the GEC, with erasure rate  $\tilde{\epsilon}$ . More precisely,  $Y_i$  is an erasure with probability  $\tilde{\epsilon}_i$  where  $\tilde{\epsilon}_i = \tilde{\epsilon}$  for all index  $i$ . Also, with some abuse of notation, let  $\mathcal{E}_1^n(t)$  be the result of passing  $X_1^n$  through a BEC extrinsic channel with erasure probability  $t$ . Similarly to [71], let  $T_n(1-t, \tilde{\epsilon}) \triangleq \frac{1}{n} \sum_{i=1}^n I(X_i; Y_1^n(\tilde{\epsilon}), \mathcal{E}_{\sim i}(t))$  denote the mutual information transfer function where  $\mathcal{E}_{\sim i}$  comprises the sequence of extrinsic bit estimates except for the  $i$ -th bit. We also let  $f_n(t, \tilde{\epsilon}) \triangleq 1 - T_n(1-t, \tilde{\epsilon})$ . By the area theorem [87, Th.

2], [77], one obtains

$$\int_0^\delta \frac{1}{n} \sum_{i=1}^n H(X_i|Y_1^n(\tilde{\epsilon}), \mathcal{E}_{\sim i}(t)) dt = \frac{1}{n} H(X_1^n|Y_1^n(\tilde{\epsilon}), \mathcal{E}_1^n(\delta)) \quad (3.17)$$

for some extrinsic erasure rate  $\delta$ .

We then have

$$\frac{d}{d\tilde{\epsilon}} \int_0^\delta f_n(t, \tilde{\epsilon}) dt = \frac{d}{d\tilde{\epsilon}} \int_0^\delta -T_n(1-t, \tilde{\epsilon}) dt \quad (3.18)$$

$$\begin{aligned} &= \frac{d}{d\tilde{\epsilon}} \int_0^\delta \left( -\frac{1}{n} \sum_{i=1}^n I(X_i; Y_1^n(\tilde{\epsilon}), \mathcal{E}_{\sim i}(t)) \right) dt \\ &= \frac{d}{d\tilde{\epsilon}} \int_0^\delta \left( \frac{1}{n} \sum_{i=1}^n H(X_i) - \frac{1}{n} \sum_{i=1}^n I(X_i; Y_1^n(\tilde{\epsilon}), \mathcal{E}_{\sim i}(t)) \right) dt \end{aligned} \quad (3.19)$$

$$\begin{aligned} &= \frac{d}{d\tilde{\epsilon}} \int_0^\delta \frac{1}{n} \sum_{i=1}^n H(X_i|Y_1^n(\tilde{\epsilon}), \mathcal{E}_{\sim i}(t)) dt \\ &= \frac{d}{d\tilde{\epsilon}} \left[ \frac{1}{n} H(X_1^n|Y_1^n(\tilde{\epsilon}), \mathcal{E}_1^n(\delta)) \right] \end{aligned} \quad (3.20)$$

$$= \frac{1}{n} \sum_{i=1}^n \frac{d}{d\tilde{\epsilon}_i} H(X_1^n|Y_i(\tilde{\epsilon}_i), \mathcal{E}_i(\delta)) \Big|_{\tilde{\epsilon}_i=\tilde{\epsilon}}$$

where (3.19) holds because  $\frac{\delta}{n} \sum_{i=1}^n H(X_i)$  is not a function of  $\tilde{\epsilon}$  while (3.20) follows from (3.17). The derivative in (3.18) exists a.e. because  $f_n(t, \tilde{\epsilon})$  is non-decreasing in  $\tilde{\epsilon}$ .

If one considers the BP estimator then

$$\frac{d}{d\tilde{\epsilon}} \int_0^{\delta_{\ell, W, n}} f_n(t, \tilde{\epsilon}) dt = \frac{1}{n} \sum_{i=1}^n h_i^{\text{BP}, \ell, W}(\tilde{\epsilon})$$

where  $\delta_{\ell, W, n}$  is erasure rate of the extrinsic channel obtained after iteration  $\ell$  of the joint BP decoder that employs the windowed BCJR detector of size  $W$ .

Taking expectation and letting  $n \rightarrow \infty$ , one has

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{d}{d\tilde{\epsilon}} \int_0^{\delta_{\ell, W, n}} f_n(t, \tilde{\epsilon}) dt \right] &= \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{1}{n} \sum_{i=1}^n h_i^{\text{BP}, \ell, W}(\tilde{\epsilon}) \right] \\ &= h^{\text{BP}, \ell, W}(\tilde{\epsilon}) \end{aligned} \quad (3.21)$$

Furthermore, it can be shown that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{d}{d\tilde{\epsilon}} \int_0^{\delta_{\ell, W, n}} f_n(t, \tilde{\epsilon}) dt \right] = \frac{d}{d\tilde{\epsilon}} \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \int_0^{\delta_{\ell, W, n}} f_n(t, \tilde{\epsilon}) dt \right] \quad (3.22)$$

$$= \frac{d}{d\tilde{\epsilon}} \int_0^{\delta_{\ell, W}} f(t, \tilde{\epsilon}) dt \quad (3.23)$$

where  $\delta_{\ell, W}$  is the extrinsic erasure rate obtained from the DE equation after iteration  $\ell$ . Here, (3.22) holds because  $\tilde{\epsilon}$  does not depend on  $\mathcal{C}_n$  and the expectation is a finite sum while (3.23) follows from the following facts: 1) for fixed  $W, \ell$  and by letting  $n \rightarrow \infty$ , one can invoke the standard analysis of concentration around ensemble average and analyze the computation tree implied by the DE equation, 2)  $f_n(t, \tilde{\epsilon})$  converges point-wise to  $f(t, \tilde{\epsilon})$  by using the superadditivity of the sequence  $\sum_{i=1}^n I(X_i; Y_1^n(\tilde{\epsilon}), \mathcal{E}_{\sim i}(t))$ , and an application of Lebesgue's dominated convergence theorem. Also, derivatives exist a.e. because  $f_n(t, \tilde{\epsilon})$  and  $f(t, \tilde{\epsilon})$  are non-decreasing in  $\tilde{\epsilon}$ .

Next, by letting  $\ell \rightarrow \infty$  and  $W \rightarrow \infty$ , one reaches a FP where  $\delta_{\ell, W} \rightarrow L(y)$  and finally obtains from (3.21) and (3.23) that

$$\frac{d}{d\tilde{\epsilon}} \int_0^{L(y)} f(t, \tilde{\epsilon}) dt \Big|_{\tilde{\epsilon}=\epsilon} = h^{\text{BP}}(\epsilon).$$

□

**Example 12.** Applying (3.16) to (3.3), (3.4), and (3.5) gives the following BP-EXIT functions

$$h_{\text{DEC}}^{\text{BP}}(\epsilon) = \frac{2\epsilon L(y)(4 - L(y)(2 - \epsilon))}{(2 - L(y)(1 - \epsilon))^2}, \quad (3.24)$$

$$h_{\text{pDEC}}^{\text{BP}}(\epsilon) = \frac{2\epsilon L^2(y)(2 - \epsilon(1 - 2L(y)))}{(1 - \epsilon(1 - 2L(y)))^2}, \quad (3.25)$$

and

$$h_{\text{PR2EC}}^{\text{BP}}(\epsilon) = \frac{2\epsilon^2 L(y) (24 - 4(6 - 4\epsilon - \epsilon^2)L(y) + 2(3 - 4\epsilon + 2\epsilon^2 + 2\epsilon^3)L^2(y))}{((4 - 2(1 - \epsilon^2)L(y) - (1 - \epsilon)\epsilon^2 L^2(y))^2)} \quad (3.26)$$

for the DEC, pDEC, and PR2EC, respectively. Here,  $y$  is a short notation of  $y(x)$  where  $x$  is the DE FP at channel erasure rate  $\epsilon$ . The formula (3.24) for the DEC case is equivalent to the result shown in [78] by analyzing the BCJR algorithm. Following a similar approach, we also computed the BP-EXIT functions for the pDEC and PR2EC and verified that the results are consistent with (3.25) and (3.26), respectively. One can also apply (3.16) for the BEC to obtain the known result  $h_{\text{BEC}}^{\text{BP}}(\epsilon) = \frac{\partial}{\partial \tilde{\epsilon}} \int_0^{L(y)} \tilde{\epsilon} dt \Big|_{\tilde{\epsilon}=\epsilon} = L(y)$ .

Using an approach similar to [77, Sec. III-B] and taking care of (3.8) and (3.16), one gets the following parametric form for the BP-EXIT function. This involves in defining

$$\mathcal{I} \triangleq \bigcup_{i=1, \dots, J} [\underline{x}^i, \bar{x}^i) \cup \{1\}$$

as the unique finite union of disjoint intervals that represent all stable and achievable FPs of DE equations. Please note that  $J$  represents the number of discontinuities in the BP-EXIT function and  $\epsilon(x)$  is monotonically increasing as  $x$  is increasing in  $\mathcal{I}$  (see [77, Sec. III-B]). The joint BP decoding threshold, denoted as  $\epsilon^{\text{BP}}$ , is the supremum of all channel parameters  $\epsilon$  such that  $h^{\text{BP}}(\epsilon) = 0$ .

**Lemma 14.** *Given a d.d. pair  $(\lambda, \rho)$ , the BP-EXIT function for a GEC is given parametrically by*

$$(\epsilon, h^{\text{BP}}(\epsilon)) = \begin{cases} (\epsilon, 0), & \epsilon \in [0, \epsilon^{\text{BP}}) \\ \left( \epsilon(x), \frac{\text{d}}{\text{d}\tilde{\epsilon}} \int_0^{L(y(x))} f(t, \tilde{\epsilon}) dt \Big|_{\tilde{\epsilon}=\epsilon(x)} \right) \forall x \in \mathcal{I}, & \epsilon \in (\epsilon^{\text{BP}}, 1] \end{cases}$$

where  $\epsilon(x)$  is given in (3.8).

In [77], the extended BP (EBP) EXIT curve for the BEC was introduced as the hidden bridge between the BP threshold and its MAP counterpart. In a similar manner, the EBP-EXIT curve for GECs is given below with its own area theorem.

**Definition 11.** For a given d.d. pair  $(\lambda, \rho)$ , the EBP-EXIT curve for the GEC is defined by the pair

$$\left( \epsilon(x), \frac{d}{d\tilde{\epsilon}} \int_0^{L(y(x))} f(t, \tilde{\epsilon}) dt \Big|_{\tilde{\epsilon}=\epsilon(x)} \right), x \in [0, 1]$$

where  $\epsilon(x)$  is given in (3.8).

**Example 13.** For the DEC case, using (3.9) and (3.24), the EBP-EXIT curve is given by

$$\left( \frac{2 - L(y(x))}{2\sqrt{\frac{\lambda(y(x))}{x}} - L(y(x))}, L(y(x)) \left( 2\sqrt{\frac{x}{\lambda(y(x))}} - \frac{xL(y(x))}{2\lambda(y(x))} \right) \right), x \in [0, 1].$$

For example, the EBP-EXIT curves for various d.d. pairs  $(\lambda, \rho)$  together with their BP thresholds can be seen in Fig. 17.

**Lemma 15.** Consider the GEC and a d.d. pair  $(\lambda, \rho)$ . Define the “trial entropy” as

$$P(x) \triangleq \int_0^x h^{\text{EBP}}(t) \epsilon'(t) dt \quad (3.27)$$

where  $h^{\text{EBP}}(x)$  is the second coordinate the EBP-EXIT curve. Then, we have

$$P(x) = \int_0^{L(y)} f(t, \epsilon(x)) dt - \frac{L'(1)}{R'(1)} (1 - R(1-x) - xR'(1-x)). \quad (3.28)$$

*Proof.* First, we let

$$\begin{aligned} Q(x) &\triangleq \int_0^{L(y)} f(t, \epsilon(x)) dt - \frac{L'(1)}{R'(1)} (1 - R(1-x) - xR'(1-x)) \\ &= \int_0^{L(y)} f(t, \epsilon(x)) dt - L'(1) \int_0^x u dy(u) \end{aligned} \quad (3.29)$$

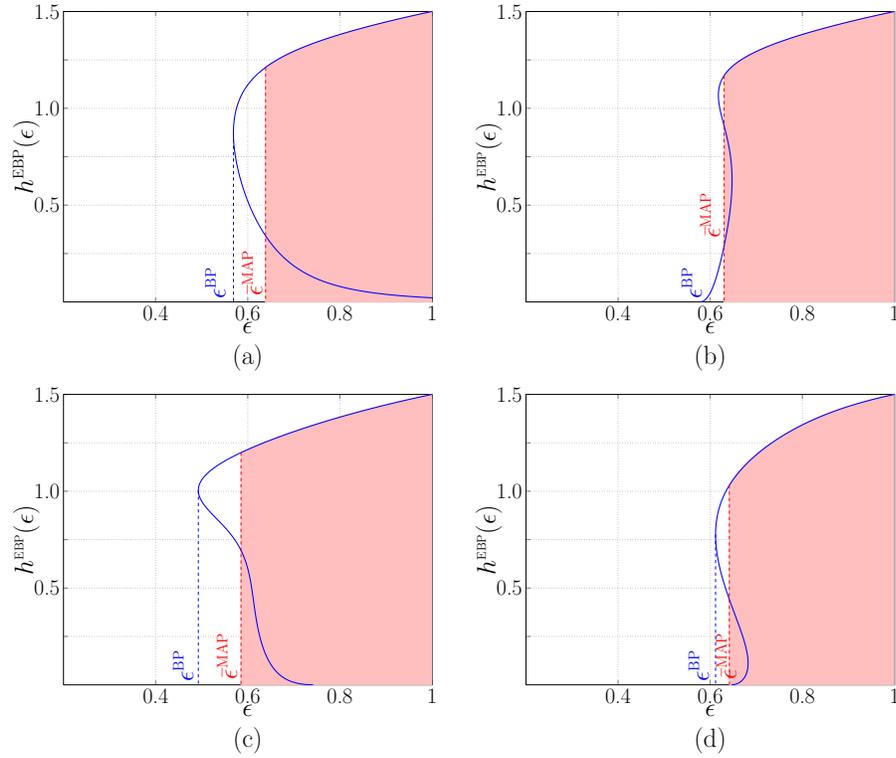


Fig. 17. EBP-EXIT curves and BP thresholds for various LDPC ensembles over the DEC: (a)  $(\lambda, \rho) = (x^2, x^5)$ , (b)  $(\lambda, \rho) = (0.6x + 0.4x^9, x^5)$ , (c)  $(\lambda, \rho) = (0.2x + 0.3x^2 + 0.5x^{15}, x^9)$ , (d)  $(\lambda, \rho) = (0.4x + 0.6x^4, 0.2x^2 + 0.8x^7)$ . Also, by setting the area of the shaded region equal to the design rate of the corresponding ensemble, one obtains the upper bound  $\bar{\epsilon}^{\text{MAP}}$  on the MAP threshold using the technique in Section 2

where in (3.29), integration by parts is used.

Then, one can use Leibniz's rule to get

$$\begin{aligned}
Q'(x) &= f(L(y), \epsilon(x))y'L'(y) + \int_0^{L(y)} \frac{\partial}{\partial x} f(t, \epsilon(x))dt - L'(1)xy' \\
&= \int_0^{L(y)} \frac{\partial}{\partial x} f(t, \epsilon(x))dt \\
&= \int_0^{L(y)} \frac{\partial}{\partial \epsilon(x)} f(t, \epsilon(x)) \frac{d}{dx} \epsilon(x) dt \\
&= \epsilon'(x) \int_0^{L(y)} \frac{\partial}{\partial \epsilon(x)} f(t, \epsilon(x)) dt \\
&= \epsilon'(x) h^{\text{EBP}}(x) \\
&= P'(x)
\end{aligned} \tag{3.30}$$

where (3.30) follows from the DE equation  $f(L(y), \epsilon(x))\lambda(y) = x$  and the fact that  $\lambda(y) = \frac{L'(y)}{L'(1)}$  while (3.31) follows by taking derivative on both sides of (3.27).

Thus,  $Q(x)$  and  $P(x)$  may differ only by a constant. But,  $P(0) = Q(0) = 0$  implies that one must have  $P(x) = Q(x)$ .  $\square$

**Example 14.** For the DEC, explicit calculation gives

$$P(x) = \frac{2\epsilon^2(x)L(y)}{2 - L(y)(1 - \epsilon(x))} - \frac{L'(1)}{R'(1)}(1 - R(1 - x) - xR'(1 - x)).$$

Also, one can see that, for the BEC, this gives same formula as [6, p. 124].

**Theorem 9.** (Area Theorem for EBP) Consider a d.d. pair  $(\lambda, \rho)$  of design rate  $r$ . Then the EBP-EXIT curve for the GEC satisfies

$$\int_0^1 h^{\text{EBP}}(x) d\epsilon(x) = r.$$

*Proof.* Using the result in Lemma 15, a direct calculation reveals that

$$\int_0^1 h^{\text{EBP}}(x) d\epsilon(x) = P(1) = \int_0^1 f(t, 1) dt - \frac{L'(1)}{R'(1)} = 1 - \frac{L'(1)}{R'(1)} = r$$

since  $\int_0^1 f(t, 1) dt = 1 - I_s(1) = 1$  and we conclude the proof.  $\square$

## 2. Upper Bound on the MAP Threshold

The MAP threshold, called  $\epsilon^{\text{MAP}}$  and defined as the supremum of all channel parameters  $\epsilon$  such that  $h(\epsilon) = 0$ , is generally hard to compute. However, because of the optimality of the MAP decoder in the sense that  $h^{\text{MAP}} \leq h^{\text{BP}}$  (see Lemma 12), one can obtain an upper bound on the MAP threshold by first finding the largest value  $\bar{\epsilon}^{\text{MAP}}$  such that  $\int_{\bar{\epsilon}^{\text{MAP}}}^1 h^{\text{BP}}(\epsilon) d\epsilon = r$  and then bound the MAP threshold by the inequality  $\epsilon^{\text{MAP}} \leq \bar{\epsilon}^{\text{MAP}}$ . This technique was introduced by Méasson *et al.* in [77] for the BEC and conjectured to be tight in many scenarios. In fact, for the whole class of regular LDPC ensembles over the BEC, this bound was analytically proven to be tight [88].

Using the ingredients provided by our analysis above, this technique can also be extended to GECs. More specifically, using Lemma 15 and Theorem 9, one has the following corollary after a few steps.

**Corollary 4.** *Assume that  $x^{\text{MAP}}$  is the solution of  $P(x) = 0$  in  $(0, 1]$  such that there is no  $x' \in (x^{\text{MAP}}, 1]$  satisfying  $\epsilon(x') = \epsilon(x^{\text{MAP}})$ . Then, one obtains the upper bound  $\epsilon^{\text{MAP}} \leq \epsilon(x^{\text{MAP}}) = \bar{\epsilon}^{\text{MAP}}$ .*

It is also clear that,  $\bar{\epsilon}^{\text{MAP}}$  for the case of regular LDPC ensembles quickly approaches  $\epsilon^{\text{SIR}}$  of GECs which is formalized by the following theorem.

**Theorem 10.** *Consider the  $(d_l, d_r)$ -regular ensemble and a GEC. Consider a fixed design rate  $r = 1 - \frac{d_l}{d_r}$ . Then*

$$\lim_{d_l, d_r \rightarrow \infty, r \text{ fixed}} \bar{\epsilon}^{\text{MAP}}(d_l, d_r) = \epsilon^{\text{SIR}}(r),$$

where  $\epsilon^{\text{SIR}}(r) = I_s^{-1}(r)$  is the erasure rate when the SIR defined in (3.2) equals  $r$ .

*Proof.* As defined above,  $x^{\text{MAP}}(d_l, d_r)$  must be a root of  $P(x) = 0$ . For a fixed rate  $r$ ,  $x^{\text{MAP}}(d_l, d_r)$  is bounded away from zero for  $d_l$  large enough (one can show that  $x^{\text{MAP}}(d_l, d_r)$  for the GEC is always greater than  $x_{\text{BEC}}^{\text{MAP}}(d_l, d_r)$  for the BEC and the latter converges to  $1 - r$  [20, Lm. 8]). Suppose that all the limits are taken when  $d_l, d_r \rightarrow \infty$  while  $r$  is kept fixed. Then, we have  $(1 - x^{\text{MAP}}(d_l, d_r))^{d_r - 1} \rightarrow 0$  exponentially fast in  $d_r$ .

Next, one also sees that

$$L(y(x^{\text{MAP}}(d_l, d_r))) = (1 - (1 - x^{\text{MAP}}(d_l, d_r))^{d_r - 1})^{d_l} \rightarrow 1 \text{ and } \lambda(y(x^{\text{MAP}}(d_l, d_r))) \rightarrow 1 \quad (3.32)$$

which can be obtained from

$$\frac{\log(1 - (1 - x^{\text{MAP}}(d_l, d_r))^{d_r - 1})}{1/(d_r - 1)} \rightarrow 0. \quad (3.33)$$

To see (3.33), we apply L'Hôpital's rule and use the fact that

$$\frac{(1 - x^{\text{MAP}}(d_l, d_r))^{d_r - 1}}{(1 - (1 - x^{\text{MAP}}(d_l, d_r))^{d_r - 1})/(d_r - 1)^2} \rightarrow 0$$

because the numerator vanishes exponentially while the denominator only vanishes quadratically in  $d_r$ .

Note that for  $(d_l, d_r)$ -regular ensemble, (3.28) can be rewritten as

$$P(x) = \int_0^{L(y)} f(t, \epsilon(x)) dt + \frac{d_l}{d_r} (1 - x)^{d_r - 1} (1 + (d_r - 1)x) - \frac{d_l}{d_r} = 0. \quad (3.34)$$

Therefore, we can use  $P(x^{\text{MAP}}(d_l, d_r)) = 0$  and (3.34), (3.32) to have

$$\int_0^1 f(t, \epsilon(x^{\text{MAP}}(d_l, d_r))) dt \rightarrow \frac{d_l}{d_r} = 1 - r.$$

In addition, from the definition of SIR in (3.2), we have  $\int_0^1 f(t, \bar{\epsilon}^{\text{MAP}}(d_l, d_r)) = 1 -$

Table III. Thresholds of  $(d_l, d_r)$ -regular ensembles over the DEC, pDEC and PR2EC.

$(d_l, d_r)$ - regular	DEC			pDEC			PR2EC		
	$\epsilon^{\text{BP}}$	$\bar{\epsilon}^{\text{MAP}}$	$\epsilon^{\text{SIR}}$	$\epsilon^{\text{BP}}$	$\bar{\epsilon}^{\text{MAP}}$	$\epsilon^{\text{SIR}}$	$\epsilon^{\text{BP}}$	$\bar{\epsilon}^{\text{MAP}}$	$\epsilon^{\text{SIR}}$
(3, 6)	0.5689	0.6387	0.6404	0.5288	0.6388	0.6404	0.7056	0.7515	0.7530
(5, 10)	0.4647	0.6404	0.6404	0.4377	0.6404	0.6404	0.6275	0.7530	0.7530
(3, 27)	0.2318	0.2642	0.2651	0.2143	0.2642	0.2651	0.4221	0.4423	0.4446
(5, 45)	0.1921	0.2651	0.2651	0.1808	0.2651	0.2651	0.3831	0.4445	0.4446

$I_s(\bar{\epsilon}^{\text{MAP}}(d_l, d_r))$  and  $1 - I_s(\epsilon^{\text{SIR}}(r)) = 1 - r$ . Therefore,

$$I_s(\bar{\epsilon}^{\text{MAP}}(d_l, d_r)) \rightarrow I_s(\epsilon^{\text{SIR}}(r))$$

and one has  $\bar{\epsilon}^{\text{MAP}}(d_l, d_r) \rightarrow \epsilon^{\text{SIR}}(r)$  as  $I_s(\cdot)$  is a continuous and monotone function.  $\square$

**Example 15.** Let us consider the DEC. For rate one-half ensembles,  $\bar{\epsilon}^{\text{MAP}}(3, 6) \approx 0.638659$ ,  $\bar{\epsilon}^{\text{MAP}}(4, 8) \approx 0.640163$ ,  $\bar{\epsilon}^{\text{MAP}}(5, 10) \approx 0.640355$ ,  $\bar{\epsilon}^{\text{MAP}}(7, 14) \approx 0.640387$ ,  $\bar{\epsilon}^{\text{MAP}}(8, 16) \approx 0.640388$  that quickly approach  $\epsilon^{\text{SIR}}(\frac{1}{2}) \approx 0.640388$ . This can be partially seen in Fig. 18 where  $\bar{\epsilon}^{\text{MAP}}(4, 8)$  is already very close to  $\epsilon^{\text{SIR}}$ . Estimates of  $\epsilon^{\text{BP}}$ ,  $\bar{\epsilon}^{\text{MAP}}$  and  $\epsilon^{\text{SIR}}$  for several regular LDPC ensembles over the DEC, pDEC, and PR2EC can be found in Table III.

### 3. Tightness of the Upper Bound

In this section, we discuss the tightness of the  $\bar{\epsilon}^{\text{MAP}}$  bounding technique. Assume that the joint BP decoder is run on the joint graph of the LDPC code and GEC. Since one never gets errors in the GEC, the joint BP decoder must reach a FP where no more bit nodes can be decoded. At this FP, one obtains a residual graph (see [6, Ch. 3]) by removing all the known bit nodes as well as their neighboring check nodes and

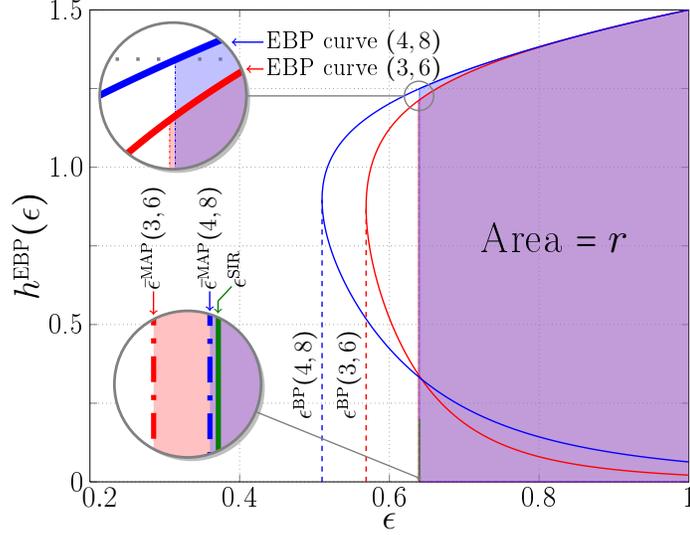


Fig. 18. EBP-EXIT curves for (3, 6) and (4, 8) regular LDPC ensembles over the DEC. Projection of the left most point of the curves on to the  $\epsilon$ -axis allows one to determine  $\epsilon^{\text{BP}}$ . Setting the area under the EBP curves to be equal to the design rate  $r$  allows one to find the upper bound  $\bar{\epsilon}^{\text{MAP}}$ .

the edges connecting them. Then, one can follow the general procedure to show that the MAP bounding technique is tight, i.e., by seeing at channel erasure rate  $\bar{\epsilon}^{\text{MAP}}$ , the design rate of the residual graph is zero and providing numerical evidence that for this residual graph, the actual rate converges to the design rate as the blocklength  $n \rightarrow \infty$ .

We start with the following lemma.

**Lemma 16.** *Consider a d.d. pair  $(\lambda, \rho)$  and the GEC with channel erasure rate  $\epsilon$ . First, run the joint BP decoder until it reaches a FP so that we obtain a residual graph. Next, use the remaining channel constraints to merge all bit nodes that must*

have the same value. The expected check node d.d. of the residual graph<sup>2</sup> is given by

$$\tilde{R}_\epsilon(z) = R(1 - x + zx) - R(1 - x) - zxR'(1 - x) \quad (3.35)$$

where  $x$  is the FP of DE and  $y = 1 - \rho(1 - x)$ . Furthermore, if the expected bit node d.d. is

$$\tilde{L}_\epsilon(z) = \int_0^{L(yz)} f(t, \epsilon) dt \quad (3.36)$$

then at  $\epsilon = \bar{\epsilon}^{MAP}$ , the design rate of the residual graph  $\tilde{r}_{\bar{\epsilon}^{MAP}}$  equals zero.

*Proof.* Consider the original graph at the FP and let  $x$  be the average erasure rate from a bit node to a check node. Pick a check node of degree  $j$  in the original graph. We can obtain a check node of degree  $i \leq j$  in the residual graph by removing all  $(j - i)$  edges with known values. Note that  $i \geq 2$  since a check node of degree one must not be in the residual graph. The remaining  $i$  edges of this check node must contain erasure messages. The probability for this event is  $\binom{j}{i}(1 - x)^{j-i}x^i$ . Thus, the check node d.d. for the residual graph (normalized by the number of check nodes in the original graph) is<sup>3</sup>

$$\begin{aligned} \tilde{R}_\epsilon(z) &= \sum_{j \geq 2} R_j \sum_{i=2}^j \binom{j}{i} (1 - x)^{j-i} (xz)^i \\ &= R(1 - x + zx) - R(1 - x) - zxR'(1 - x) \end{aligned}$$

and (3.35) holds.

Suppose that the bit node d.d. is given by (3.36),  $\tilde{L}'_\epsilon(z) = y'L'(yz)f(L(yz), \epsilon)$

---

<sup>2</sup>The check node and bit node d.d. are normalized with respect to the number of nodes in the original graph.

<sup>3</sup>This formula is the same as the check node d.d. for residual graph left by the peeling decoder for the BEC, obtained via solving a differential equation in [15].

and  $\tilde{R}'_\epsilon(z) = xR'(1-x+zx) - xR'(1-x)$ . Therefore, one obtains

$$\begin{aligned} \frac{\tilde{L}'_\epsilon(1)}{\tilde{R}'_\epsilon(1)} &= \frac{yL'(y)f(L(y), \epsilon)}{xR'(1)(1-\rho(1-x))} \\ &= \frac{L'(1)}{R'(1)} \cdot \frac{\lambda(y)f(L(y), \epsilon)}{x} \\ &= \frac{L'(1)}{R'(1)} \end{aligned} \quad (3.37)$$

by using (3.7),  $y = 1 - \rho(1-x)$  and the known facts that  $L'(y) = L'(1)\lambda(y)$  and  $R'(1-x) = R'(1)\rho(1-x)$ .

Note that the standard d.d. pair from the node perspective of the residual graph is  $\left(\frac{\tilde{L}_\epsilon(z)}{\tilde{L}_\epsilon(1)}, \frac{\tilde{R}_\epsilon(z)}{\tilde{R}_\epsilon(1)}\right)$  and the corresponding design rate is then

$$\tilde{r}_\epsilon = 1 - \frac{\tilde{L}'_\epsilon(1)}{\tilde{R}'_\epsilon(1)} \cdot \frac{\tilde{R}_\epsilon(1)}{\tilde{L}_\epsilon(1)}.$$

Using (3.37), it now is clear that

$$\tilde{r}_\epsilon = 1 - \frac{L'(1)}{R'(1)} \cdot \frac{\tilde{R}_\epsilon(1)}{\tilde{L}_\epsilon(1)} = \frac{P(x)}{\tilde{L}_\epsilon(1)}$$

where the last equality follows from (3.35), (3.36) and (3.28).

For the special case  $\epsilon = \bar{\epsilon}^{\text{MAP}}$ , one has  $\tilde{r}_{\bar{\epsilon}^{\text{MAP}}} = P(x^{\text{MAP}})/\tilde{L}_{\bar{\epsilon}^{\text{MAP}}}(1) = 0$ .  $\square$

**Remark 17.** For the BEC, the bit node d.d. given in (3.36) matches the known result in [6, Th. 3.106]. In fact, this also holds for the DEC case which can be shown by the following lemma.

**Lemma 17.** Consider a d.d. pair  $(\lambda, \rho)$  and the DEC with erasure probability  $\epsilon$ . The expected bit node d.d. in this case follows the form (3.36), i.e.,

$$\tilde{L}_\epsilon(z) = \frac{2\epsilon^2 L(yz)}{2 - L(yz)(1-\epsilon)} = \sum_{k=0}^{\infty} \epsilon^2 \left(\frac{1-\epsilon}{2}\right)^k L(yz)^{k+1} \quad (3.38)$$

Consequently, at  $\epsilon = \bar{\epsilon}^{\text{MAP}}$  the design rate of the residual graph equals zero.

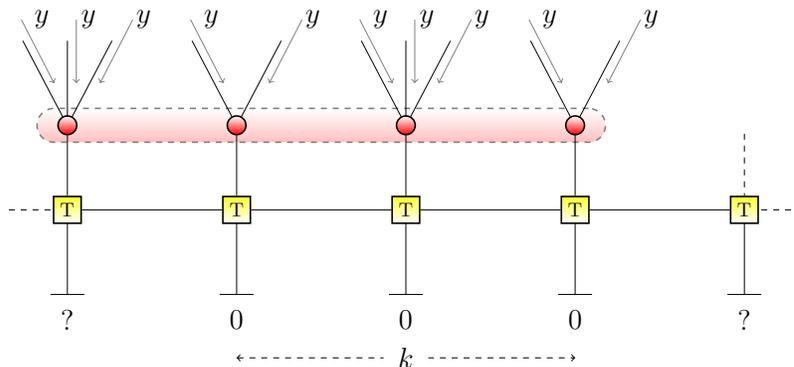


Fig. 19. A trellis section in the residual graph for the DEC. The notation “?” denotes that an erasure is received at the channel output. One can form a larger bit node by merging all the bit nodes that attach to this trellis section.

*Proof.* Consider a bit node in the original graph. To remain in the residual graph, this bit node must connect to a trellis section of the form depicted in Fig. 19 for some  $k \in \mathbb{N}$ . More specifically, the observation sequence must be  $(?, 0, \dots, 0, ?)$  and all the check-to-bit messages to this section must be erasures. Otherwise, the joint BP decoder can still decode all the bit nodes in this section.

The probability of such a observation sequence is  $\epsilon^2 \left(\frac{1-\epsilon}{2}\right)^k$ . Also, if all messages from check nodes to the bit nodes that attach to this trellis section are “?” then *all* these bit nodes remain in the residual graph. On the other hand, if at least one of the messages is not “?”, then the joint BP decoder can decode and then remove *all* these bit nodes from the residual graph. Therefore, one can consider all the bit nodes that attach to such a trellis section as one larger bit node whose degree is the sum of the  $k + 1$  component degrees. The generating function for this sum of  $k + 1$  i.i.d. random variables is  $L(z)^{k+1}$ . This is quite similar to the graph reduction technique discussed in [89] for IRA/ARA codes.

Therefore, since each edge is associated with erasure rate  $y$ , the d.d. (normalized by the number of bit nodes in the original graph) of residual graph after graph

reduction is then given by

$$\tilde{L}_\epsilon(z) = \sum_{k=0}^{\infty} \epsilon^2 \left(\frac{1-\epsilon}{2}\right)^k L(yz)^{k+1} = \frac{2\epsilon^2 L(yz)}{2 - L(yz)(1-\epsilon)}.$$

□

From the above analysis, once one has  $\tilde{r}_{\bar{\epsilon}^{\text{MAP}}} = 0$ , the final missing piece to prove the tightness<sup>4</sup> of the MAP upper bound is to show that the actual rate of the residual graph equals its design rate with high probability as the blocklength tends to  $\infty$ . While a general proof for this still requires some analytic work, one can use the following test to numerically verify if this is true.

**Lemma 18.** [77, Lm. 7-8] *Let  $\mathcal{C}_n$  be chosen uniformly at random from the ensemble LDPC( $n, \lambda, \rho$ ) and let  $r(\mathcal{C}_n)$  be its rate. Let  $r(\lambda, \rho)$  be the design rate of the ensemble. Consider the function*

$$\Psi(u) = -L'(1) \log_2 \left(\frac{1+uv}{1+v}\right) + \sum_i L_i \log_2 \left(\frac{1+u^i}{2}\right) + \frac{L'(1)}{R'(1)} \sum_j R_j \log_2 \left[1 + \left(\frac{1-v}{1+v}\right)^j\right]$$

where  $v = \left(\sum_i \frac{\lambda_i u^{i-1}}{1+u^i}\right) / \left(\sum_i \frac{\lambda_i}{1+u^i}\right)$ .

Assume that  $\Psi(u)$  takes on its maximum in the range  $u \in [0, 1]$  at  $u = 1$ . Then there exists  $B > 0$  such that, for any  $\xi > 0$ , and  $n > n_0(\xi, \lambda, \rho)$ , sufficiently large,

$$\Pr\{|r(\mathcal{C}_n) - r(\lambda, \rho)| > \xi\} \leq e^{-Bn\xi}.$$

Moreover, there exists  $C > 0$  such that, for  $n > n_0(\xi, \lambda, \rho)$

$$\mathbb{E}[|r(\mathcal{C}_n) - r(\lambda, \rho)|] \leq C \frac{\log n}{n}.$$

Therefore, to show the tightness of the upper bound, one just needs to show that

---

<sup>4</sup>If this is true, then the MAP decoder can decode perfectly for all  $\epsilon < \bar{\epsilon}^{\text{MAP}}$  and therefore  $\bar{\epsilon}^{\text{MAP}} = \epsilon^{\text{MAP}}$ .

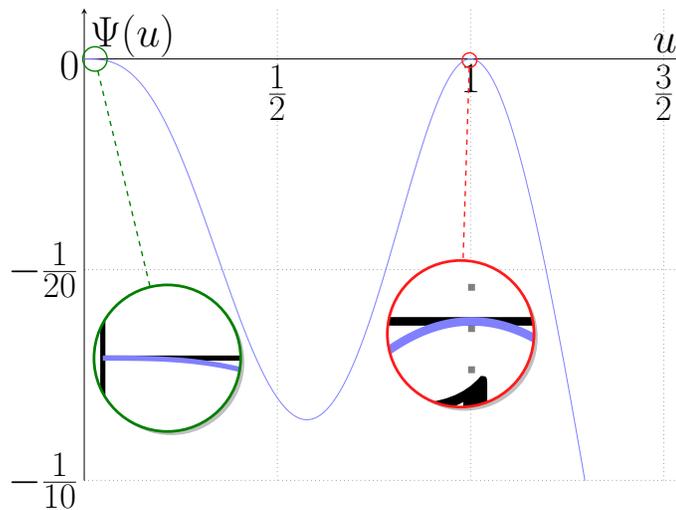


Fig. 20. Function  $\Psi(u)$  for the residual graph obtained after joint BP decoding of the  $(3,6)$ -regular LDPC ensemble over the DEC. This shows numerically that the MAP upper bound is tight in this case.

the function  $\Psi(u)$  in Lemma 18 for the residual graph has the following property:  $\Psi(u) \leq 0$  in the interval  $[0, 1]$  with equality only at  $u = 0$  and  $u = 1$ .

**Remark 18.** For a graph whose design rate is zero, one can use the following procedure to examine if  $\Psi(u)$  satisfies the property in Lemma 18. For some small  $\delta > 0$ , one can numerically verify  $\Psi(u) < 0$  for all  $u \notin [0, \delta] \cup [1 - \delta, 1]$  but at  $u = 0, 1$  one can instead show that  $\Psi'(u) = 0$  and  $\Psi''(u) < 0$ .

For our case, the bit node d.d. for the residual graph from (3.36) typically have unbounded degrees as seen in (3.38) for the DEC case. However, for this DEC and  $(d_l, d_r)$ -regular ensemble, the fraction of bit nodes with degree  $d_l(k+1)$  is upper bounded by  $(\frac{1}{2})^k$ . Therefore,  $\tilde{L}_\epsilon(z)$  has an exponentially vanishing tail, and one can truncate the series  $\tilde{L}_\epsilon(z)$  to obtain the result with a negligible error. For example, one can truncate  $\tilde{L}_\epsilon(z)$  at  $k = 20$  and for the  $(3,6)$ -regular ensemble, the truncated version of  $\Psi(u)$  is numerically shown in Fig. 20 to satisfy the desired property.

#### 4. Spatially-Coupled Codes for the GEC

Consider the  $(d_l, d_r, L, w)$  spatially-coupled ensemble over the GEC. The joint code and channel graph is similar to the one in Fig. 16 which is for the  $(d_l, d_r, L)$  ensemble. We also follow the DE equation discussed in [62] to compute the BP thresholds of the coupled ensembles. The main difference is that we use well-defined EBP curves with area theorems instead of the EXIT-like curves as in [62]. Let  $x_i^{(\ell)}$  denote the expected erasure rate at iteration  $\ell$  from bit nodes at position  $i$  to check nodes. For  $i \notin [1, L]$ , we set  $x_i^{(\ell)} = 0$ . Let us define

$$g(x_{i-w+1}, \dots, x_{i+w-1}) \triangleq \left( 1 - \frac{1}{w} \sum_{j=0}^{w-1} \left( 1 - \frac{1}{w} \sum_{k=0}^{w-1} x_{i+j-k} \right)^{d_r-1} \right)^{d_l-1},$$

$$\Gamma(x_{i-w+1}, \dots, x_{i+w-1}) \triangleq \left( 1 - \frac{1}{w} \sum_{j=0}^{w-1} \left( 1 - \frac{1}{w} \sum_{k=0}^{w-1} x_{i+j-k} \right)^{d_r-1} \right)^{d_l}.$$

The DE equation for the joint BP decoder can be written as

$$x_i^{(\ell+1)} = f(\Gamma(x_{i-w+1}^{(\ell)}, \dots, x_{i+w-1}^{(\ell)}), \epsilon) \cdot g(x_{i-w+1}^{(\ell)}, \dots, x_{i+w-1}^{(\ell)})$$

for  $i \in [1, L]$ . To compute both the stable and unstable FPs of DE, one can use the fixed entropy DE procedure outlined in [90, Sec. VIII] where the normalized entropy of a constellation  $\underline{x}^{(\ell)} = (x_1^{(\ell)}, \dots, x_L^{(\ell)})$ , which is defined as  $\chi(\underline{x}^{(\ell)}) = \frac{1}{L} \sum_{i=1}^L x_i^{(\ell)}$ , is kept constant at every iteration by varying the channel parameter. With each FP  $\underline{x}$  obtained, one can compute the EBP-EXIT value of the spatially-coupled ensemble as  $\frac{1}{L} \sum_{i=1}^L h^{\text{EBP}}(x_i)$ .

The threshold saturation effect of coupling can be nicely seen by plotting the (E)BP-EXIT curves for the uncoupled and coupled codes. For the DEC, Fig. 21 shows the EBP curves for the  $(3, 6, L, 5)$  ensembles with various  $L$  along with the EBP curve of the underlying  $(3, 6)$ -regular ensemble. From the EBP curves, one can de-

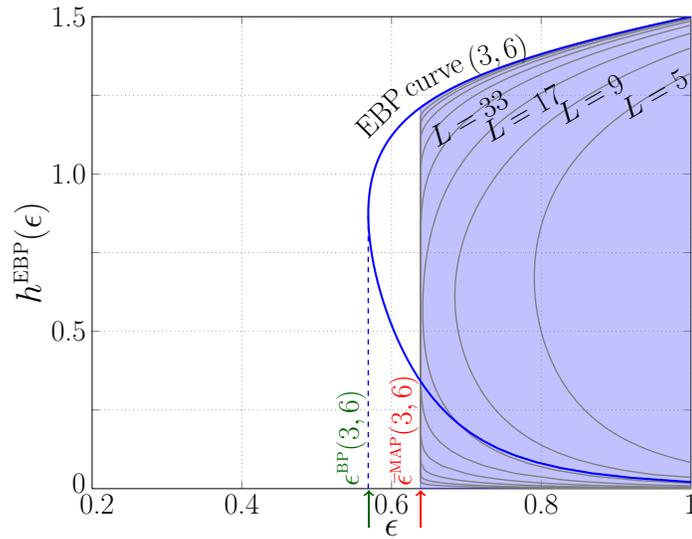


Fig. 21. EBP-EXIT curves for  $(3, 6, L, 5)$  over the DEC with  $L = 2\hat{L} + 1$  where  $\hat{L} = 2, 4, 8, 16, 32, 64, 128, 246$ . For small values of  $L$ , the increase in threshold can be explained by the large rate-loss. As  $L$  grows larger, the rate loss becomes negligible and the curves keep moving left, but they saturate at the MAP threshold of the underlying regular ensemble.

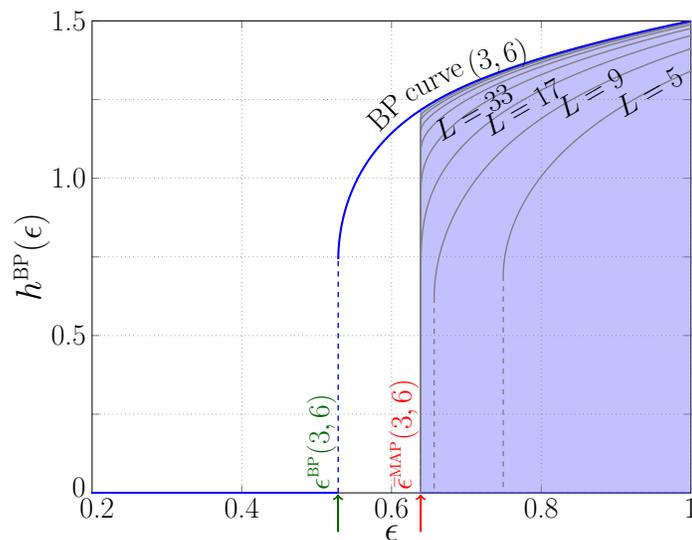


Fig. 22. BP-EXIT curves for  $(3, 6, L, 5)$  over the pDEC with  $L = 2\hat{L} + 1$  where  $\hat{L} = 2, 4, 8, 16, 32, 64, 128, 246$ . Threshold saturation can also be observed for this case.

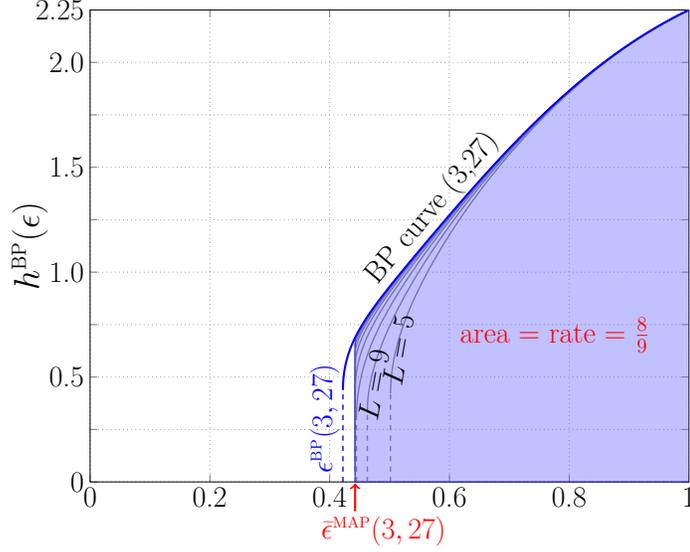


Fig. 23. BP-EXIT curves for  $(3, 27, L, 5)$  over the PR2EC with  $L = 2\hat{L} + 1$  where  $\hat{L} = 2, 4, 8, 16, 32, 64, 128, 246$ .

termine  $\epsilon^{\text{BP}}(3, 6) \approx 0.56892$  and  $\bar{\epsilon}^{\text{MAP}}(3, 6) \approx 0.63866$ . The BP thresholds of spatially-coupled ensembles for small  $L$  can be even larger due to rate-loss, e.g.,  $\epsilon^{\text{BP}}(3, 6, 17, 6) \approx 0.64170 > \bar{\epsilon}^{\text{MAP}}(3, 6)$ . However, for a wide range of  $L$ , i.e.,  $L = 33, 65, 129, 257, 513$ , we observe that  $\epsilon^{\text{BP}}(3, 6, L, 5) \approx 0.63866$  which is essentially  $\bar{\epsilon}^{\text{MAP}}(3, 6)$  while the rate loss gradually becomes insignificant. In [62], Kudekar and Kasai provided a similar plot but here we include the MAP threshold estimate  $\bar{\epsilon}^{\text{MAP}}$  and use the EXIT function  $h^{\text{EBP}}$  instead of the EXIT-like  $L(y)$  in [62]. Similarly, one can also verify the threshold saturation over the pDEC as seen in Fig. 22. For this channel, the BP threshold for  $(3, 6)$ -regular ensemble is  $\epsilon^{\text{BP}}(3, 6) \approx 0.52877$  and, with spatial coupling, the BP threshold is improved to  $\epsilon^{\text{BP}}(3, 6, L, 5) \approx \bar{\epsilon}^{\text{MAP}}(3, 6) \approx 0.63877$  with a negligible loss in rate as  $L \rightarrow \infty$ . In a similar fashion, Fig. 23, plotted for a high-rate example based on the  $(3, 27)$ -regular ensemble, strongly suggests that the threshold saturation effect also occurs for the PR2EC.

Even though the threshold saturation effect has been only shown numerically

for several example channels, the above procedure that includes computing the BP thresholds of spatially-coupled codes and the MAP threshold estimates of their underlying ensembles is readily applicable to the entire family of GECs. Still, the analytic proof for threshold saturation remains open for the GEC. Combining such a proof with Theorem 10 would demonstrate the SIR-achieving capability and universality of spatially-coupled ensembles.

#### D. General ISI Channels

In this section, we shift our focus to ISI channels with more realistic noise models. The MAP upper bound for general binary *memoryless* symmetric channels was presented by Méasson *et al.* and conjectured to be tight [90]. For general ISI channels, we apply a similar technique to give an estimate of the MAP threshold of the underlying uncoupled ensemble by first constructing the BP-GEXIT curve that follows an area theorem. While our method can be used for a wide range of noise models, we particularly focus on the case of AWGN. The BP thresholds of the corresponding coupled ensembles are then computed via DE and the threshold saturation effect is also observed. In addition, simulations of the joint BP decoder for SC codes of finite length are described that validate these thresholds.

##### 1. GEXIT Curves for the ISI channels

Consider an ISI channel of memory  $\nu$ . When the channel input  $X_1^n$  is chosen uniformly at random from a suitable<sup>5</sup> binary linear code  $\mathcal{C}_n$ , the ISI channel output without noise  $Z_i$  at some index  $i$  is a discrete random variable characterized by its probability mass

---

<sup>5</sup>The code is proper [6, p. 14] and its dual code contains no codewords involving only 0's and a run of  $(\nu + 1)$  1's.

function  $p_{Z_i}(z)$  for all values  $z$  in the alphabet  $\mathcal{Z}$ . For example, in the case of a di-code channel,  $\mathcal{Z} = \{0, +2, -2\}$  and  $p_{Z_i}(0) = \frac{1}{2}, p_{Z_i}(+2) = p_{Z_i}(-2) = \frac{1}{4}$ . The channel from  $Z_i$  to  $Y_i$  is a  $|\mathcal{Z}|$ -ary input memoryless channel characterized by its transition probability density  $p_{Y_i|Z_i}(y|z)$ . Without specifying the index, we denote  $\mathbf{h} \triangleq H(Z|Y)$  and get

$$\begin{aligned} \mathbf{h} &= H(Z) - I(Z; Y) \\ &= H(Z) - \int_{-\infty}^{\infty} \sum_z p(z) p(y|z) \log_2 \left\{ \frac{p(y|z)}{\sum_{z'} p(z') p(y|z')} \right\} dy, \end{aligned}$$

where  $p(z)$  and  $p(y|z)$  are the shorthand notations of  $p_Z(z)$  and  $p_{Y|Z}(y|z)$ , respectively.

Instead of looking at a particular channel, we assume that the channel from  $Z_i$  to  $Y_i$  is from a smooth family  $\{M(\mathbf{h}_i)\}_{\mathbf{h}_i}$  of  $|\mathcal{Z}|$ -ary input memoryless channels characterized by conditional entropy  $\mathbf{h}_i$ . A further assumption is made that all individual channel families are parametrized in a smooth way by a common parameter<sup>6</sup>  $\epsilon$ , i.e.,  $\mathbf{h}_i = H(Z_i|Y_i)(\epsilon)$ .

With the convention that  $y_{\sim i} \triangleq y_1^n \setminus y_i$ , define  $\phi_i(y_{\sim i}) \triangleq \{P_{Z_i|Y_{\sim i}}(z|y_{\sim i}) : z \in \mathcal{Z}\}$  and the random vector  $\Phi_i \triangleq \phi_i(Y_{\sim i})$ . Each value of  $\phi_i$  is a vector of length  $|\mathcal{Z}|$  in the  $(|\mathcal{Z}| - 1)$ -dimensional probability simplex. The index of the vector associated with  $z \in \mathcal{Z}$  is denoted by  $[z]$ . It is easy to verify that  $\Phi_i$  is a sufficient statistic for estimating  $Z_i$  given  $Y_{\sim i}$ , i.e.,  $Z_i \rightarrow \Phi_i(Y_{\sim i}) \rightarrow Y_{\sim i}$  forms a Markov chain<sup>7</sup>.

**Definition 12.** *Suppose the initial state in the trellis is  $S_0$ . Let  $X_1^n$ , chosen from*

<sup>6</sup>For AWGN case, a convenient choice for  $\epsilon$  is  $\epsilon = -\frac{1}{2\sigma^2}$ .

<sup>7</sup>To see this, write

$$P_{Y_{\sim i}|Z_i}(y_{\sim i}|z_i) = \frac{P_{Z_i|Y_{\sim i}}(z_i|y_{\sim i})}{P_{Z_i}(z_i)} P_{Y_{\sim i}}(y_{\sim i}) = \frac{\Phi_i \cdot e_{[z_i]}^T}{P_{Z_i}(z_i)} P_{Y_{\sim i}}(y_{\sim i}),$$

where  $e_{[z]}^T$  is the standard basis column vector with a 1 in the index  $[z]$ , and apply the result from [6, p. 29].

code  $\mathcal{C}_n$  according to  $p_{X_1^n}(x_1^n)$ , be the input sequence,  $Z_1^n$  be the ISI output sequence without noise and  $Y_1^n$  be the final channel output sequence, i.e.,  $Y_i$  is the result of transmitting  $Z_i$  over the smooth family  $\{M(\mathbf{h}_i)\}_{\mathbf{h}_i}$  of memoryless channels. Then the  $i$ -th GEXIT function is

$$\mathbf{G}_i(\mathbf{h}_1, \dots, \mathbf{h}_n) = \frac{\partial}{\partial \mathbf{h}_i} H(X_1^n | Y_1^n(\mathbf{h}_1, \dots, \mathbf{h}_n), S_0) \quad (3.39)$$

and the average GEXIT function is defined by

$$\mathbf{G}(\mathbf{h}_1, \dots, \mathbf{h}_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{G}_i(\mathbf{h}_1, \dots, \mathbf{h}_n).$$

For the case where all channel families are the same, i.e.,  $\mathbf{h}_i = \mathbf{h}$ , we have

$$\mathbf{G}(\mathbf{h}) = \frac{1}{n} \cdot \frac{d}{d\mathbf{h}} H(X_1^n | Y_1^n(\mathbf{h}), S_0).$$

**Theorem 11.** *The above form of the GEXIT function naturally conforms with a generalized area theorem that gives*

$$\frac{1}{n} H(X_1^n | Y_1^n(\mathbf{h}^*), S_0) = \int_0^{\mathbf{h}^*} \mathbf{G}(\mathbf{h}) d\mathbf{h}.$$

*If one considers a special case that assumes  $\mathbf{h}^* = H(Z)$  and a uniform prior on the set of the codewords then one has*

$$\int_0^{H(Z)} \mathbf{G}(\mathbf{h}) d\mathbf{h} = r$$

*which tells that the area under the GEXIT curve equal to the rate of the code.*

After defining the GEXIT function that follows the generalized area theorem, it is now possible to analyze the GEXIT curve and use the MAP bounding technique discussed above.

**Lemma 19.** Assume that all the channel families are the same<sup>8</sup>, i.e.,  $\mathbf{h}_i = \mathbf{h}$ . The  $i$ -th GEXIT function is given by

$$\mathbf{G}_i(\mathbf{h}) = \sum_z p(z) \int_{\underline{v}} \mathbf{a}_{i,z}(\underline{v}) \kappa_{i,z}(\underline{v}) d\underline{v}$$

where  $\mathbf{a}_{i,z}$  is the distribution of the vector  $\Phi_i$  given  $Z_i = z$ ,  $\underline{v}$  is a vector of length  $|\mathcal{Z}|$  in the  $(|\mathcal{Z}| - 1)$ -dimensional probability simplex and the GEXIT kernel (for  $i$  and  $z$ ) is<sup>9</sup>

$$\kappa_{i,z}(\underline{v}) = \frac{\int_{-\infty}^{\infty} \frac{d}{d\epsilon} p(y_i|z) \log_2 \left\{ \frac{\sum_{z'} v_{[z']} p(y_i|z')}{v_{[z]} p(y_i|z)} \right\} dy_i}{\int_{-\infty}^{\infty} \sum_z p(z) \frac{d}{d\epsilon} p(y_i|z) \log_2 \left\{ \frac{\sum_{z'} p(z') p(y_i|z')}{p(z) p(y_i|z)} \right\} dy_i}.$$

*Proof.* Suppose the initial state is  $S_0$ , we start by writing

$$\begin{aligned} H(X_1^n | Y_1^n, S_0) &= H(Z_1^n | Y_1^n, S_0) \\ &= H(Z_i | Y_1^n, S_0) + H(Z_{\sim i} | Y_1^n, Z_i, S_0). \end{aligned} \quad (3.40)$$

To simplify notation, we drop  $S_0$  in all the expressions although the dependency on  $S_0$  is always implied. From (3.39) and (3.40), it is clear that

$$\mathbf{G}_i(\mathbf{h}) = \frac{\partial}{\partial \mathbf{h}_i} H(Z_i | Y_1^n). \quad (3.41)$$

since  $H(Z_{\sim i} | Y_1^n, Z_i, S_0)$  does not depend on  $\mathbf{h}_i$ .

We also have

$$\begin{aligned} H(Z_i | Y_1^n) &= H(Z_i | Y_i, \Phi_i(Y_{\sim i})) \\ &= - \int_{\phi_i} \int_{y_i} \sum_{z_i} p(z_i) p(\phi_i | z_i) p(y_i | z_i) \log_2 \left\{ \frac{p(z_i | \phi_i) p(y_i | z_i)}{\sum_{z'_i} p(z'_i | \phi_i) p(y_i | z'_i)} \right\} dy_i d\phi_i \end{aligned} \quad (3.42)$$

---

<sup>8</sup>Note that, for the case of different channel families, one can still compute the  $i$ -th GEXIT function as a function of the common parameter  $\epsilon$ .

<sup>9</sup> $p(y_i|z)$  is dependent on  $\mathbf{h}_i$  and hence is dependent on  $\epsilon$ .

where (3.42) follows from the Bayes' theorem and the fact that

$$p(z_i, \phi_i, y_i) = p(z_i, \phi_i)p(y_i|\phi_i, z_i) = p(z_i)p(\phi_i|z_i)p(y_i|z_i). \quad (3.43)$$

Note that (3.43) is true since  $Y_i$  and  $\Phi_i(Y_{\sim i})$  are independent given  $Z_i$ , i.e.,  $Y_i \rightarrow Z_i \rightarrow \Phi_i(Y_{\sim i})$ .

Taking derivative and using  $p(z_i|\phi_i) = p(z_i|y_{\sim i})$ , we get<sup>10</sup>

$$\begin{aligned} \mathbf{G}_i(\mathbf{h}) &= \sum_{z_i} p(z_i) \int_{\phi_i} p(\phi_i|z_i) \int_{y_i} \frac{d}{d\mathbf{h}_i} p(y_i|z_i) \log_2 \left\{ \sum_{z'_i} \frac{p(z'_i|y_{\sim i})p(y_i|z'_i)}{p(z_i|y_{\sim i})p(y_i|z_i)} \right\} dy_i d\phi_i \\ &= \sum_z p(z) \int_{\underline{v}} \mathbf{a}_{i,z}(\underline{v}) \kappa_{i,z}(\underline{v}) d\underline{v}. \end{aligned}$$

where

$$\begin{aligned} \kappa_{i,z}(\underline{v}) &= \int_{y_i} \frac{d}{d\mathbf{h}_i} p(y_i|z) \log_2 \left\{ \frac{\sum_{z'} v_{[z']} p(y_i|z')}{v_{[z]} p(y_i|z)} \right\} dy_i \\ &= \int_{y_i} \frac{d}{d\epsilon} p(y_i|z) \log_2 \left\{ \frac{\sum_{z'} v_{[z']} p(y_i|z')}{v_{[z]} p(y_i|z)} \right\} dy_i \left( \frac{d\mathbf{h}_i}{d\epsilon} \right)^{-1}. \end{aligned}$$

Finally, by seeing that

$$\begin{aligned} \frac{d\mathbf{h}_i}{d\epsilon} &= \frac{dH(Z_i|Y_i(\epsilon))}{d\epsilon} \\ &= \sum_z \int_{y_i} p(z) \frac{d}{d\epsilon} p(y_i|z) \log_2 \left\{ \frac{\sum_{z'} p(z') p(y_i|z')}{p(z) p(y_i|z)} \right\} dy_i. \end{aligned}$$

we obtain the result.  $\square$

**Remark 19.** For erasure noise and the GEC in particular,  $\mathbf{h} = H(Z|Y) = \epsilon H(Z)$  (scaling  $\epsilon$  by  $H(Z)$ ) and since in this case

$$\kappa_{i,z}(\underline{v}) = \frac{1}{H(Z)} \log_2 \left\{ 1 + \frac{\sum_{z' \neq z} v_{[z']}}{v_{[z]}} \right\},$$

---

<sup>10</sup>One can verify that the terms obtained by taking derivative with respect to the channel inside the  $\log_2$  vanish.

$\mathbf{G}(\mathbf{h}) = \frac{h(\epsilon)}{H(Z)}$  (scaling  $h(\epsilon)$  by  $\frac{1}{H(Z)}$ ) where  $h(\epsilon)$  is the EXIT function for the GEC.

**Remark 20.** For AWGN with  $\sigma = 0$  (or erasure noise with  $\epsilon = 0$ ),  $\mathbf{h} = 0$  and  $\mathbf{a}_{i,z}$  is a “delta at  $\underline{v} = e_{[z]}$ ” where  $e_{[z]}$  is the standard basis vector. At this extreme,  $\mathbf{G}(0) = 0$  because  $\kappa_{i,z}(\underline{v}) = 0$ . At the  $\sigma \rightarrow \infty$  extreme for AWGN (or  $\epsilon = 1$  for erasure noise),  $\mathbf{h} = H(Z)$  (e.g., 1.5 for the dicode channel) and  $\mathbf{G}(\mathbf{h}) = 1$  since  $\mathbf{a}_{i,z}$  is a “delta at  $v_{[z']} = p(z') \forall z'$ ”.

a. BP-GEXIT Curve (with AWGN)

In this section, we are particularly interested in computing the asymptotic BP-GEXIT function for ISI channels with AWGN. In this case, let  $\Phi_i^{\text{BP},\ell,W}$  denote the extrinsic estimate of  $Z_i$  at the  $\ell$ th round of joint BP decoding that employs a windowed BCJR detector of size  $W$ . If  $\Phi_i^{\text{BP},\ell,W}$  is used instead of  $\Phi_i$  in the above formulas then one can compute the (asymptotic) BP-GEXIT function at the  $\ell$ th round  $\mathbf{G}^{\text{BP},\ell,W}$  in a similar manner to [90], i.e.,

$$\mathbf{G}^{\text{BP},\ell,W}(\mathbf{h}) = \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{G}_i^{\text{BP},\ell,W}(\mathbf{h}) \right] \quad (3.44)$$

where the expectation is taken over all code  $\mathcal{C}_n$  from the ensemble LDPC( $n, \lambda, \rho$ ).

The overall (asymptotic) BP-GEXIT is defined as

$$\mathbf{G}^{\text{BP}}(\mathbf{h}) = \lim_{W \rightarrow \infty} \lim_{\ell \rightarrow \infty} \mathbf{G}^{\text{BP},\ell,W}(\mathbf{h}). \quad (3.45)$$

The existence of the limit in (3.44) is implied by the fact that for a fixed window size  $W$ , the computation graph of depth  $\ell$  becomes tree-like as  $n \rightarrow \infty$  and one can apply the standard analysis of concentration around ensemble average similarly to [90, Th. 3]. Meanwhile, the limit in (3.45) also exists because of the monotonicity of the extrinsic BP estimate  $\Phi_i^{\text{BP},\ell,W}$  with respect to  $L$  and  $W$ . Also, notice that the

two extremes in Remark 20 still apply when the BP decoder is used instead of the MAP decoder.

Next, AWGN implies that  $p(y_i|z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i-z)^2}{2\sigma^2}}$  and then  $\frac{d}{d\epsilon} p(y_i|z) = ((y_i-z)^2 - \sigma^2)p(y_i|z)$  where we choose  $\epsilon = -\frac{1}{2\sigma^2}$ . Therefore, the corresponding  $i$ -th BP-GEXIT is  $G_i^{\text{BP},\ell}(\mathcal{C}, \mathbf{h}) = \frac{A}{B}$  where

$$A = \sum_z p(z) \int_{\underline{v}} \mathbf{a}_{i,z}^{\text{BP},\ell}(\underline{v}) \int_{-\infty}^{\infty} p(y_i|z) \left\{ \frac{(y_i-z)^2}{\sigma^2} - 1 \right\} \log_2 \left\{ \sum_{z'} \frac{v_{[z']}}{v_{[z]}} e^{\frac{(z'-z)(2y_i-z-z')}{2\sigma^2}} \right\} dy_i d\underline{v}$$

and

$$B = \sum_z p(z) \int_{-\infty}^{\infty} p(y_i|z) \left\{ \frac{(y_i-z)^2}{\sigma^2} - 1 \right\} \log_2 \left\{ \sum_{z'} \frac{p(z')}{p(z)} e^{\frac{(z'-z)(2y_i-z-z')}{2\sigma^2}} \right\} dy_i.$$

In the limit of  $\ell \rightarrow \infty$ , one can run the DE for ISI channels [66] to obtain the DE-FP and compute the quantities  $A$  and  $B$  at this FP. With some abuse of notation, let  $\mathbf{a}^{(\ell)}$ ,  $\mathbf{b}^{(\ell)}$ ,  $\mathbf{c}^{(\ell)}$  and  $\mathbf{d}^{(\ell)}$  denote the average density of the bit-to-check, check-to-bit, bit-to-trellis and trellis-to-bit messages, respectively (see Fig. 15), at iteration  $\ell$  with initial values (at  $\ell = 0$ ) being  $\Delta_0$ , the delta function at 0. Also, let  $\mathbf{n}$  denote the density of channel noise. The DE update equation for joint BP decoding of a general binary-input ISI channels is

$$\mathbf{a}^{(\ell)} = \mathbf{d}^{(\ell-1)} \otimes \lambda(\mathbf{b}^{(\ell-1)}),$$

$$\mathbf{b}^{(\ell)} = \rho(\mathbf{a}^{(\ell)}),$$

$$\mathbf{c}^{(\ell)} = L(\mathbf{b}^{(\ell)}),$$

$$\mathbf{d}^{(\ell)} = \Gamma(\mathbf{c}^{(\ell)}, \mathbf{n})$$

where for a density  $\mathbf{x}$ ,  $\lambda(\mathbf{x}) = \sum_i \lambda_i \mathbf{x}^{\otimes(i-1)}$ ,  $\rho(\mathbf{x}) = \sum_i \rho_i \mathbf{x}^{\boxtimes(i-1)}$  and  $L(\mathbf{x}) = \sum_i L_i \mathbf{x}^{\otimes i}$ . The operators  $\otimes$  and  $\boxtimes$  are the standard density transformations used in [6, p. 181]. The map  $\Gamma(\cdot, \cdot)$  is not easy to compute in closed form for general trellises and often

one needs to resort to the Monte Carlo methods (i.e., running the windowed BCJR algorithm with window parameter  $W$  on a long enough trellis - see details in [66]) to give the estimates. A similar method was used to upper bound the MAP threshold for turbo codes over BMS channels [91].

The denominator  $B$  can be computed either by numerical integration or by Monte Carlo methods. Meanwhile, the numerator  $A$  involves in the quantity  $v_{[z]} = p(Z_i = z | \mathbf{T}_i^\ell)$  where  $\mathbf{T}_i^\ell$  denotes the computation tree of depth  $\ell$ , rooted at index  $i$ , which includes all channel and code constraints associated with  $\ell$  iterations of decoding. This computation tree  $\mathbf{T}_i^\ell$  excludes the observation  $y_i$  from the root and is implied by the decoding schedule in the DE equation. Due to complications from the trellis, the quantity  $v_{[z]}$  is not easy to obtain in closed form. However, one can readily compute  $v_{[z]}$  as an extra output of the BCJR algorithm, which is required by DE, as

$$v_{[z]} \propto \sum_{s_i, s_{i-1}: Z_i=z} \alpha_{i-1}(s_{i-1}) \cdot \gamma_i(s_{i-1}, s_i) \cdot \beta_i(s_i),$$

where  $\gamma_i(s_{i-1}, s_i)$  is probability of the input  $x_i$  that corresponds to the transition from state  $s_{i-1}$  (at time index  $i-1$ ) to state  $s_i$  at (time index  $i$ ) given the computation tree  $\mathbf{T}_i^\ell$ . Here,  $\alpha_i(\cdot)$  and  $\beta_i(\cdot)$  are the standard forward and backward state probabilities in the BCJR algorithm. Note that the scaling constant can be chosen so that  $\sum_z v_{[z]} = 1$ .

## 2. Upper Bound on the MAP Threshold

As briefly discussed before, the above-mentioned GEXIT function (associated with a code  $\mathcal{C}$ ) naturally follows the area theorem that gives

$$\int_0^{H(Z)} \mathbf{G}(\mathbf{h}) d\mathbf{h} = r.$$

One important consequence of this area theorem is to give a good estimate of the threshold of MAP decoding, which is defined as

$$\mathbf{h}^{\text{MAP}} \triangleq \inf \left\{ \mathbf{h} \in [0, H(Z)] : \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathcal{C}} [H(X_1^n | Y_1^n(\mathbf{h}), S_0)] > 0 \right\}.$$

To do this, one can apply [90, Lm. 4] to the BMS channel from  $Z_1^n$  to  $Y_1^n$  and obtain

$$\frac{\partial H(Z_i | Y_1^n)}{\partial \mathbf{h}_i} \leq \frac{\partial H(Z_i | Y_i, \Phi_i^{\text{BP}, \ell})}{\partial \mathbf{h}_i}.$$

Consequently, by invoking (3.41), one has the optimality of the MAP decoder in the sense that  $\mathbf{G}(\mathbf{h}) \leq \mathbf{G}^{\text{BP}}(\mathbf{h})$ . Therefore, one can use the discussed bounding technique, i.e., by finding the largest value  $\bar{\mathbf{h}}^{\text{MAP}}$  such that the area under the BP-GEXIT curve equals the code rate,

$$\int_{\bar{\mathbf{h}}^{\text{MAP}}}^{H(Z)} \mathbf{G}^{\text{BP}}(\mathbf{h}) d\mathbf{h} = r,$$

to obtain the MAP upper bound  $\bar{\mathbf{h}}^{\text{MAP}} \geq \mathbf{h}^{\text{MAP}}$  (see similar arguments for the BMS case in [90, Th. 5]).

For example, the BP-GEXIT curve, for the (3,6)-regular LDPC code over an AWGN dicode channel with  $a(D) = (1 - D)/\sqrt{2}$ , was computed using the analysis in Section 1 and is shown in Fig. 24. In this case,  $\mathbf{h}^{\text{BP}}(3,6) \approx 0.851 \pm 0.001$  (the corresponding threshold measured in dB<sup>11</sup> is  $\sigma^{\text{BP}}(3,6) \approx 1.703 \pm 0.001$  dB) while  $\bar{\mathbf{h}}^{\text{MAP}}(3,6) \approx 0.920 \pm 0.001$  (or  $\bar{\sigma}^{\text{MAP}}(3,6) \approx 0.959 \pm 0.001$  dB). Similarly, for the (5,10)-regular LDPC code, one has  $\mathbf{h}^{\text{BP}}(5,10) \approx 0.716 \pm 0.001$  and  $\bar{\mathbf{h}}^{\text{MAP}}(5,10) \approx 0.931 \pm 0.001$ . For a high-rate example, the BP-EXIT curve for the (3,27)-regular LDPC ensemble is plotted in Fig. 25. The estimated BP and MAP thresholds (measured in dB) for various regular LDPC ensembles over the dicode and PR2 channels with AWGN can

---

<sup>11</sup>We adopt the convention that  $\sigma$  is the SNR threshold measured in dB, i.e.,  $\sigma = 10 \log_{10} \frac{\sum_{t=0}^{\nu} a_t^2}{\text{var}}$  where  $\text{var}$  is the noise variance.

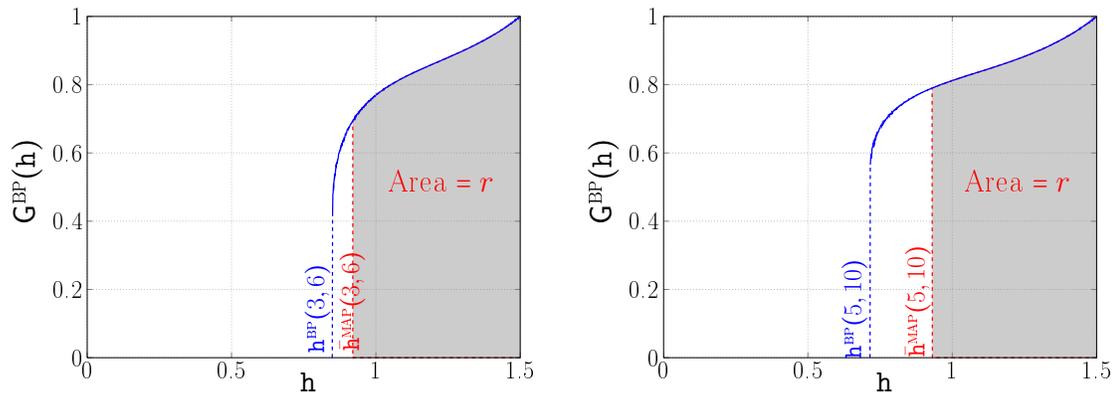


Fig. 24. The BP-GEXIT curve for  $(3, 6)$ -regular and  $(5, 10)$ -regular LDPC codes over an AWGN dicode channel with  $a(D) = (1 - D)/\sqrt{2}$ . The upper bound  $\bar{h}^{\text{MAP}}$  is obtained by setting the area under the BP-GEXIT curve (the shaded region) equal to the code rate.

be found in Table IV.

Table IV. Threshold estimates, measured in dB, of  $(d_l, d_r)$ -regular ensembles over the dicode AWGN and PR2 AWGN channels.

$(d_l, d_r)$ - regular	Dicode AWGN			$(d_l, d_r)$ - regular	PR2 AWGN		
	$\sigma^{\text{BP}}$	$\bar{\sigma}^{\text{MAP}}$	$\sigma^{\text{SIR}}$		$\sigma^{\text{BP}}$	$\bar{\sigma}^{\text{MAP}}$	$\sigma^{\text{SIR}}$
$(3, 6)$	1.73	0.96	0.82	$(3, 27)$	7.96	7.29	7.28
$(5, 10)$	3.03	0.83	0.82	$(5, 45)$	8.59	7.28	7.28

### 3. Spatially-Coupled Codes on General ISI Channels

Consider the  $(d_l, d_r, L)$  spatially-coupled ensemble. For general ISI channels, the DE equation for this ensemble can be obtained from the protograph chain in a similar manner to the case of memoryless channels discussed in [92]. For each  $i, j \in [1 - \hat{d}_l, L + \hat{d}_l]$ , let  $\mathbf{a}_{i \rightarrow j}^{(\ell)}$  (and  $\mathbf{b}_{i \leftarrow j}^{(\ell)}$ ) denote the average density of the messages from bit

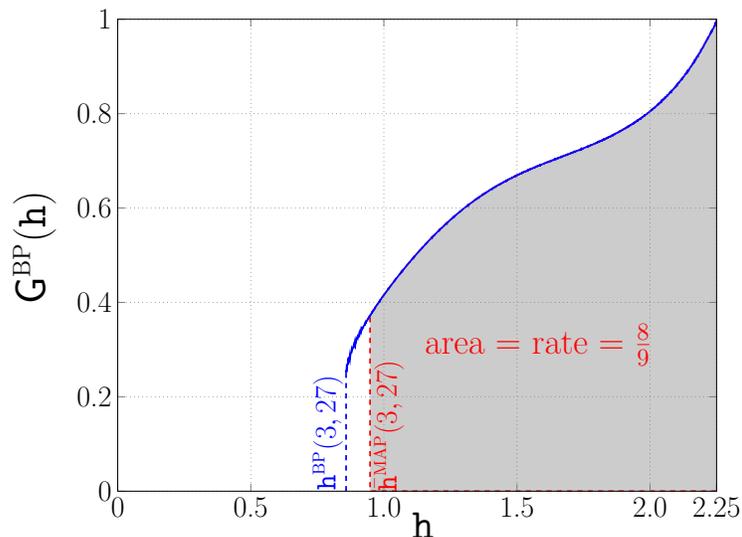


Fig. 25. A high-rate example: the BP-GEXIT curve for  $(3, 27)$ -regular LDPC codes over an AWGN PR2 channel with  $a(D) = (1+2D+D^2)/\sqrt{6}$ . The upper bound  $\bar{h}^{\text{MAP}}$  is determined by the left border of the shaded region.

nodes at position  $i$  to check nodes at position  $j$  (and the other way around)<sup>12</sup>. With all the initial message densities (at  $\ell = 0$ ) being  $\Delta_0$ , the DE update equation (for all  $i \in [1, L]$ ) is

$$\begin{aligned} \mathbf{a}_{i \rightarrow j}^{(\ell)} &= \mathbf{d}_i^{(\ell-1)} \otimes \left\{ \bigotimes_{j' \in [i-\hat{d}_l, i+\hat{d}_l] \setminus j} \mathbf{b}_{i \leftarrow j'}^{(\ell-1)} \right\}, \forall j \in [i-\hat{d}_l, i+\hat{d}_l], \\ \mathbf{b}_{i \leftarrow j}^{(\ell)} &= \bigotimes_{i' \in [j-\hat{d}_l, j+\hat{d}_l] \setminus i} \mathbf{a}_{i' \rightarrow j}^{(\ell)}, \forall j \in [i-\hat{d}_l, i+\hat{d}_l], \\ \mathbf{c}_i^{(\ell)} &= \bigotimes_{j' \in [i-\hat{d}_l, i+\hat{d}_l]} \mathbf{b}_{i \leftarrow j'}^{(\ell)}, \\ \mathbf{d}_i^{(\ell)} &= \Gamma(\mathbf{c}_i^{(\ell)}, \mathbf{n}) \end{aligned}$$

where  $\bigotimes_{j \in \{j_1, \dots, j_t\}} \mathbf{x}_j$  and  $\bigotimes_{i \in \{i_1, \dots, i_t\}} \mathbf{x}_i$  denote the operations  $\mathbf{x}_{j_1} \otimes \mathbf{x}_{j_2} \otimes \dots \otimes \mathbf{x}_{j_t}$  and  $\mathbf{x}_{i_1} \otimes \mathbf{x}_{i_2} \otimes \dots \otimes \mathbf{x}_{i_t}$ , respectively.

<sup>12</sup>For  $i \notin [1, L]$ , set  $\mathbf{a}_{i \rightarrow j}^{(\ell)} = \Delta_{+\infty}$ , the delta function at  $+\infty$ .

#### 4. Simulation Results

In this section, we start with the  $(d_l, d_r, L)$  circular ensemble obtained by considering all the positions  $i > L$  of the protograph chain to be the same as position  $i - L$  (similar to [59]). The order of bit transmissions is “left to right” in each length- $L$  row and then start with the next row (in a total of  $M$  rows, see Fig. 16). The  $I \triangleq \max(\nu, d_l - 1)$  first bits in each row are known. These known bits “break” the circular ensemble into the  $(d_l, d_r, L - I)$  ensemble and also serve as the pilot bits to fix the trellis state. As a consequence of this fixing, one only needs to run the BCJR independently in each row and this can be done in a parallel manner [73, 74].

In our experiments, we conduct simulations over the AWGN dicode channel with  $a(D) = (1 - D)/\sqrt{2}$  and memory  $\nu = 1$ . First, we use the DE in Sec. 3 to compute the BP thresholds of the spatially-coupled coding scheme. The results in Fig. 26 reveals that  $\sigma^{\text{BP}}(3, 6, 22)$  is roughly  $0.959 \pm 0.001$  dB and approximately the same as  $\sigma^{\text{BP}}(3, 6, 44)$  whose rate loss is smaller. Notice that this is also roughly  $\bar{\sigma}^{\text{MAP}}(3, 6)$  - the MAP threshold estimate of the underlying  $(3, 6)$ -regular ensemble, obtained by the bounding technique, and is a significant improvement over  $\sigma^{\text{BP}}(3, 6) \approx 1.703 \pm 0.001$  dB. This suggests that threshold saturation occurs for regular ensembles. Since MAP decoding of regular ensembles can achieve the SIR [79], it also suggests that one can universally approach the SIR of general ISI channels using coupled codes with joint iterative decoding. To support this, one can also see that for the  $(5, 10, 44)$  ensemble of the same rate as the  $(3, 6, 22)$  one, the threshold  $\sigma^{\text{BP}}(5, 10, 44) \approx 0.834 \pm 0.001$  dB (which is also roughly  $\bar{\sigma}^{\text{MAP}}(5, 10)$ ) gets very close to the signal-to-noise ratio (SNR) corresponding to the SIR ( $\sigma^{\text{SIR}} \approx 0.823 \pm 0.001$  dB using the numerical method in [67, 68]). Similar effects can also be observed for other SC codes based on regular LDPC ensembles considered in Table IV for dicode and PR2 channels with AWGN.

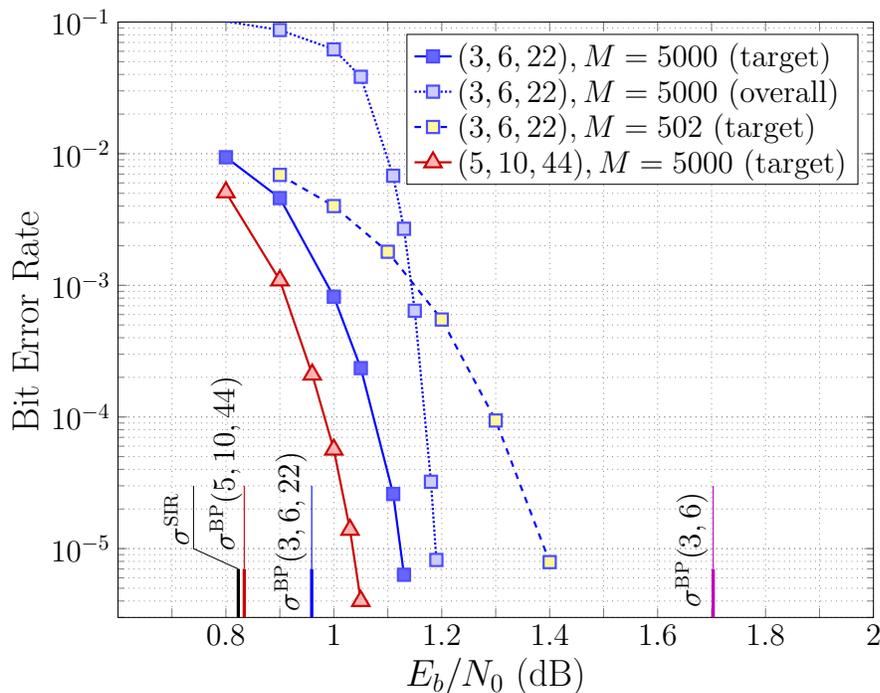


Fig. 26. BER and BP thresholds for the  $(3, 6)$ -regular,  $(3, 6, 22)$ -SC and  $(5, 10, 44)$ -SC LDPC codes over the AWGN dicode channel.

Also shown in Fig. 26 is the bit error rate (BER) versus SNR plot for the ensembles derived from the  $(d_l, d_r, L)$  circular ensembles of finite  $M = 502$  and  $M = 5000$ . For each simulation, we use  $\ell_{\text{outer}} = 20$  channel updates and between two such channel updates, we run  $\ell_{\text{inner}} = 5$  BP iterations on the code part alone. The curves labeled “target” give the BER for the bits at position  $I + 1$  (right after the known bits) in the coupled chain while the curve labeled “overall” is the average BER for all positions  $[I + 1, L]$  together. We expect that the “overall” BER will get closer to the “target” BER for large enough  $M$  and large enough number of iterations. From Fig. 26, one can also observe that the “overall” BER for  $(3, 6, 22)$  and  $M = 5000$  keeps getting “closer” to the “target” BER as SNR slightly increases. Those BER curves are far to the left of  $\epsilon^{\text{BP}}(3, 6)$  - the BP threshold for the underlying  $(3, 6)$ -regular ensemble.

## CHAPTER IV

## ON THE MAP THRESHOLD OF MULTIUSER SYSTEMS WITH ERASURES

## A. Introduction

If the factor graph representing an LDPC code has no cycles, then the BP decoder provides an optimal decoding solution whose complexity scales linearly with the block-length. On the other hand, the maximum *a posteriori* (MAP) decoding is globally optimal but its complexity is prohibitively large in many scenarios. Associated with each decoder is a noise threshold, below which the decoder achieves arbitrarily reliable communication as  $n \rightarrow \infty$ . In many cases, there is a gap between the BP and MAP thresholds [6]. Evaluating these two thresholds, for an iterative decoding system, is an important part of understanding the codes and decoding algorithms.

An interesting example is the threshold saturation phenomenon of spatially-coupled (SC) LDPC codes, whereby the BP threshold can be improved to the MAP threshold [20]. While determining the BP threshold is straightforward via density-evolution (DE) analysis, evaluating the MAP threshold directly is problematic due to complexity issues. Fortunately, for LDPC codes over the binary erasure channel (BEC), a fundamental relationship between the BP and MAP thresholds can be found using extrinsic information transfer (EXIT) curves. One can use this connection to upper bound the MAP threshold [77]. This bounding technique has now been used in various point-to-point communication problems [90, 78, 81] and also to evaluate the MAP threshold of turbo codes [91].

In this chapter, we use a similar analysis to evaluate the MAP thresholds of LDPC codes over two multiuser systems: the noisy Slepian-Wolf (SW) problem and the two-user multiple access channel (MAC). For more realistic noise models, one

can use generalized EXIT (GEXIT) analysis to compute upper bounds on the MAP thresholds [64, 65]. This chapter focuses, however, on erasure noise models. The simplicity of these toy models allows one to perform a thorough analysis and provide insights into the general case. For each problem, we derive the appropriate EXIT curve and use the natural area theorem to obtain an upper bound on the MAP threshold. We then provide a counting argument that allows numerical verification of the bound's tightness.

As mentioned earlier, one direct application of this analysis is verification of the threshold saturation phenomenon, the underlying mechanism behind the impressive performance of SC codes. While the BP thresholds of SC were recently observed to get close to the Shannon limits of these considered problems [64, 63], with the MAP threshold evaluated here, one can see that the BP thresholds of SC codes saturate not just to some value that can be close to the “capacity” but this value turns out to be exactly the MAP threshold of the underlying ensemble. Since the MAP thresholds can be shown to converge to the Shannon limit as the node degrees increase, it is not surprising that SC codes can achieve the entire capacity region of the corresponding problems.

## 1. Preliminaries

Besides the standard LDPC ensemble  $(\lambda(z), \rho(z))$ , for analysis, we also consider two-edge-type LDPC ensembles [93] whose degree distribution (d.d.) can be given by  $(L(z_1, z_2), R^{[1]}(z), R^{[2]}(z)) = (\sum_{i_1, i_2} L_{i_1, i_2} z_1^{i_1} z_2^{i_2}, \sum_i R_i^{[1]} z^i, \sum_i R_i^{[2]} z^i)$  where  $L_{i_1, i_2}$  gives the fraction of bit nodes with  $i_j$  outgoing edges of type  $j$  while  $R_i^{[j]}$  gives the fraction of check nodes with  $i$  edges of type  $j$  for  $j \in \{1, 2\}$ .

Throughout this chapter,  $\mathbf{X}^{[j]}$  is used to denote a vector of bits where  $[j]$  indicates user (or channel)  $j$  for  $j \in \{1, 2\}$ . Likewise,  $X_i^{[j]}$  represents the  $i$ -th bit and

$X^{[j]}$  is sometimes used if the bit index is not emphasized. For simplicity,  $\mathbf{X}_{\sim i}^{[j]}$  is used when the  $i$ -th bit is omitted from the vector  $\mathbf{X}^{[j]}$ . Erasures are denoted by  $?$ .

## B. Slepian-Wolf Problem with Erasures

### 1. Channel Model

Two correlated discrete memoryless sources are encoded by two independent linear encoding functions of identical design rate  $r$ . These encoders map  $k$  input symbols ( $\mathbf{U}^{[1]}$  and  $\mathbf{U}^{[2]}$ ) to  $n$  output symbols ( $\mathbf{X}^{[1]}$  and  $\mathbf{X}^{[2]}$ ) which are then transmitted through two independent channels. A central location receives  $(\mathbf{Y}^{[1]}, \mathbf{Y}^{[2]})$  and jointly decodes them to  $(\mathbf{U}^{[1]}, \mathbf{U}^{[2]})$ . In the model we consider, the two channels are BECs with erasure rate  $\epsilon^{[1]}$  and  $\epsilon^{[2]}$ , respectively. That is

$$Y_i^{[j]} = \begin{cases} X_i^{[j]} & \text{with probability } 1 - \epsilon_i^{[j]}, \\ ? & \text{with probability } \epsilon_i^{[j]}, \end{cases}$$

for  $j \in \{1, 2\}$  and  $i \in \{1, 2, \dots, n\}$  where we assume  $\epsilon_i^{[1]} = \epsilon^{[1]}$  and  $\epsilon_i^{[2]} = \epsilon^{[2]}$  for all  $i$ . We also consider an erasure correlation model between the two sources. More specifically, let  $Z$  be a Bernoulli- $p$  random variable and  $X$  and  $X'$  be i.i.d. Bernoulli- $\frac{1}{2}$  random variables. The sources  $U_1$  and  $U_2$  are defined by

$$(U_1, U_2) = \begin{cases} (X, X') & \text{if } Z = 0, \\ (X, X) & \text{if } Z = 1. \end{cases}$$

This gives  $H(U_1|U_2) = H(U_2|U_1) = 1-p$  and  $H(U_1, U_2) = 2-p$ . The decoder is assumed to have access to the side information  $Z$ . In [64], The Slepian-Wolf region is found to

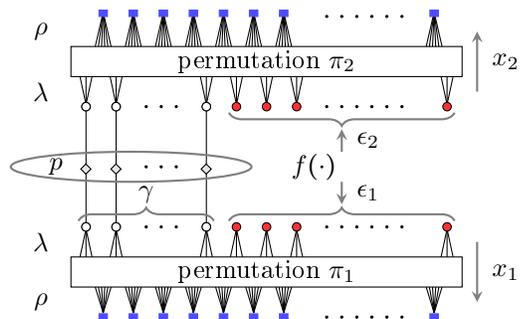


Fig. 27. Tanner graph for an LDPC code and the SWE

be

$$\begin{aligned}\epsilon^{[1]} &\leq 1 - H(U_2|U_1)r = 1 - (1-p)r, \\ \epsilon^{[2]} &\leq 1 - H(U_1|U_2)r = 1 - (1-p)r, \\ \epsilon^{[1]} + \epsilon^{[2]} &\leq 2 - H(U_1, U_2)r = 2 - (2-p)r.\end{aligned}$$

Assume that the sequences  $\mathbf{U}^{[1]}$  and  $\mathbf{U}^{[2]}$  are encoded by LDPC codes with the same d.d.  $(\lambda, \rho)$  with a punctured systematic encoder. The fraction of punctured systematic bits is  $\gamma = 1 - \frac{L'(1)}{R'(1)}$  (see more discussion in [64]). After puncturing, the two codes have rate  $r = \frac{\gamma}{1-\gamma}$ . The Tanner graph for an LDPC code and the SW problem with erasures (SWE) is given by Fig. 27.

If the joint BP decoder is used, one has the following fixed point (FP) equation based on DE (see [64])

$$x_1 = [\gamma f(L(y(x_2))) + (1-\gamma)\epsilon^{[1]}] \lambda(y(x_1)), \quad (4.1)$$

$$x_2 = [\gamma f(L(y(x_1))) + (1-\gamma)\epsilon^{[2]}] \lambda(y(x_2)), \quad (4.2)$$

where  $f(t) \triangleq (1-p) + pt$  and  $y(t) \triangleq 1 - \rho(1-t)$ . Here,  $x_1$  (resp.  $x_2$ ) is the average erasure rate of messages from bit nodes to check nodes corresponding to source 1 (resp. 2) in the limit of infinite block length and infinite number of BP iterations.

From this, one can write

$$\begin{aligned}\epsilon^{[1]}(x_1, x_2) &= \frac{1}{1-\gamma} \left[ \frac{x_1}{\lambda(y(x_1))} - \gamma f(L(y(x_2))) \right], \\ \epsilon^{[2]}(x_1, x_2) &= \frac{1}{1-\gamma} \left[ \frac{x_2}{\lambda(y(x_2))} - \gamma f(L(y(x_1))) \right].\end{aligned}$$

Let us express  $x_1(x)$  and  $x_2(x)$  according to a common parameter  $x$ , say  $x = x_1$ , and consider a smooth curve  $\mathfrak{C}$  from  $(x_1(1), x_2(1)) = (1, 1)$  and decreases in both arguments. This curve  $\mathfrak{C}$  can be characterized by a single parameter  $\epsilon \in [0, 1]$ , say  $\epsilon = \frac{\epsilon^{[1]} + \epsilon^{[2]}}{2}$ , and it can be seen that  $\epsilon = 1$  corresponds to  $(\epsilon^{[1]}, \epsilon^{[2]}) = (1, 1)$ . Many steps in analysis, presented in Section B, are based on this assumption for the curve  $\mathfrak{C}$ .

## 2. EXIT Functions

**Definition 13.** Consider a sequence of LDPC( $n, \lambda, \rho$ ) ensembles. For each  $n$ , pick  $\mathcal{C}_n$  uniformly at random from LDPC( $n, \lambda, \rho$ ) and let  $\mathbf{X}^{[1]}, \mathbf{X}^{[2]}$  be chosen uniformly from  $\mathcal{C}_n$ . Let  $Y_1^n$  be the received sequence after transmission over the SWE with erasure rate pair  $(\epsilon^{[1]}, \epsilon^{[2]})$  characterized by a common parameter  $\epsilon$ . The (MAP-)EXIT function associated with  $\mathcal{C}_n$  is defined by

$$h_{\mathcal{C}_n}(\epsilon) \triangleq \frac{1}{n} \cdot \frac{d}{d\epsilon} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}(\epsilon^{[1]}(\epsilon)), \mathbf{Y}^{[2]}(\epsilon^{[2]}(\epsilon))).$$

and the (asymptotic) EXIT function is given by

$$h(\epsilon) \triangleq \limsup_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n} [h_{\mathcal{C}_n}(\epsilon)].$$

**Theorem 12.** The above definition of the EXIT function naturally gives an area theorem as follows

$$\int_0^{\epsilon^*} h_{\mathcal{C}_n}(\epsilon) d\epsilon = \frac{1}{n} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}(\epsilon^{[1]}(\epsilon^*)), \mathbf{Y}^{[2]}(\epsilon^{[2]}(\epsilon^*))).$$

As a consequence, if  $\epsilon^* = 1$  then this gives the area  $\int_0^1 h_{C_n}(\epsilon) d\epsilon = H(U^{[1]}, U^{[2]})r = (2-p)r$  given uniform priors on the codeword sets.

**Lemma 20.** *For the SWE, the EXIT function becomes*

$$h_{C_n}(\epsilon) = \frac{1}{n} \sum_{i=1}^n \left( H(X_i^{[1]} | \mathbf{Y}_{\sim i}^{[1]}, \mathbf{Y}^{[2]}) \frac{d\epsilon^{[1]}}{d\epsilon} + H(X_i^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}_{\sim i}^{[2]}) \frac{d\epsilon^{[2]}}{d\epsilon} \right) \quad (4.3)$$

where  $\mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}, \mathbf{Y}_{\sim i}^{[1]}, \mathbf{Y}_{\sim i}^{[2]}, \epsilon^{[1]}, \epsilon^{[2]}$  can all be written as functions of  $\epsilon$ .

*Proof.* Since  $H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]})$  depends on  $\epsilon_i^{[1]}$  and  $\epsilon_i^{[2]}$ , one has

$$\begin{aligned} \frac{d}{d\epsilon} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) &= \sum_{i=1}^n \left( \frac{\partial}{\partial \epsilon_i^{[1]}} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) \frac{d\epsilon_i^{[1]}}{d\epsilon} \right. \\ &\quad \left. + \frac{\partial}{\partial \epsilon_i^{[2]}} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) \frac{d\epsilon_i^{[2]}}{d\epsilon} \right). \end{aligned} \quad (4.4)$$

Using the entropy chain rule, one can write

$$\begin{aligned} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) &= H(\mathbf{X}^{[1]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) + H(\mathbf{X}^{[2]} | \mathbf{X}^{[1]}, \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) \\ &= H(X_i^{[1]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) + H(\mathbf{X}_{\sim i}^{[1]} | X_i^{[1]}, \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) \\ &\quad + H(\mathbf{X}^{[2]} | \mathbf{X}^{[1]}, \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) \\ &= \epsilon_i^{[1]} H(X_i^{[1]} | \mathbf{Y}_{\sim i}^{[1]}, \mathbf{Y}^{[2]}) + H(\mathbf{X}_{\sim i}^{[1]} | X_i^{[1]}, \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) \\ &\quad + H(\mathbf{X}^{[2]} | \mathbf{X}^{[1]}, \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}). \end{aligned} \quad (4.5)$$

Now, one obtains

$$\frac{\partial}{\partial \epsilon_i^{[1]}} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) = H(X_i^{[1]} | \mathbf{Y}_{\sim i}^{[1]}, \mathbf{Y}^{[2]}) \quad (4.6)$$

since the second and third summands on the RHS of (4.5) do not depend on  $\epsilon_i^{[1]}$ .

Similarly, it can be shown that

$$\frac{\partial}{\partial \epsilon_i^{[2]}} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}^{[2]}) = H(X_i^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}_{\sim i}^{[2]}) \quad (4.7)$$

and the lemma follows directly from (4.4), (4.6) and (4.7).  $\square$

When the BP estimator is used instead of the MAP estimator, one also has the BP-EXIT function. The asymptotic BP-EXIT function can be obtained by taking the average BP-EXIT function over all the codes  $\mathcal{C}_n$  and taking  $n \rightarrow \infty$  and followed by the number of BP iterations  $\ell \rightarrow \infty$ . Using a concentration theorem and the fact that, for a fixed number of iterations, the computation graph for a specific bit becomes tree-like as  $n \rightarrow \infty$ , one can compute the asymptotic BP-EXIT function as follows.

**Lemma 21.** *The (asymptotic) BP-EXIT function is given by*

$$h^{\text{BP}}(\epsilon) = L(y(x_1)) \frac{d\epsilon^{[1]}}{d\epsilon} + L(y(x_2)) \frac{d\epsilon^{[2]}}{d\epsilon}$$

where  $(x_1, x_2)$  is the FP pair at channel erasure rate pair  $(\epsilon^{[1]}(\epsilon), \epsilon^{[2]}(\epsilon))$ .

### 3. MAP Threshold

#### a. Upper Bound on the MAP Threshold

Generally speaking, the MAP threshold (along the curve  $\mathfrak{C}$ ) can be defined as the supremum of all parameters  $\epsilon$  such that  $h(\epsilon) = 0$ . Likewise, one can define the BP threshold. By the optimality of the MAP decoder, one can invoke the data processing inequality [41] and have the following lemma.

**Lemma 22.** *For the SWE, one has  $0 \leq h(\epsilon) \leq h^{\text{BP}}(\epsilon) \leq 2$ .*

*Proof.* By the data processing inequality [41], the entropies  $H(X_i^{[1]} | \mathbf{Y}_{\sim i}^{[1]}, \mathbf{Y}^{[2]})$  and  $H(X_i^{[2]} | \mathbf{Y}^{[1]}, \mathbf{Y}_{\sim i}^{[2]})$  in (4.3) reduce if one replaces the optimal MAP estimator with the BP estimator. They are also upper bounded by 1. The lemma then follows immediately.  $\square$

**Remark 21.** *With the above analysis, one can use a similar approach to [77] to obtain an upper bound on the MAP threshold. More specifically, by finding the largest  $\bar{\epsilon}^{\text{MAP}}$  such that  $\int_{\bar{\epsilon}^{\text{MAP}}}^1 h^{\text{BP}}(\epsilon) d\epsilon = H(U^{[1]}, U^{[2]})r$ , one has  $\epsilon^{\text{MAP}} \leq \bar{\epsilon}^{\text{MAP}}$  which follows from  $\int_{\bar{\epsilon}^{\text{MAP}}}^1 h(\epsilon) d\epsilon \leq \int_{\bar{\epsilon}^{\text{MAP}}}^1 h^{\text{BP}}(\epsilon) d\epsilon = H(U^{[1]}, U^{[2]})r \leq \int_{\epsilon^{\text{MAP}}}^1 h(\epsilon) d\epsilon$ .*

To compute the area under the BP-EXIT curve, it is more convenient to consider the extended BP (EBP) EXIT curve that extends the BP-EXIT by also including the unstable FPs.

**Definition 14.** *The EBP-EXIT curve for the SWE is defined by*

$$\left( \epsilon(x), L(y(x_1(x))) \frac{d\epsilon^{[1]}}{d\epsilon}(x) + L(y(x_2(x))) \frac{d\epsilon^{[2]}}{d\epsilon}(x) \right)$$

for  $x \in [0, 1]$  where the second coordinate is called the EBP-EXIT function  $h^{\text{EBP}}(x)$ .

The area under the BP-EXIT curve can be computed with the help of a “trial entropy” as follows.

**Lemma 23.** *Let  $P(x) \triangleq \int_0^x h^{\text{EBP}}(t) d\epsilon(t)$  denote the “trial entropy”. Then, we have*

$$\begin{aligned} P(x) = & \frac{1}{1-\gamma} \left\{ L(y(x_1)) [(1-\gamma)\epsilon^{[1]} + (1-p)\gamma] + L(y(x_2)) [(1-\gamma)\epsilon^{[2]} + (1-p)\gamma] \right. \\ & + \gamma p L(y(x_1)) L(y(x_2)) - \frac{L'(1)}{R'(1)} \left[ 2 - \left( R(1-x_1) + R(1-x_2) \right. \right. \\ & \left. \left. + x_1 R'(1-x_1) + x_2 R'(1-x_2) \right) \right] \left. \right\} \end{aligned}$$

where  $\epsilon^{[j]}$  and  $x_j$  are also functions of  $x$  for  $j \in \{1, 2\}$ .

*Proof.* We start by realizing that

$$\begin{aligned}
P(x) &= \int_0^x \left( L(y(x_1(t))) \frac{d\epsilon^{[1]}(t)}{d\epsilon(t)} + L(y(x_1(t))) \frac{d\epsilon^{[1]}(t)}{d\epsilon(t)} \right) d\epsilon(t) \\
&= \int_0^x L(y(x_1(t))) d\epsilon^{[1]}(t) + \int_0^x L(y(x_2(t))) d\epsilon^{[2]}(t) \\
&= \frac{1}{1-\gamma} \int_0^x L(y(x_1(t))) d\left( \frac{x_1(t)}{\lambda(y(x_1(t)))} \right) + \frac{1}{1-\gamma} \int_0^x L(y(x_2(t))) d\left( \frac{x_2(t)}{\lambda(y(x_2(t)))} \right) \\
&\quad - \frac{p\gamma}{1-\gamma} \left[ \int_0^x L(y(x_1(t))) dL(y(x_2(t))) + \int_0^x L(y(x_2(t))) dL(y(x_1(t))) \right] \quad (4.8) \\
&= \frac{1}{1-\gamma} \left\{ L(y(x_1)) \frac{x_1}{\lambda(y(x_1))} - \frac{L'(1)}{R'(1)} [1 - R(1-x_1) - x_1 R'(1-x_1)] \right\} \\
&\quad + \frac{1}{1-\gamma} \left\{ L(y(x_2)) \frac{x_2}{\lambda(y(x_2))} - \frac{L'(1)}{R'(1)} [1 - R(1-x_2) - x_2 R'(1-x_2)] \right\} \\
&\quad - \frac{p\gamma}{1-\gamma} L(y(x_1)) L(y(x_2)) \quad (4.9)
\end{aligned}$$

where (4.9) follows from applying the trial entropy formula of the BEC in [6, p. 124] for the first and second summands of (4.8) and using integration by parts for the third summand. Here, in (4.9), we write  $x_1$  and  $x_2$  as shorthand notations for  $x_1(x)$  and  $x_2(x)$ , respectively.

Next, from (4.1) and (4.2), one can substitute  $\frac{x_1}{\lambda(y(x_1))}$  by  $(1-\gamma)\epsilon^{[1]}(x_1, x_2) + \gamma((1-p) + pL(y(x_2)))$  and  $\frac{x_2}{\lambda(y(x_2))}$  by  $(1-\gamma)\epsilon^{[2]}(x_1, x_2) + \gamma((1-p) + pL(y(x_1)))$ , respectively, in (4.9) and obtain the lemma after a few simplifications.  $\square$

**Corollary 5.** *The EBP-EXIT curve also satisfies an area property that says*

$$\int_0^1 h^{\text{EBP}}(t) dt = (2-p)r = H(U^{[1]}, U^{[2]})r.$$

*Proof.* It follows immediately from Lemma 23 that

$$\int_0^1 h^{\text{EBP}}(t) dt = P(1) = \frac{1}{1-\gamma} \left\{ 2 \left( 1 - \frac{L'(1)}{R'(1)} \right) - \gamma p \right\}.$$

Then, one obtains the result by realizing that  $1 - \frac{L'(1)}{R'(1)} = \gamma$  and  $\frac{\gamma}{1-\gamma} = r$ .  $\square$

Using Corollary 5, the bounding technique in Remark 21 can also be discussed in the following way.

**Corollary 6.** *By finding a positive root  $x^{\text{MAP}}$  of  $P(x) = 0$  that gives the largest  $\epsilon(x^{\text{MAP}})$ , one can obtain the upper bound on the MAP threshold  $\epsilon^{\text{MAP}} \leq \epsilon(x^{\text{MAP}}) \triangleq \bar{\epsilon}^{\text{MAP}}$ .*

If one considers regular LDPC ensembles of a fixed rate, using the property  $P(x^{\text{MAP}}) = 0$ , the following lemma shows that, for sufficiently high degrees,  $\bar{\epsilon}^{\text{MAP}}$  approaches the Shannon limit.

**Lemma 24.** *Consider the  $(d_l, d_r)$ -regular ensembles and let  $d_l \rightarrow \infty$  so that the design rate  $r = 1 - \frac{d_l}{d_r}$  is fixed. Then, we have  $\bar{\epsilon}^{\text{MAP}}(d_l, d_r) \rightarrow \epsilon^{\text{Sh}}(r)$  where  $\epsilon^{\text{Sh}}(r)$  corresponds to the boundary of the SW region along the considered curve  $\mathfrak{C}$ .*

*Proof.* For simplicity, we denote  $x_j(x^{\text{MAP}}(d_l, d_r))$  as  $x_j^{\text{MAP}}(d_l, d_r)$  for  $j \in \{1, 2\}$ . Let  $\alpha_j$  be the limit of  $x_j^{\text{MAP}}(d_l, d_r)$  as  $d_l, d_r \rightarrow \infty$  with  $r$  constant. Similarly to the proof of Theorem 10,  $L(y(x_j^{\text{MAP}}(d_l, d_r)))$ ,  $\lambda(y(x_j^{\text{MAP}}(d_l, d_r)))$ ,  $R(1 - x_j^{\text{MAP}}(d_l, d_r))$  and  $R'(1 - x_j^{\text{MAP}}(d_l, d_r))$  all converge to 1 if  $\alpha_j \neq 0$  and converge to 0 if  $\alpha_j = 0$ . Then, based on the observation that  $P(x^{\text{MAP}}(d_l, d_r)) = 0$ , one either obtains  $(\epsilon^{[1] + \epsilon^{[2]}}) \rightarrow 2 - (2-p)r$  for the case of  $\alpha_1 \neq 0$  and  $\alpha_2 \neq 0$ ,  $\epsilon^{[1]} \rightarrow 1 - (1-p)r$  for the case of  $\alpha_1 \neq 0$  and  $\alpha_2 = 0$ , or  $\epsilon^{[2]} \rightarrow 1 - (1-p)r$  for the case of  $\alpha_1 = 0$  and  $\alpha_2 \neq 0$  which depends on the curve  $\mathfrak{C}$  being considered.  $\square$

#### b. Tightness of the Upper Bound

**Lemma 25.** *Assume the joint BP decoder is run until it reaches a FP. Remove all known bits and merge two aligned bits that have the same value into a larger bit nodes and obtain a residual graph. The residual graph can be seen as a two edge type LDPC*

ensemble and its expected d.d. (normalized with respect to the original graph) is<sup>1</sup>

$$\begin{aligned}\tilde{L}_\epsilon(z_1, z_2) &= L(y(x_1)z_1)(\epsilon^{[1]}(1-\gamma) + (1-p)\gamma) + L(y(x_2)z_2)(\epsilon^{[2]}(1-\gamma) + (1-p)\gamma) \\ &\quad + \gamma p L(y(x_1)z_1)L(y(x_2)z_2),\end{aligned}$$

$$\tilde{R}_\epsilon^{[1]}(z) = R(1-x_1+zx_1) - R(1-x_1) - zx_1R'(1-x_1),$$

$$\tilde{R}_\epsilon^{[2]}(z) = R(1-x_2+zx_2) - R(1-x_2) - zx_2R'(1-x_2),$$

where  $(x_1, x_2)$  is the FP at erasure rate pair  $(\epsilon^{[1]}(\epsilon), \epsilon^{[2]}(\epsilon))$  and  $y_1 \triangleq y(x_1)$  and  $y_2 \triangleq y(x_2)$ .

*Proof.* The check node d.d.  $\tilde{R}_\epsilon^{[j]}(z)$  can be shown in the same way as in the proof of Lemma 16. For the d.d. of bit nodes, we use the following counting argument.

One first considers unpunctured bit nodes whose fraction is  $1-\gamma$ . For these unpunctured bit nodes to remain in the residual graph, the messages from the channel as well as from the corresponding check nodes must be erasures. This happens with probability  $\epsilon^{[1]}L(y(x_1))$  and  $\epsilon^{[2]}L(y(x_2))$  for the source 1 and source 2, respectively.

Next, one considers punctured bit nodes whose fraction is  $\gamma$ . Among punctured bit nodes, there are ones connected by correlation nodes, with probability  $p$ , and ones not connected, with probability  $1-p$ . For the punctured bit nodes not connected by correlation nodes to remain in the residual graph, the messages from the corresponding check nodes must be erasures. This happens with probability  $(1-p)L(y(x_1))$  for bits corresponding to the source 1 and  $(1-p)L(y(x_2))$  for bits corresponding to the source 2. Meanwhile, every two punctured bit nodes which are connected by a correlation node are merged into a larger bit nodes because they must have the same value. For each larger bit node to remain in the residual graph, the messages from

---

<sup>1</sup>The two edge type standard d.d. is  $\left(\frac{\tilde{L}_\epsilon(z_1, z_2)}{\tilde{L}_\epsilon(1, 1)}, \frac{\tilde{R}_\epsilon^{[1]}(z)}{\tilde{R}_\epsilon^{[1]}(1)}, \frac{\tilde{R}_\epsilon^{[2]}(z)}{\tilde{R}_\epsilon^{[2]}(1)}\right)$ .

check nodes in both sources must be erasures. This event happens with probability  $pL(y(x_1))L(y(x_2))$ .

Therefore, we can write the d.d. of the bit nodes in the residual graph as

$$\begin{aligned} \tilde{L}_\epsilon(z_1, z_2) &= (1 - \gamma) \left( \epsilon^{[1]} L(y(x_1)z_1) + \epsilon^{[2]} L(y(x_2)z_2) \right) \\ &\quad + \gamma \left( (1 - p) [L(y(x_1)z_1) + L(y(x_2)z_2)] + pL(y(x_1)z_1)L(y(x_2)z_2) \right), \end{aligned}$$

because each edge of type  $j$  is associated with a erasure probability  $y(x_j)$  from check nodes to bit nodes for  $j \in \{1, 2\}$ . The result then follows immediately.  $\square$

**Theorem 13.** *At  $\epsilon = \bar{\epsilon}^{\text{MAP}}$ , the design rate of the residual graph equals zero.*

*Proof.* Let the number of bit nodes in the original graph be  $n$ . Then, then number of bit nodes in the residual graph is  $n\tilde{L}_\epsilon(1, 1)$ . Also, the number of check nodes with outgoing edges of type  $j$ , for  $j \in \{1, 2\}$ , in the original graph is  $n(1 - r_j) = n\frac{L'(1)}{R'(1)}$ . Thus, the number of check nodes with outgoing edges of type  $j$  in the residual graph is  $n\frac{L'(1)}{R'(1)}\tilde{R}_\epsilon^{[j]}(1)$ . Therefore, the total number of check nodes in the residual graph is  $n\frac{L'(1)}{R'(1)}\left(\tilde{R}_\epsilon^{[1]}(1) + \tilde{R}_\epsilon^{[2]}(1)\right)$ .

Consequently, the design rate of the residual graph is

$$\tilde{r}_\epsilon = 1 - \frac{L'(1)}{R'(1)} \left( \tilde{R}_\epsilon^{[1]}(1) + \tilde{R}_\epsilon^{[2]}(1) \right) / \tilde{L}_\epsilon(1, 1). \quad (4.10)$$

Finally, one can see from (4.10), Lemma 23 and Corollary 6 that

$$\tilde{r}_{\bar{\epsilon}^{\text{MAP}}} = (1 - \gamma)P(x^{\text{MAP}}) / \tilde{L}_{\bar{\epsilon}^{\text{MAP}}}(1, 1) = 0.$$

and obtain the theorem.  $\square$

**Remark 22.** *To show the tightness of the MAP upper bound, it remains to be shown that the actual rate of the residual graph is zero because if this is true, the MAP decoder*

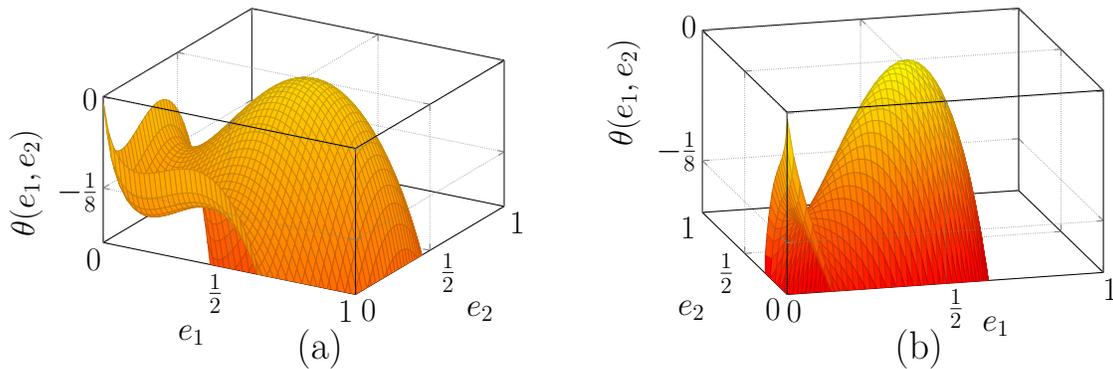


Fig. 28.  $\theta(e_1, e_2)$  of the residual graph for: (a) the SWE in Remark 22 and (b) the EMAC in Remark 23.

should decode perfectly for all  $\epsilon < \bar{\epsilon}^{\text{MAP}}$ . One can use the test in [93] to numerically verify this fact. For simplicity, we focus on the curves  $\mathfrak{C}$  that extends linearly from  $(1, 1)$  according to  $1 - \epsilon^{[1]} = A(1 - \epsilon^{[2]})$  for some  $A > 0$  where, e.g.,  $A = 1$  represents the symmetric channel condition. In Fig. 28 (a), the function  $\theta(e_1, e_2)$  (see [93, p. 11] for definition) for the residual graph d.d. is plotted for the case  $A = \frac{3}{2}$  and one can see that the maximum of  $\theta(e_1, e_2)$  over the unit square is zero. Thus, the actual rate of the residual graph equals its design rate, which is zero, with high probability as  $n \rightarrow \infty$ . This implies that the upper bound is tight and  $\epsilon^{\text{MAP}} = \bar{\epsilon}^{\text{MAP}}$ .

## C. Multiple Access Channel with Erasures

### 1. Channel Model

In this section, we consider the two-user MAC channel with erasure noise (EMAC) discussed in [63] and evaluate the MAP threshold when the two users transmit LDPC codes. For the inputs  $X_i^{[1]}, X_i^{[2]} \in \{\pm 1\}$ , let the output be given by

$$Y_i = \begin{cases} X_i^{[1]} + X_i^{[2]} \triangleq Z_i & \text{with probability } 1 - \epsilon_i, \\ ? & \text{with probability } \epsilon_i, \end{cases}$$

where erasure rate  $\epsilon_i = \epsilon$  for all  $i \in \{1, 2, \dots, n\}$ . The achievable rate region for the design rate pair  $(r_1, r_2)$  of the two users is characterized by

$$\begin{aligned} r_1 &\leq I(X^{[1]}; Y|X^{[2]}) = 1 - \epsilon, \\ r_2 &\leq I(X^{[2]}; Y|X^{[1]}) = 1 - \epsilon, \\ r_1 + r_2 &\leq I(X^{[1]}, X^{[2]}; Y) = \frac{3}{2}(1 - \epsilon), \end{aligned}$$

and therefore the Shannon limit, i.e., the supremum of all erasure rates  $\epsilon$  that allows reliable communication for both users, is  $\epsilon^{\text{Sh}}(r_1, r_2) = \min\{1 - r_1, 1 - r_2, 1 - \frac{2}{3}(r_1 + r_2)\}$ .

We consider the input sequences  $\mathbf{X}^{[1]}$  and  $\mathbf{X}^{[2]}$  to be chosen uniformly at random from LDPC( $n, \lambda^{[1]}, \rho^{[1]}$ ) and LDPC( $n, \lambda^{[2]}, \rho^{[2]}$ ) ensembles, respectively. If one uses the joint BP decoder operating on the Tanner graph in Fig. 29, the FP equation based on DE is given by

$$\begin{aligned} x_1 &= (\epsilon + (1 - \epsilon)L^{[2]}(y_2(x_2)/2)\lambda^{[1]}(y_1(x_1))), \\ x_2 &= (\epsilon + (1 - \epsilon)L^{[1]}(y_1(x_1)/2)\lambda^{[2]}(y_2(x_2))), \end{aligned}$$

where  $x_j$  and  $y_j(x_j) \triangleq 1 - \rho^{[j]}(1 - x_j)$  are the expected erasure rate of the messages from bit nodes to check nodes and check nodes to bit nodes, respectively, corresponding to user  $j$  (in the limit of infinite block length and infinite number of BP iterations). One can express the FP pair  $(x_1, x_2)$  as a function of a common parameter  $x$ , say  $x = x_1$ . Thus, one can write  $\epsilon(x)$ ,  $x_1(x)$  and  $x_2(x)$  to emphasize the dependence on  $x$  and note that  $\epsilon(1) = 1$ . With some abuse of notation, we write  $y_j(x)$  as a shorthand notation of  $y_j(x_j(x))$  for  $j \in \{1, 2\}$ .

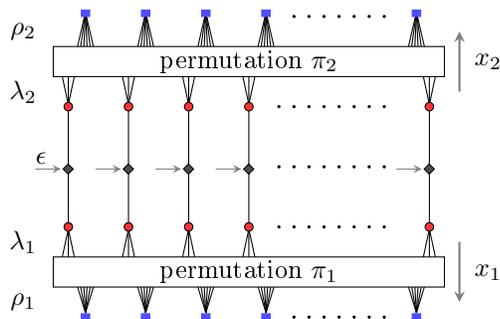


Fig. 29. Tanner graph of the joint decoder for the EMAC

## 2. EXIT Functions

**Definition 15.** Consider sequences of LDPC( $n, \lambda^{[1]}, \rho^{[1]}$ ) and LDPC( $n, \lambda^{[2]}, \rho^{[2]}$ ) ensembles. For each  $n$ , pick  $\mathcal{C}_n^{[j]}$  uniformly at random from LDPC( $n, \lambda_j, \rho_j$ ). Let  $\mathbf{X}^{[j]}$  be chosen uniformly from  $\mathcal{C}_n^{[j]}$  for  $j \in \{1, 2\}$  and  $\mathbf{Y}_1^n$  be the received sequence at the output of the EMAC with erasure probability  $\epsilon$ . The (MAP-)EXIT function associated with  $\mathcal{C}_n^{[1]}$  and  $\mathcal{C}_n^{[2]}$  is defined by

$$h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) = \frac{1}{n} \cdot \frac{d}{d\epsilon} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}(\epsilon)). \quad (4.11)$$

and the (asymptotic) EXIT function is given by

$$h(\epsilon) = \limsup_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}} \left[ h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) \right].$$

**Theorem 14.** The definition of the EXIT function leads to an area theorem that says

$$\int_0^{\epsilon^*} h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) d\epsilon = \frac{1}{n} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}(\epsilon^*)).$$

Consequently, if  $\epsilon^* = 1$  then one has  $\int_0^1 h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) d\epsilon = r_1 + r_2$  given a uniform prior on the set of the codewords.

**Lemma 26.** *The EXIT function for the EMAC is*

$$h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) = \frac{1}{n} \sum_{i=1}^n \left( H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}(\epsilon)) - \frac{1}{2} H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}(\epsilon), Z_i = 0) \right). \quad (4.12)$$

*Proof.* First, we write

$$H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}) = H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}) + H(\mathbf{X}_{\sim i}^{[1]}, \mathbf{X}_{\sim i}^{[2]} | \mathbf{Y}, X_i^{[1]}, X_i^{[2]})$$

where the second term of the RHS does not depend on  $\epsilon_i$ .

Thus, this gives

$$\begin{aligned} \frac{d}{d\epsilon} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}) &= \sum_{i=1}^n \frac{\partial}{\partial \epsilon_i} H(\mathbf{X}^{[1]}, \mathbf{X}^{[2]} | \mathbf{Y}) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \epsilon_i} H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}). \end{aligned} \quad (4.13)$$

Now, one has

$$\begin{aligned} H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}) &= \epsilon_i H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}) + (1 - \epsilon_i) H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}, Z_i) \\ &= \epsilon_i H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}) + \frac{1 - \epsilon_i}{2} H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}, Z_i = 0) \end{aligned} \quad (4.14)$$

since  $H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}, Z_i \neq 0) = 0$ .

Finally, combining (4.13), (4.14), and (4.11) gives the result.  $\square$

Replacing the MAP estimator with the BP estimator, one has the BP-EXIT function for this problem as follows.

**Definition 16.** *The BP-EXIT function after iteration  $\ell$  for the EMAC is given by*

$$h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}^{\text{BP}, \ell}(\epsilon) = \frac{1}{n} \sum_{i=1}^n \left( H(X_i^{[1]}, X_i^{[2]} | \mathcal{E}_i^{\text{BP}, \ell}(\mathbf{Y}_{\sim i}(\epsilon))) - \frac{1}{2} H(X_i^{[1]}, X_i^{[2]} | \mathcal{E}_i^{\text{BP}, \ell}(\mathbf{Y}_{\sim i}(\epsilon)), Z_i = 0) \right)$$

where  $\mathcal{E}_i^{\text{BP}, \ell}(\mathbf{Y}_{\sim i})$  is the extrinsic BP estimate of  $(X_i^{[1]}, X_i^{[2]})$  after iteration  $\ell$  of the BP decoder.

The (asymptotic) BP-EXIT function for the EMAC is defined as

$$h^{\text{BP}}(\epsilon) \triangleq \lim_{\ell \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}} \left[ h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}^{\text{BP}, \ell}(\epsilon) \right]$$

where the expectation is taken over all codes  $\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}$ .

By invoking a concentration theorem and the tree-like property of the computation graph for a bit node (as  $n \rightarrow \infty$  and the number of iterations is fixed), the limits in Definition 16 exist and, furthermore, one can express the asymptotic BP-EXIT function conveniently as follows.

**Lemma 27.** *For BP estimator, the (asymptotic) BP-EXIT function is*

$$h^{\text{BP}}(\epsilon) = L^{[1]}(y_1(x_1)) + L^{[2]}(y_2(x_2)) - \frac{1}{2} L^{[1]}(y_1(x_1)) L^{[2]}(y_2(x_2)) \quad (4.15)$$

where  $(x_1, x_2)$  is the FP pair at channel erasure rate  $\epsilon$ .

### 3. MAP Threshold

#### a. Upper Bound on the MAP Threshold

Similarly to the case of SWE, the MAP threshold  $\epsilon^{\text{MAP}}$  is defined as the supremum of all channel parameters  $\epsilon$  such that  $h(\epsilon) = 0$ . The BP threshold  $\epsilon^{\text{BP}}$  can also be defined correspondingly. One can also show that the (MAP-)EXIT function lies below the BP-EXIT function.

**Lemma 28.** *For this EMAC problem, one has*

$$0 \leq h(\epsilon) \leq h^{\text{BP}}(\epsilon) \leq \frac{3}{2}.$$

*Proof.* From (4.14), one has

$$\frac{1}{2} H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}(\epsilon), Z_i = 0) = \frac{1}{1 - \epsilon} \left( H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}) - \epsilon H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}(\epsilon)) \right).$$

Combining this with (4.12) gives

$$\begin{aligned} h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) &= \frac{1}{(1-\epsilon)n} \sum_{i=1}^n \left( H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}_{\sim i}(\epsilon)) - H(X_i^{[1]}, X_i^{[2]} | \mathbf{Y}) \right) \\ &= \frac{1}{(1-\epsilon)n} \sum_{i=1}^n I(X_i^{[1]}, X_i^{[2]}; Y_i(\epsilon) | \mathbf{Y}_{\sim i}(\epsilon)) \end{aligned}$$

Now, we claim that  $I(X_i^{[1]}, X_i^{[2]}; Y_i | \mathbf{Y}_{\sim i}) \leq I(X_i^{[1]}, X_i^{[2]}; Y_i | \mathcal{E}_i^{\text{BP}, \ell}(\mathbf{Y}_{\sim i}))$  using a similar argument to the data processing inequality [41] as follows.

By the chain rule, we can expand mutual information in two different ways

$$\begin{aligned} I(X_i^{[1]}, X_i^{[2]}, \mathbf{Y}_{\sim i}; Y_i | \mathcal{E}_i^{\text{BP}, \ell}) &= I(X_i^{[1]}, X_i^{[2]}; Y_i | \mathcal{E}_i^{\text{BP}, \ell}) + I(\mathbf{Y}_{\sim i}; Y_i | X_i^{[1]}, X_i^{[2]}, \mathcal{E}_i^{\text{BP}, \ell}) \\ &= I(\mathbf{Y}_{\sim i}; Y_i | \mathcal{E}_i^{\text{BP}, \ell}) + I(X_i^{[1]}, X_i^{[2]}; Y_i | \mathbf{Y}_{\sim i}, \mathcal{E}_i^{\text{BP}, \ell}). \quad (4.16) \end{aligned}$$

Next,  $Y_i \rightarrow (X_i^{[1]}, X_i^{[2]}, \mathcal{E}_i^{\text{BP}, \ell}) \rightarrow \mathbf{Y}_{\sim i}$  forms a Markov chain since the channel is memoryless and therefore  $I(\mathbf{Y}_{\sim i}; Y_i | X_i^{[1]}, X_i^{[2]}, \mathcal{E}_i^{\text{BP}, \ell}) = 0$ . Furthermore, one also has  $I(X_i^{[1]}, X_i^{[2]}; Y_i | \mathbf{Y}_{\sim i}, \mathcal{E}_i^{\text{BP}, \ell}) = I(X_i^{[1]}, X_i^{[2]}; Y_i | \mathbf{Y}_{\sim i})$  and  $I(\mathbf{Y}_{\sim i}; Y_i | \mathcal{E}_i^{\text{BP}, \ell}) \geq 0$ . Using this and (4.16), one proves the claim.

As a consequence, it is clear that  $h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}(\epsilon) \leq h_{\mathcal{C}_n^{[1]}, \mathcal{C}_n^{[2]}}^{\text{BP}, \ell}(\epsilon)$  and then  $h(\epsilon) \leq h^{\text{BP}}(\epsilon)$ .

For simplicity of notations, let  $v_j \triangleq L^{[j]}(y_j(x_j))$ . One can rewrite (4.15) to obtain  $h^{\text{BP}}(\epsilon) = v_1(1-v_2) + v_2(1-v_1) + \frac{3}{2}v_1v_2 \geq 0$  because  $0 \leq v_1, v_2 \leq 1$ . For a similar reason, one also has  $h^{\text{BP}}(\epsilon) = \frac{3}{2} - \frac{1}{4}[(1-v_1)(3-v_2) + (1-v_2)(3-v_1)] \leq \frac{3}{2}$ .  $\square$

With the above analysis, one can invoke the bounding technique, i.e., finding the largest  $\bar{\epsilon}^{\text{MAP}}$  such that  $\int_{\bar{\epsilon}^{\text{MAP}}}^1 h^{\text{BP}}(\epsilon) d\epsilon = r_1 + r_2$  and obtaining  $\epsilon^{\text{MAP}} \leq \bar{\epsilon}^{\text{MAP}}$  as a result.

To conveniently compute the area under the BP-EXIT curve, we define the EBP-EXIT curve and compute the ‘‘trial entropy’’ as follows.

**Definition 17.** *The EBP-EXIT curve for the EMAC is defined by*

$$\left( \epsilon(x), L^{[1]}(y_1(x)) + L^{[2]}(y_2(x)) - \frac{1}{2}L^{[1]}(y_1(x))L^{[2]}(y_2(x)) \right)$$

for  $x \in [0, 1]$  where the second coordinate is called the EBP-EXIT function  $h^{\text{EBP}}(x)$ .

**Lemma 29.** Let  $P(x) \triangleq \int_0^x h^{\text{EBP}}(t) d\epsilon(t)$  denote the “trial entropy”. Then, one has

$$\begin{aligned} P(x) &= \epsilon(x) \left[ L^{[1]}(y_1(x)) + L^{[2]}(y_2(x)) \right] + \frac{1 - \epsilon(x)}{2} L^{[1]}(y_1(x)) L^{[2]}(y_2(x)) \\ &\quad - \frac{L^{[1]}'(1)}{R^{[1]}'(1)} \left[ 1 - R^{[1]}(1 - x_1(x)) - x_1(x) R^{[1]}'(1 - x_1(x)) \right] \\ &\quad - \frac{L^{[2]}'(1)}{R^{[2]}'(1)} \left[ 1 - R^{[2]}(1 - x_2(x)) - x_2(x) R^{[2]}'(1 - x_2(x)) \right]. \end{aligned}$$

*Proof.* One starts with

$$P(x) = h^{\text{EBP}}(x) \epsilon(x) - \int_0^x \epsilon(t) dh^{\text{EBP}}(t) \quad (4.17)$$

$$\begin{aligned} &= h^{\text{EBP}}(x) \epsilon(x) - \int_0^x \epsilon(t) \left( 1 - \frac{1}{2} L^{[2]}(y_2(x_2(t))) \right) dL^{[1]}(y_1(x_1(t))) \\ &\quad - \int_0^x \epsilon(t) \left( 1 - \frac{1}{2} L^{[1]}(y_1(x_1(t))) \right) dL^{[2]}(y_2(x_2(t))) \end{aligned} \quad (4.18)$$

$$\begin{aligned} &= h^{\text{EBP}}(x) \epsilon(x) - \int_0^x \left( \frac{x_1(t)}{\lambda^{[1]}(y_1(x_1(t)))} - \frac{L^{[2]}(y_2(x_2(t)))}{2} \right) dL^{[1]}(y_1(x_1(t))) \\ &\quad - \int_0^x \left( \frac{x_2(t)}{\lambda^{[2]}(y_2(x_2(t)))} - \frac{L^{[1]}(y_1(x_1(t)))}{2} \right) dL^{[2]}(y_2(x_2(t))) \end{aligned} \quad (4.19)$$

$$\begin{aligned} &= h^{\text{EBP}}(x) \epsilon(x) - \int_0^x \frac{x_1(t)}{\lambda^{[1]}(y_1(x_1(t)))} dL^{[1]}(y_1(x_1(t))) \\ &\quad - \int_0^x \frac{x_2(t)}{\lambda^{[2]}(y_2(x_2(t)))} dL^{[2]}(y_2(x_2(t))) + \frac{1}{2} L^{[1]}(y_1(x_1(x))) L^{[2]}(y_2(x_2(x))) \end{aligned} \quad (4.20)$$

$$\begin{aligned} &= h^{\text{EBP}}(x) \epsilon(x) + \left[ L^{[1]}'(1) x_1(x) \rho^{[1]}(1 - x_1(x)) - \frac{L^{[1]}'(1)}{R^{[1]}'(1)} (1 - R^{[1]}(1 - x_1(x))) \right] \\ &\quad + \left[ L^{[2]}'(1) x_2(x) \rho^{[2]}(1 - x_2(x)) - \frac{L^{[2]}'(1)}{R^{[2]}'(1)} (1 - R^{[2]}(1 - x_2(x))) \right] \\ &\quad + \frac{1}{2} L^{[1]}(y_1(x_1(x))) L^{[2]}(y_2(x_2(x))) \end{aligned} \quad (4.21)$$

where (4.17) uses integration by parts, (4.18) follows from the definition of  $h^{\text{EBP}}(\cdot)$ , (4.19) holds because of the DE-FP equation while (4.20) and (4.21) both use integrations by parts.

By substituting the full formula of  $h^{\text{EBP}}(x)$ , the lemma follows after a few simplifications.  $\square$

**Corollary 7.** *The EBP-EXIT curve for the EMAC satisfies an area property as follows*

$$\int_0^1 h^{\text{EBP}}(t) dt = r_1 + r_2.$$

*Proof.* It follows from Lemma 29 that

$$\int_0^1 h^{\text{EBP}}(t) dt = P(1) = 2 - \frac{L^{[1]'}(1)}{R_{[1]}'(1)} - \frac{L^{[2]'}(1)}{R_{[2]}'(1)} = r_1 + r_2.$$

$\square$

**Corollary 8.** *By finding a positive root  $x^{\text{MAP}}$  of  $P(x) = 0$  that gives the largest  $\epsilon(x^{\text{MAP}})$ , one can obtain an upper bound on the MAP threshold  $\epsilon^{\text{MAP}} \leq \bar{\epsilon}^{\text{MAP}}$  where  $\bar{\epsilon}^{\text{MAP}} = \epsilon(x^{\text{MAP}})$ .*

Again, by considering regular LDPC ensembles of a fixed rate and using the property  $P(x^{\text{MAP}}) = 0$ ,  $\bar{\epsilon}^{\text{MAP}}$  can be shown to approach the Shannon limit.

**Lemma 30.** *Consider the  $(d_l^{[1]}, d_r^{[1]})$  and  $(d_l^{[2]}, d_r^{[2]})$ -regular ensembles for user 1 and user 2, respectively, and let  $d_l^{[j]} \rightarrow \infty$  so that  $r_j = 1 - \frac{d_l^{[j]}}{d_r^{[j]}}$  is fixed (for  $j \in \{1, 2\}$ ). Then,  $\bar{\epsilon}^{\text{MAP}}(d_l^{[1]}, d_r^{[1]}, d_l^{[2]}, d_r^{[2]}) \rightarrow \epsilon^{\text{Sh}}(r_1, r_2)$ .*

*Proof.* For simplicity, we denote  $x_j(x^{\text{MAP}}(d_l^{[1]}, d_r^{[1]}, d_l^{[2]}, d_r^{[2]}))$  as  $x_j^{\text{MAP}}$  for  $j \in \{1, 2\}$ . Let  $\alpha_j$  denote the limit of  $x_j^{\text{MAP}}$  as  $d_l^{[1]}, d_r^{[1]}, d_l^{[2]}, d_r^{[2]} \rightarrow \infty$  with  $r_1, r_2$  constant. Similarly to the proof of Theorem 10,  $L^{[j]}(y_j(x_j^{\text{MAP}}))$ ,  $\lambda^{[j]}(y_j(x_j^{\text{MAP}}))$ ,  $R^{[j]}(1 - x_j^{\text{MAP}})$  and  $R^{[j]'}(1 - x_j^{\text{MAP}})$  all converge to 1 if  $\alpha_j \neq 0$  and converge to 0 if  $\alpha_j = 0$ . Then, based on the observation that  $P(x^{\text{MAP}}) = 0$ , one either has  $\bar{\epsilon}^{\text{MAP}} \rightarrow 1 - \frac{2}{3}(r_1 + r_2)$  for the case of  $\alpha_1 \neq 0$   $\alpha_2 \neq 0$ ,  $\bar{\epsilon}^{\text{MAP}} \rightarrow 1 - r_1$  for the case of  $\alpha_1 \neq 0$  and  $\alpha_2 = 0$ , or  $\bar{\epsilon}^{\text{MAP}} \rightarrow 1 - r_2$  for the

case of  $\alpha_1 = 0$  and  $\alpha_2 \neq 0$  where  $\bar{\epsilon}^{\text{MAP}} = \epsilon(x^{\text{MAP}}(d_l^{[1]}, d_r^{[1]}, d_l^{[2]}, d_r^{[2]}))$ . The exact cases to be considered depends on the corresponding rate pair  $(r_1, r_2)$ .  $\square$

### b. Tightness of the Upper Bound

We follow a similar approach as in the case of SWE.

**Lemma 31.** *Assume the joint BP decoder is run until it reaches a FP. Next, remove all the known bits and their adjacent edges and check nodes. Further, merge any pair of bit nodes at the same index that have the same value. The residual graph can be seen as a two-edge-type LDPC ensemble and its expected d.d. (normalized with respect to the original graph) is*

$$\begin{aligned}\tilde{L}_\epsilon(z_1, z_2) &= \epsilon L^{[1]}(y_1 z_1) + \epsilon L^{[2]}(y_2 z_2) + \frac{1-\epsilon}{2} L^{[1]}(y_1 z_1) L^{[2]}(y_2 z_2), \\ \tilde{R}_\epsilon^{[1]}(z) &= R^{[1]}(1 - x_1 + z x_1) - R^{[1]}(1 - x_1) - z x_1 R^{[1]'}(1 - x_1), \\ \tilde{R}_\epsilon^{[2]}(z) &= R^{[2]}(1 - x_2 + z x_2) - R^{[2]}(1 - x_2) - z x_2 R^{[2]'}(1 - x_2),\end{aligned}$$

where  $(x_1, x_2)$  is the FP at channel erasure rate  $\epsilon$ .

*Proof.* The fraction of indices  $i$  where  $Z_i = ?$  is  $\epsilon$ . Therefore, the probability that bit nodes of the original graph at these indices remain in the residual graph is  $\epsilon L(y_1)$  and  $\epsilon L(y_2)$  corresponding to user 1 and user 2, respectively.

Meanwhile, the fraction of indices  $i$  where  $Z_i$  is not erased is  $1 - \epsilon$ . Half of these indices belongs to the case when  $Z_i \neq 0$ , i.e.,  $Z_i \in \{-2, 2\}$  where the decoder can perfectly recover  $X_i^{[1]}$  and  $X_i^{[2]}$ , i.e., the corresponding bit nodes are removed from the residual graph. Meanwhile, at each index  $i$  belonging to the other half where  $Z_i = 0$ , the corresponding bit nodes of user 1 and user 2 must have the same value and can be merged as a larger bit node. For these larger bit nodes to remain in the residual graph, the messages from the check nodes of both users must be erasures and

this happens with probability  $L(y_1)L(y_2)$ .

Thus, the d.d. of bit nodes in the residual graph is

$$\tilde{L}_\epsilon(z_1, z_2) = \epsilon \left( L^{[1]}(y_1 z_1) + L^{[2]}(y_2 z_2) \right) + \frac{1}{2}(1 - \epsilon)L^{[1]}(y_1 z_1)L^{[2]}(y_2 z_2)$$

because each edge of type  $j$ , for  $j \in \{1, 2\}$ , is associated with a erasure probability  $y_j$  from check nodes to bit nodes.  $\square$

**Theorem 15.** *At erasure rate  $\epsilon = \bar{\epsilon}^{\text{MAP}}$ , the design rate of the residual graph  $\tilde{r}_{\bar{\epsilon}^{\text{MAP}}}$  equals zero.*

*Proof.* Similarly to the proof of Theorem 13, one knows that the total number of check nodes in the residual graph is  $n \left( \frac{L^{[1]'}(1)}{R^{[1]'}(1)} \tilde{R}_\epsilon^{[1]}(1) + \frac{L^{[2]'}(1)}{R^{[2]'}(1)} \tilde{R}_\epsilon^{[2]}(1) \right)$ . Also, the number of bit nodes in the residual graph is  $n \tilde{L}_\epsilon(1, 1)$ .

Therefore, the design rate of the residual graph is

$$\tilde{r}_\epsilon = 1 - \left( \frac{L^{[1]'}(1)}{R^{[1]'}(1)} \tilde{R}_\epsilon^{[1]}(1) + \frac{L^{[2]'}(1)}{R^{[2]'}(1)} \tilde{R}_\epsilon^{[2]}(1) \right) / \tilde{L}_\epsilon(1, 1). \quad (4.22)$$

From (4.22), Lemma 29 and Corollary 8, it is clear that

$$\tilde{r}_{\bar{\epsilon}^{\text{MAP}}} = P(x^{\text{MAP}}) / \tilde{L}_{\bar{\epsilon}^{\text{MAP}}}(1, 1) = 0.$$

$\square$

**Remark 23.** *Similar to Remark 22, one can use the test in [93, p. 11] to show the tightness of the bound. For example, let us consider the case when user 1 and user 2 use the (3, 6)-regular and (3, 9)-regular LDPC ensembles, respectively. From Fig. 28 (b), the maximum of  $\theta(e_1, e_2)$ , for the corresponding residual graph, over the unit square is zero. Once this is true, the actual rate of this residual graph equals its design rate, hence equals zero, with high probability as  $n \rightarrow \infty$  and consequently, the bound is tight.*

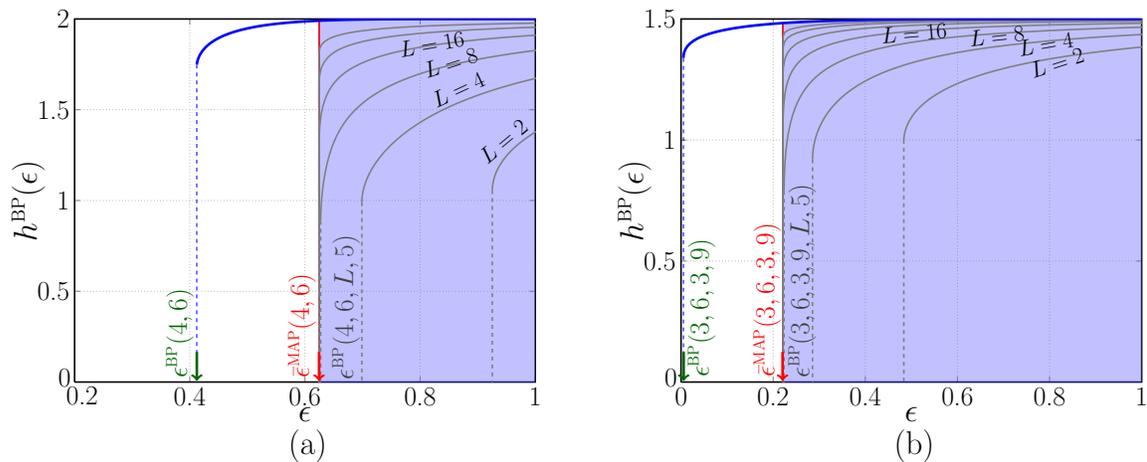


Fig. 30. BP-EXIT curves and MAP threshold for: (a) the (4,6)-regular and (4,6,  $L$ , 5) SC ensembles for the SWE where  $A = 3/2$  and  $L = 2, 4, 8, 16, 32, 64$ , (b) the (3,6,3,9) uncoupled and (3,6,3,9,  $L$ , 5) SC ensembles for the EMAC for  $L = 2, 4, 8, 16, 32, 64, 128$ .

#### D. Threshold Saturation of Spatially-Coupled Codes

One important application of our analysis is that one can compare the MAP thresholds of uncoupled LDPC ensembles with the BP thresholds of SC ensembles to observe the threshold saturation. This can be nicely seen by plotting the BP-EXIT curves for both the coupled and uncoupled systems. In Fig. 30 (a), the BP-EXIT curves for the (4,6,  $L$ , 5) SC ensembles based on the punctured (4,6)-regular ensemble (see [64]) are plotted for the SWE with an asymmetric channel condition where  $A = \frac{3}{2}$  (see Remark 22) and  $p = 0.5$ . It can be seen that as  $L$  increases, these curves saturate to the MAP threshold of the uncoupled system. Similarly, in Fig. 30 (b) for the EMAC, the BP-EXIT curves for the (3,6,3,9,  $L$ , 5) SC system, i.e., the two users use the (3,6,  $L$ , 5) and (3,9,  $L$ , 9) ensembles respectively, also saturate to the MAP threshold of the uncoupled system.

Similar plots for the SWE with symmetric channel conditions ( $A = 1$ ) and the EMAC with symmetric user rates ( $r_1 = r_2$ ) were also plotted in [64] and [63], respec-

tively, but without a rigorous consideration of the MAP thresholds. On the other hand, the main result of this chapter is not to demonstrate the impressive performance of SC codes under joint BP decoding but to focus on the MAP threshold evaluation. With this analysis on the MAP threshold, now one can observe that the saturation point of the BP thresholds of SC turns out to be the MAP threshold of the underlying ensembles for these two multiuser problems. We believe this is also a required step in any proof of threshold saturation for these systems.

## CHAPTER V

## CONCLUSIONS AND FUTURE WORK

This dissertation studies several coding techniques based on two popular classes of error-correcting codes, namely Reed-Solomon codes and LDPC codes. Data storage systems where these two families of codes prevail are among the most important applications of our work. In this chapter, we summarize the main contributions and also point out some potential future work.

#### A. A Rate-Distortion Framework to Analyze and Design Multiple Decoding Attempts of Reed-Solomon Codes

##### 1. Summary

In Chapter II, a unified framework based on rate-distortion (RD) theory is developed to analyze multiple decoding trials, with various algorithms, of RS codes in terms of performance and complexity. An important contribution is the connection made between the complexity and performance (in an asymptotic sense) of these multiple-decoding algorithms and the rate-distortion of an associated RD problem. Based on this analysis, we propose two solutions; the first is based on the RD function and the second on the RD exponent (RDE).

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 in Chapter II). Secondly, it helps estimate the error probability using exponentially tight bounds for  $n$  large enough. Further, we have shown that covering codes can also be combined with the RD approach to mitigate the suboptimality of

random codes when the effective block-length is not large enough. As part of this analysis, we also present numerical and analytical computations of the RD and RDE functions for sequences of i.n.d. sources. Finally, the simulation results show that our proposed algorithms based on the RD and RDE approaches achieve a better performance-versus-complexity trade-off than previously proposed algorithms. One key result is that, for the (255, 239) RS code, multiple-decoding using the standard Berlekamp-Massey algorithm (mBM) is as good as multiple-decoding using more complex algebraic soft-decision algorithms (mASD). However, for the (458, 410) RS code, the RDE approach improves the performance of mASD algorithms beyond that of mBM decoding.

## 2. Future Work

Simulations results suggest an interesting conjecture that, for moderate-rate RS codes, multiple ASD decoding attempts with small  $\mu$  is preferred while for low-rate RS codes, a single ASD decoding with large  $\mu$  may be preferred. This conjecture remains open for future research. Our future work will also focus on extending this framework to analyze multiple decoding attempts for intersymbol-interference channels. In this case, it will be appropriate for the decoder to consider multiple candidate error-events during decoding. Extending the RD and RDE approaches directly to this case is not straightforward since computing the RD and RDE functions for Markov sources in the large distortion regime is still an open problem. Another interesting extension is to use clever techniques to reuse the computations from one stage of errors-and-erasures decoding to the next in order to lower the complexity per decoding trial (e.g., [29]).

## B. Applications of Spatially-Coupled Codes via Threshold Saturation

### 1. Summary

In Chapter III, we consider binary communication over ISI channels and numerically show that, for spatially-coupled codes, threshold saturation occurs on several channels from the family of GECs as well as the dicode and PR2 channels with AWGN. To do this, we construct the EXIT and GEXIT curves that satisfy the area theorem and obtain an upper bound on the threshold of the MAP decoder. This upper bound is conjectured to be tight and, for the DEC, we show a numerical evidence which strongly supports this conjecture. The observed threshold saturation effect has an important implication: it suggests that universal performance under joint BP decoding is possible in practice by first finding a regular LDPC ensemble that has the performance close to the “capacity” under MAP decoding and then spatially coupling this underlying ensemble. Although numerical results are shown for these particular channels, the overall method is readily applicable to ISI channels with higher memory.

In Chapter IV, a similar analysis is extended to obtain an upper bound on the MAP thresholds of LDPC codes for two multiuser systems, namely the noisy Slepian-Wolf (SW) problem and the two-user multiple access channel (MAC). We deliberately focus on the models with erasures because this simplicity enables us to derive a rigorous analysis and show that the bound is tight in some cases. As a consequence of this analysis, threshold saturation of spatially-coupled codes is also observed over these multiuser systems. It then suggests that via spatial coupling, it is possible to design practical codes to universally achieve the entire capacity region of the two problems we consider.

## 2. Future Work

It has been known that the spatially-coupled codes (or LDPC convolutional codes) inherit some other advantages such as the typical minimum distance and the size of the smallest non-empty trapping sets both growing linearly with the protograph expansion  $M$  [94]. In addition, the convolutional structure of the codes allows one to consider a windowed decoder like the one discussed in [95, 96]. All of these properties suggest that spatially-coupled codes may be competitive in practice for systems with ISI, which are usually used to model the magnetic recording systems in data storage. Also, techniques to mitigate the rate loss induced by spatial coupling also need to be addressed to improve the finite-length performance. The detailed solutions to these practical challenges remain future lines of work we would like to consider.

Besides, we believe that applying spatial coupling to two-dimensional (2D) ISI channels will lead to substantial progress towards computing and achieving the SIR of 2D-ISI channels, which is unknown in general. Also, a general proof of threshold saturation for these systems is a challenging and important open problem.

## REFERENCES

- [1] C. E. Shannon, “A mathematical theory of communication,” *The Bell Syst. Techn. J.*, vol. 27, pp. 379–423, 623–656, Jul. / Oct. 1948.
- [2] P. Elias, “Coding for noisy channels,” in *IRE Conf. Rec.*, ser. pt. 4, 1955, pp. 37–46.
- [3] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
- [4] R. E. Blahut, *Algebraic Codes for Data Transmission*. Cambridge, UK: Cambridge University Press, 2003, ISBN-10 0521553741.
- [5] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2004, ISBN-13: 978-0130426727.
- [6] T. J. Richardson and R. L. Urbanke, *Modern Coding Theory*. Cambridge, UK: Cambridge University Press, 2008.
- [7] E. Arikan, “Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inform. Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.
- [8] A. Kavčić and A. Patapoutian, “The read channel,” in *Proceedings of the IEEE*, vol. 96, no. 11, pp. 1761–1774, Nov. 2008.
- [9] I. S. Reed and G. Solomon, “Polynomial codes over certain finite fields,” *J. Soc. Indust. Math.*, vol. 8, no. 2, pp. 300–304, 1960.

- [10] R. Singleton, “Maximum distance  $q$ -nary codes,” *IEEE Trans. Inform. Theory*, vol. 10, no. 2, pp. 116–118, Apr. 1964.
- [11] R. G. Gallager, “Low-density parity-check codes,” Ph.D. dissertation, M.I.T., Cambridge, MA, USA, 1960.
- [12] D. J. C. MacKay, “Good error-correcting codes based on very sparse matrices,” *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, Mar. 1999.
- [13] R. M. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.
- [14] M. G. Luby, M. Mitzenmacher, and M. A. Shokrollahi, “Practical loss-resilient codes,” in *Proc. 29th Annu. ACM Symp. Theory of Computing*, 1997, pp. 150–159.
- [15] M. G. Luby, M. Mitzenmacher, M. A. Shokrollahi, and D. A. Spielman, “Efficient erasure correcting codes,” *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.
- [16] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, “Design of capacity-approaching irregular low-density parity-check codes,” *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.
- [17] J. Felstrom and K. S. Zigangirov, “Time-varying periodic convolutional codes with low-density parity-check matrix,” *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2181–2191, 1999.
- [18] A. Sridharan, M. Lentmaier, D. J. Costello, and K. S. Zigangirov, “Convergence analysis of a class of LDPC convolutional codes for the erasure channel,” in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, 2004, pp. 953–962.

- [19] M. Lentmaier, A. Sridharan, K. Zigangirov, and D. J. Costello, “Terminated LDPC convolutional codes with thresholds close to capacity,” in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, 2005, pp. 1372–1376.
- [20] S. Kudekar, T. Richardson, and R. Urbanke, “Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC,” *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 803–834, Feb. 2011.
- [21] J. Thorpe, “Low-density parity-check (LDPC) codes constructed from protographs,” *IPN Progress Report*, vol. 42, no. 154, Aug. 2003. [Online]. Available: [http://tmo.jpl.nasa.gov/progress\\_report/42-154/154C.pdf](http://tmo.jpl.nasa.gov/progress_report/42-154/154C.pdf).
- [22] 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, Sep. 1999.
- [23] 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.
- [24] G. D. Forney, Jr., “Generalized minimum distance decoding,” *IEEE Trans. Inform. Theory*, vol. 12, no. 2, pp. 125–131, Apr. 1966.
- [25] 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.
- [26] 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.

- [27] Y. L. M. F. H. Tang 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.
- [28] H. Tokushige, I. Hisadomi, and T. Kasami, "Selection of test patterns in an iterative erasure and error decoding algorithm for non-binary block codes," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E89-A, no. 11, pp. 3355–3359, Nov. 2006.
- [29] J. Bellorado and A. Kavcic, "Low-complexity soft-decoding algorithms for Reed-Solomon codes - part I: An algebraic soft-in hard-out chase decoder," *IEEE Trans. Inform. Theory*, vol. 56, no. 3, pp. 945–959, Mar. 2010.
- [30] 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, Jul. 2007.
- [31] 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.
- [32] J. Jiang and K. R. Narayanan, "Iterative soft-input-soft-output decoding of Reed-Solomon codes by adapting the parity-check matrix," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3746–3756, Aug. 2006.
- [33] M. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on order statistics," *IEEE Trans. Inform. Theory*, vol. 41, no. 5, pp. 1379–1396, 1995.
- [34] F. Parvaresh and A. Vardy, "Multiplicity assignments for algebraic soft-decoding of Reed-Solomon codes," in *Proc. IEEE Int. Symp. Inform. Theory*, Yokohama,

Japan, Jul. 2003, p. 205.

- [35] 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.
- [36] 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, Nov. 2005.
- [37] 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. Inform. Theory*, Seoul, Korea, Jun. 2009, pp. 1248–1252.
- [38] 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, pp. 1235–1242.
- [39] —, “A rate-distortion exponent approach to multiple decoding attempts for Reed-Solomon codes,” in *Proc. IEEE Int. Symp. Inform. Theory*, Austin, TX, Jun. 2010, pp. 1798–1802.
- [40] —, “On multiple decoding attempts for Reed-Solomon codes: A rate-distortion approach,” *IEEE Trans. Inform. Theory*, vol. 57, no. 2, pp. 668–691, Feb. 2011.
- [41] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Wiley Series in Telecommunications. Hoboken, NJ, USA: Wiley, 1991.

- [42] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 4, pp. 405–417, July 1974.
- [43] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 197–199, Mar. 1974.
- [44] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*. Akademiai Kiado, Budapest, Hungary, 1981.
- [45] R. E. Blahut, *Principles and practice of information theory*. Reading, MA, USA: Addison-Wesley, 1987, ISBN-0201107090.
- [46] ———, "Computation of channel capacity and rate distortion functions," *IEEE Trans. Inform. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.
- [47] T. Berger, *Rate Distortion Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, Inc., 1971, ISBN 13-753103-6.
- [48] I. Csiszar, "On the computation of rate-distortion functions," *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 122–124, Jan. 1974.
- [49] S. Arimoto, "Computation of random coding exponent functions," *IEEE Trans. Inform. Theory*, vol. 22, no. 6, pp. 665–671, Nov. 1976.
- [50] G. Matz and P. Duhamel, "Information geometric formulation and interpretation of accelerated Blahut-Arimoto-type algorithms," in *Proc. IEEE Inform. Theory Workshop*, San Antonio, TX, Oct. 2004, pp. 66–70.
- [51] R. J. McEliece, "The Guruswami-Sudan decoding algorithm for Reed-Solomon codes," *JPL Interplanetary Network Progress Report*, pp. 42–153, May 2003.

- [52] W. J. Gross, F. R. Kschischang, R. Koetter, and P. G. Gulak, "Applications of algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1224–1234, 2006.
- [53] X. Zhang, "Reduced complexity interpolation architecture for soft-decision Reed-Solomon decoding," *IEEE Trans. VLSI Syst.*, vol. 14, no. 10, pp. 1156–1161, 2006.
- [54] J. Ma, A. Vardy, and Z. Wang, "Low-latency factorization architecture for algebraic soft-decision decoding of Reed-Solomon codes," *IEEE Trans. VLSI Syst.*, vol. 15, no. 11, pp. 1225–1238, 2007.
- [55] 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, Sep. 2008.
- [56] H. Tokushige, T. Koumoto, M. Fossorier, and T. Kasami, "Selection method of test patterns in soft-decision iterative bounded distance decoding algorithms," *IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86-A, no. 10, pp. 2445–2451, Oct. 2003.
- [57] S. Chung, G. D. Forney, Jr., T. J. Richardson, and R. L. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Letters*, vol. 5, no. 2, pp. 58–60, Feb. 2001.
- [58] M. Lentmaier and G. Fettweis, "On the thresholds of generalized LDPC convolutional codes based on protographs," in *Proc. IEEE Int. Symp. Inform. Theory*. IEEE, 2010, pp. 709–713.
- [59] S. Kudekar, C. Méasson, T. Richardson, and R. Urbanke, "Threshold saturation on BMS channels via spatial coupling," in *Proc. Int. Symp. on Turbo Codes &*

*Iterative Inform. Proc.*, Sep. 2010, pp. 309–313.

- [60] S. Kudekar and H. Pfister, “The effect of spatial coupling on compressive sensing,” in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Oct. 2010, pp. 347–353.
- [61] S. Hassani, N. Macris, and R. Urbanke, “Coupled graphical models and their thresholds,” in *Proc. IEEE Inform. Theory Workshop*, Dublin, Ireland, 2010, pp. 1–5.
- [62] S. Kudekar and K. Kasai, “Threshold saturation on channels with memory via spatial coupling,” in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 2562–2566.
- [63] ———, “Spatially coupled codes over the multiple access channel,” in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 2816–2820.
- [64] A. Yedla, H. Pfister, and K. Narayanan, “Universality for the noisy Slepian-Wolf problem via spatial coupling,” in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 2567–2571.
- [65] A. Yedla, P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, “Universal codes for the Gaussian MAC via spatial coupling,” in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Sep. 2011.
- [66] A. Kavčić, X. Ma, and M. Mitzenmacher, “Binary intersymbol interference channels: Gallager codes, density evolution and code performance bounds,” *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1636–1652, Jul. 2003.
- [67] D. Arnold and H. Loeliger, “On the information rate of binary-input channels with memory,” in *Proc. IEEE Int. Conf. Commun.*, Helsinki, Finland, Jun. 2001,

pp. 2692–2695.

- [68] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, “On the achievable information rates of finite state ISI channels,” in *Proc. IEEE Global Telecom. Conf.*, San Antonio, Texas, USA, Nov. 2001, pp. 2992–2996.
- [69] B. M. Kurkoski, P. H. Siegel, and J. K. Wolf, “Joint message-passing decoding of LDPC codes and partial-response channels,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1410–1422, Jun. 2002.
- [70] H. D. Pfister and P. H. Siegel, “Joint iterative decoding of LDPC codes and channels with memory,” in *Proc. 3rd Int. Symp. on Turbo Codes & Related Topics*, Brest, France, Sep. 2003, pp. 15–18.
- [71] —, “Joint iterative decoding of LDPC codes for channels with memory and erasure noise,” *IEEE J. Select. Areas Commun.*, vol. 26, no. 2, pp. 320–337, Feb. 2008.
- [72] N. Varnica and A. Kavčić, “Optimized low-density parity-check codes for partial response channels,” *IEEE Commun. Letters*, vol. 7, no. 4, pp. 168–170, 2003.
- [73] K. R. Narayanan and N. Nangare, “A BCJR-DFE based receiver for achieving near capacity performance on inter symbol interference channels,” in *Proc. 43rd Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Oct. 2004, pp. 763–772.
- [74] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, “Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding,” *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1416–1429, Apr. 2007.

- [75] H. D. Pfister, “On the capacity of finite state channels and the analysis of convolutional accumulate- $m$  codes,” Ph.D. dissertation, University of California, San Diego, La Jolla, CA, USA, Mar. 2003.
- [76] S. Sekido, K. Kasai, and K. Sakaniwa, “Threshold saturation on PR2 erasure channel via spatial coupling,” in *Proc. 34th Symp. Inform. Theory and its Appl.*, Ousyuku, Iwate, Japan, Nov. 2011, in Japanese.
- [77] C. Méasson, A. Montanari, and R. L. Urbanke, “Maxwell construction: The hidden bridge between iterative and maximum a posteriori decoding,” *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5277–5307, Dec. 2008.
- [78] C. Wang and H. D. Pfister, “Upper bounds on the MAP threshold of iterative decoding systems with erasure noise,” in *Proc. Int. Symp. on Turbo Codes & Related Topics*, Lausanne, Switzerland, Sep. 2008, pp. 7–12.
- [79] J. H. Bae and A. Anastasopoulos, “Capacity-achieving codes for finite-state channels with maximum-likelihood decoding,” *IEEE J. Select. Areas Commun.*, vol. 27, no. 6, pp. 974–984, Aug. 2009.
- [80] A. Yedla, Y. Y. Jian, P. S. Nguyen, and H. D. Pfister, “A simple proof of threshold saturation for coupled scalar recursions,” Arxiv preprint, 2012.
- [81] P. S. Nguyen, A. Yedla, H. D. Pfister, and K. R. Narayanan, “Threshold saturation of spatially-coupled codes on intersymbol-interference channels,” June 2012, accepted to *IEEE Int. Conf. Commun.*, Ottawa, Canada, Jun. 2012.
- [82] —, “Spatially-coupled codes and threshold saturation on intersymbol-interference channels,” 2012, to be submitted to *IEEE Trans. on Inform. Theory*, [Online]. Available: <http://arxiv.org/abs/1107.3253>.

- [83] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, “Optimal decoding of linear codes for minimizing symbol error rate,” *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 284–287, Mar. 1974.
- [84] D. Arnold, H. A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, “Simulation-based computation of information rates for channels with memory,” *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.
- [85] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, P. Didier, and A. Glavieux, “Iterative correction of intersymbol interference: Turbo equalization,” *Eur. Trans. Telecom.*, vol. 6, no. 5, pp. 507–511, Sep. – Oct. 1995.
- [86] J. Hou, P. H. Siegel, L. B. Milstein, and H. D. Pfister, “Capacity-approaching bandwidth-efficient coded modulation schemes based on low-density parity-check codes,” *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2141–2155, Sep. 2003.
- [87] A. Ashikhmin, G. Kramer, and S. ten Brink, “Extrinsic information transfer functions: model and erasure channel properties,” *IEEE Trans. Inform. Theory*, vol. 50, no. 11, pp. 2657–2674, Nov. 2004.
- [88] C. Méasson, A. Montanari, and R. Urbanke, “Asymptotic rate versus design rate,” in *Proc. IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 1541–1545.
- [89] H. D. Pfister and I. Sason, “Accumulate–repeat–accumulate codes: Capacity-achieving ensembles of systematic codes for the erasure channel with bounded complexity,” *IEEE Trans. Inform. Theory*, vol. 53, no. 6, pp. 2088–2115, Jun. 2007.
- [90] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke, “The generalized

- area theorem and some of its consequences,” *IEEE Trans. Inform. Theory*, vol. 55, no. 11, pp. 4793–4821, Nov. 2009.
- [91] ———, “Maximum a posteriori decoding and turbo codes for general memoryless channels,” in *Proc. IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, 2005, pp. 1241–1245.
- [92] M. Lentmaier, A. Sridharan, D. J. Costello, and K. S. Zigangirov, “Iterative decoding threshold analysis for LDPC convolutional codes,” *IEEE Trans. Inform. Theory*, vol. 56, no. 10, pp. 5274–5289, Oct. 2010.
- [93] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund, “Performance analysis and design of two edge type LDPC codes for the BEC wiretap channel,” Sep. 2010, [Online]. Available: <http://arxiv.org/abs/1009.4610>.
- [94] D. G. M. Mitchell, A. E. Pusane, M. Lentmaier, and D. J. Costello, “Exact free distance and trapping set growth rates for LDPC convolutional codes,” in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 1096–1100.
- [95] A. R. Iyengar, M. Papaleo, P. H. Siegel, J. K. Wolf, A. Vanelli-Coralli, and G. E. Corazza, “Windowed decoding of protograph-based LDPC convolutional codes over erasure channels,” Oct. 2010, submitted to *IEEE Trans. on Inform. Theory* [Online]. Available: <http://arxiv.org/abs/1010.4548>.
- [96] A. R. Iyengar, P. H. Siegel, R. L. Urbanke, and J. K. Wolf, “Windowed decoding of spatially coupled codes,” in *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, Jul. 2011, pp. 2552–2556.

## VITA

Phong Sy Nguyen (Nguyễn Sỹ Phong in Vietnamese) received the B.Eng. degree in electronics and telecommunications from Hanoi University of Technology, Hanoi, Vietnam, in 2004 and the Ph.D. degree in electrical and computer engineering from Texas A&M University, College Station, in 2012.

From 2004 to 2006, he was a system engineer at Vietnam Data-communication Company in Hanoi, Vietnam. Upon graduation, he will be joining Marvell Semiconductor, Inc., in Santa Clara, CA, as a senior DSP engineer. His research interests include information theory, signal processing, and channel coding with applications in wireless communications and data storage. Mr. Nguyen is a Fellow of the Vietnam Education Foundation, cohort 2006.

He can be reached at Department of Electrical and Computer Engineering, 214 ZEC, College Station, Texas, 77843-3128 or at [phongce@gmail.com](mailto:phongce@gmail.com).

The typist for this thesis was Phong Sy Nguyen.