Distinguishing Codes from Noise: Fundamental Limits and Applications to Sparse Communication

by

Da Wang

B.A.Sc, Electrical Engineering
University of Toronto, 2008

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author . . . .
Department of Electrical Engineering and Computer Science
May 21, 2010

Certified by . . . .
Gregory W. Wornell
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by . . . .
Terry P. Orlando
Chairman, Department Committee on Graduate Students

Distinguishing Codes from Noise: Fundamental Limits and Applications to Sparse Communication

by

Da Wang

Submitted to the Department of Electrical Engineering and Computer Science
on May 21, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

This thesis investigates the problem of distinguishing codes from noise. We develop a slotted channel model where in each time slot, the channel input is either a codeword or a noise sequence. In this model, successful communication requires both correctly detecting the presence of a codeword and decoding it to the correct message. While the decoding problem has been extensively studied, the problem of distinguishing codes from noise is relatively new, and we ask the following question regarding the "distinguishability" of a channel code: given a noisy channel and a code with a certain rate, what are the fundamental limits of distinguishing this code from noise at the output of the channel?

The problem of distinguishing codes from noise involves both detection and decoding. In our analysis, we first extend the classical channel coding problem to incorporate the requirement of detection, which admits both miss and false alarm errors. Then we investigate the fundamental limits of code distinguishing in terms of the error exponents of miss and false alarm error probabilities. In a scenario that miss probability is required to vanish asymptotically but not necessarily exponentially, we characterize the maximum false alarm error exponent at each rate, and show that an i.i.d. codebook with typicality decoding is sufficient to achieve the maximum exponent. In another scenario that requires certain miss error exponent, we show that for DMC channels, the i.i.d. codebook is suboptimal and the constant composition codebook achieves the best known performance. For AWGN channels, we develop a clustered spherical codebook that achieves the best known performance in all operating regimes.

This code distinguishability problem is strongly motivated by the synchronization problem in sparse communication, a new communication paradigm where transmissions take place intermittently and each transmission consists of a small amount of data. Our results show that, in sparse communication, the traditional approach of conducting synchronization and coding separately is suboptimal, and our approach of designing codes for joint synchronization and information transmission achieves better performance, especially at high rates. Therefore, for systems with sparse transmissions such as sensor networks, it is beneficial to adopt the joint sync–coding architecture instead of the traditional separate sync–coding architecture.

Thesis Supervisor: Gregory W. Wornell
Title: Professor of Electrical Engineering and Computer Science

Acknowledgments

My first and foremost appreciation goes to my advisor, Greg Wornell, for his guidance, insights and support throughout my graduate studies. Building the lab as an intellectual community, Greg provides a nurturing environment for one's academic growth, and I truly admire his vision and dedication. Besides Greg, this work could not have been completed without Sae-Young Chung and Venkat Chandar, who provided me with much help on the problem formulation and problem solving. And my sincere thanks to Yuval Kochman and Charles Swannack for all the helpful discussions on n-dimensional geometry and beyond. It is my tremendous fortune that I can pick the brains of so many excellent researchers! I would also like to thank the other members of the Signals, Information and Algorithms Laboratory: Anthony Accardi, Vijay Divi, Qing He, Ying-zong Huang, Ashish Khisti, James Krieger, Urs Niesen, Tricia O'Donnell, and Maryam Shanechi for making this group wonderful! And my special thanks to James Krieger and Venkat Chandar for being cool officemates!

I am also very grateful for the wonderful community at MIT. My sincere thanks to Lizhong Zheng for sharing his insights and giving vivid lectures, and to Al Oppenheim for doing a lot more than an academic advisor. The "36-6th floor" community is amazing as well. Among them, I especially thank John Sun for many valuable advices and discussions, and Dennis Wei for the comradeship during the visit to HP Labs at Palo Alto.

Beyond people at MIT, I would like to thank those who introduced me to the world of intellectual discovery at the University of Toronto: Frank Kschischang, Ben Liang, Danilo Silva and Wei Yu. My academic career is simply impossible without their help and encouragement. I am also very grateful to Jonathan Yedidia from MERL, Mitch Trott and Pascal Vontobel from HP Labs, who hosted me during my visits.

Finally, I want to thank my parents for raising me, educating me and supporting me, and my greatest appreciation goes to Cathy, who enlightens and enriches my life every day.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Synchronization has long been recognized as a fundamental aspect of communication (e.g., [1, 2]). More specifically, to establish communication, a transmitter sends a codeword based on an input message, and a receiver must detect the presence of the codeword, locate its starting position, and then determine which message it corresponds to. The process of detecting and locating a codeword is usually called *initial/one-shot frame synchronization* [3], and we simply refer to it as "synchronization" in this thesis.

The traditional solution to the problem of synchronization is training, where a specific pattern of symbols, known as *sync word*, is used to identify the start of transmitted data (or codeword). This approach separates the communication problem into two sub-problems, *synchronization* and *coding*. Separate sync–coding provides great engineering convenience in terms of both design and analysis. It is appropriate when communication is dense and hence the cost of synchronization can be "amortized", since many codewords are transmitted after synchronization is established (Figure 1-1). Indeed, many studies on synchronization adopt this separation approach and focus on designing good sync words and/or detection rules to improve synchronization performance (e.g., [4, 5, 6]).

However, as pointed out in [7], the above separation approach becomes difficult to justify for certain emerging applications such as the sensor network, where transmissions take place intermittently, with each transmission consisting of a small amount of data. We term this new type of communication paradigm *sparse communication* (Figure 1-2).

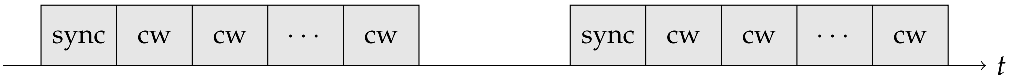


Figure 1-1: Dense communication: many codewords are transmitted after each synchronization.

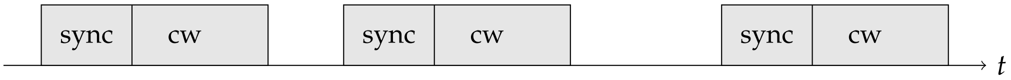


Figure 1-2: Sparse communication: each transmission consists a small amount of data.

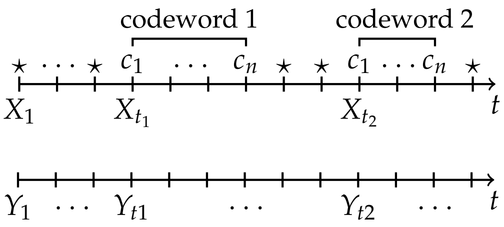

codeword 1
codeword 2
⋆ ⋯ ⋆ c_1 ⋯ c_n ⋆ ⋆ c_1 ⋯ c_n ⋆
X_1
X_{t_1}
X_{t_2}
t
Y_1 ⋯ Y_{t1} ⋯ Y_{t2} ⋯
t

Figure 1-3: Asynchronous channel inputs ($X_1, X_2, \cdots$) and outputs ($Y_1, Y_2, \cdots$).

In [7], the authors analyze the synchronization and information transmission aspects of communication jointly and develop an asynchronous channel model to capture not only the transmission from transmitter to receiver but also the channel condition when the transmitter is silent. Specifically, a special $\star$ symbol is introduced as a pseudo-input to the channel when nothing is sent from the transmitter. In this channel, both idle and transmission activities of the transmitter induce a sequence of channel inputs $X_1, X_2, \cdots$, where codewords with length n are sent only at time $t_1, t_2, \cdots$. Accordingly, the receiver observes a sequence of channel outputs $Y_1, Y_2, \cdots$, as shown in Figure 1-3.

In this thesis, we adopt the above channel model with a key simplification, where by we require the communication to be slotted, as shown in Figure 1-4. This simplification allows us to use a set of performance metrics and analysis techniques that are different from those in [7], while maintaining the essence of the synchronization problem.

Remark:

The slotted channel model introduces certain amount of synchrony between the transmitter and the receiver, and reduces the problem of synchronization from both detecting and locating the codeword to detecting the codeword only. However, as shown in [8], using a prefix with sublinear length is sufficient to locate the codeword, once we know a codeword is indeed present. Therefore, there is no essential performance difference between the slotted and unslotted model

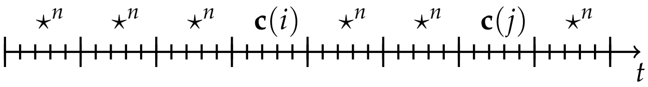


Figure 1-4: Slotted asynchronous channel, with each time slot containing either a codeword or a noise sequence with length n.

in terms of error exponents, which are the performance metric we adopt in this thesis, as discussed later.

In the slotted asynchronous channel, the channel output in each time slot is induced by either a codeword $c^n(i)$ or a noise sequence $\star^n$. In this model, successful communication requires both correctly detecting the presence of a codeword and decoding it to the correct message. To accommodate the analysis of both detection and decoding, we extend the classical coding problem defined in [9] to incorporate the requirement of detection, and ask the following *central question* regarding the "distinguishability" of a channel code:

> Given a noisy channel and a code with rate R, what are the fundamental limits of distinguishing this code from noise at the output of the channel?

Intuitively, we want the channel outputs corresponding to codewords be as "different" from the channel outputs of noise as possible, and answering this central question tells us how different they can be. In addition, the analysis process can uncover some useful insights on designing codes that are easy to distinguish from noise.

Fundamentally, distinguishing codes from noise, or detecting codewords with respect to noise, like decoding, is a hypothesis testing problem. Therefore, the problem of joint detection and decoding admits three types of errors: 1) miss (where we detect a codeword as noise), 2) false alarm (where we detect noise as a codeword), and 3) decoding error (where after correctly detecting the presence of the codeword, we decode it to an incorrect message). We denote the probabilities of these three events by P_{m}, P_{f} and P_{e} respectively.

In most applications, we are interested in the regime that P_{m}, P_{f} and P_{e} can be made arbitrarily small by increasing the codeword block length n. Then we can define the error exponents E_{m}, E_{f} and E_{d} for these three probabilities, which tell us asymptotically how fast these error probabilities decay as the codeword block length increases. As shown later, in most regimes of practical interest, these error probabilities indeed decrease exponentially and hence error exponents are useful measures of the system performance. Furthermore, given certain error probability requirements, error exponents are useful in obtaining a coarse estimate of the required codeword block length [10], an important system design parameter.

In general, there are trade-offs between the three exponents E_{m}, E_{f} and E_{d}, and we study these exponents and their trade-offs in several scenarios of practical interest.

First, we consider the case that communication is very sparse, i.e., we send very few messages over a large time interval. In this scenario, false alarm is the dominant error event, because the majority of the slots have input $\star^n$, which can only cause false alarm errors. Therefore, we may be interested in maximizing the false alarm error exponent E_f, while keeping P_m and P_e small. This is our focus of **Chapter 3**.

In addition, the above scenario of sparse communication can also be viewed as an extension of pulse position modulation (PPM), if we assume there is a predefined communication window between the transmitter and the receiver. In this setting, our results have an indication on the maximum throughput we can achieve when communicating by both timing and codeword, which is elaborated on in **Section 3.6**.

Second, we consider the above problem in another scenario, where users may have requirements on the error exponent of miss P_m. For example, in a fire alarm system or an enemy detection system, messages are so critical that one may want the probability of miss to decay with exponent at least E_m. This is analogous to the Neyman-Pearson test, but here the efficient frontier is characterized via the trade-off between the error exponents E_m and E_f, rather than the probabilities P_m and P_f. This problem, characterizing the optimal false alarm exponent E_f subject to E_m constraint, is investigated in both **Chapter 4** (for the DMC) and **Chapter 5** (for the AWGN channel).

It is worth pointing out that the above characterization of E_f given E_m constraint is equivalent to characterizing the achievable region of the $(E_\mathrm{m}, E_\mathrm{f})$ pairs. This actually provides us with a fundamental limit on sparse communication, when we require the total miss probability, total false alarm probability, and the average decoding error probability to be vanishing asymptotically. Specifically, if given $E_\mathrm{m} \geq \beta$, the optimal E_f is α, then we can at most transmit $e^{n\beta}$ messages over $e^{n\beta} + e^{n\alpha}$ many time slots, otherwise either the total miss probability or total false alarm probability is not vanishing asymptotically [1].

Based on our analysis, we compare the detection performance of joint sync–coding with separate sync–coding (training). This problem has been investigated in [11], where training is shown to be suboptimal at high rate. In our work, we show training is suboptimal at almost all rates and give a more precise quantification on the performance loss due to training.

Finally, we present a summary on the performance of different schemes in various regimes of interest in **Chapter 6**. This helps system designers to choose the proper scheme based on the sys-

[1] If we want the total decoding error probability to be vanishing as well, we need to impose the constraint that the decoding error exponent E_d is larger than β, and we need to analyze the achievable region of the triplet $(E_\mathrm{m}, E_\mathrm{f}, E_\mathrm{d})$.

tem operating regime and performance requirements. In addition, we mention a few related open problems that extend and connect the problem of distinguishing codes from noise to a broader context.

Chapter 2
Background

This chapter introduces the mathematical notations, definitions and models used in this thesis.

■ 2.1 Notations

We use lower case letters (e.g. x) to denote a particular value of the corresponding random variable denoted in capital letters (e.g. X). We use $\mathcal{P}(\mathcal{X})$ to denote all the probability distributions on the alphabet $\mathcal{X}$.

Below assumes that we have distributions $P(\cdot), Q(\cdot) \in \mathcal{P}(\mathcal{X})$ and conditional distributions $W(\cdot|\cdot) : \mathcal{X} \to \mathcal{Y}$, $V(\cdot|\cdot) : \mathcal{X} \to \mathcal{Y}$, and define

$$[P \cdot Q](x, y) \triangleq P(x)Q(y) \tag{2.1}$$

$$[P \cdot W](x, y) \triangleq W(y|x)P(x) \tag{2.2}$$

$$[P \cdot W]_Y(y) \triangleq \sum_x W(y|x)P(x). \tag{2.3}$$

We define the *shell* of a probability distribution as

$$[P]_\delta \triangleq \left\{P' \in \mathcal{P}(\mathcal{X}) : \|P - P'\|_\infty < \delta\right\}, \tag{2.4}$$

where $\|P(\cdot)\|_\infty \triangleq \max_{x \in \mathcal{X}} P(x)$ is the infinity norm.

We denote the support of a distribution P on $\mathcal{X}$ by

$$\text{Support}(P) \triangleq \{x \in \mathcal{X} : P(x) > 0\}, \tag{2.5}$$

and extend this definition to a set of distributions $\mathcal{P}$:

$$\text{Support}(\mathcal{P}) \triangleq \bigcup_{P \in \mathcal{P}} \text{Support}(P). \tag{2.6}$$

And we would use $\text{Support}(P_1, P_2, \cdots, P_n)$ as a shorthand for $\text{Support}(\{P_1, P_2, \cdots, P_n\})$.

For a collection of disjoint sets $\{\mathcal{A}_i\}$, we denote their union by $\bigsqcup \mathcal{A}_i$, where "$\bigsqcup$" is used to emphasize that the union is over disjoint sets.

■ 2.1.1 Probability distributions

This thesis refers to a few probability distributions, and we use $\mathsf{N}(\cdot,\cdot)$ to denote the Gaussian distribution [1] , $\mathsf{Bern}(\cdot)$ the Bernoulli distribution, and $\mathsf{Binom}(\cdot,\cdot)$ the Binomial distribution.

■ 2.2 Synchronous DMC and Block Codes

In this section we define the conventional (synchronous) discrete memoryless channel (DMC) $W : \mathcal{X} \to \mathcal{Y}$, with input alphabet $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$ and output alphabet $\mathcal{Y} = \{1, 2, \ldots, |\mathcal{Y}|\}$. The conditional distribution of output letter Y when the channel input letter X equals $x \in \mathcal{X}$ is denoted by $W_{Y|X}(\cdot|x)$:

$$\mathbb{P}\left[Y = y | X = x\right] = W_{Y|X}(y|x) \qquad \forall\, x \in \mathcal{X},\, \forall\, y \in \mathcal{Y}.$$

When the input and output alphabets are clear from context, W is used instead of $W_{Y|X}$.

A length n block code for the synchronous channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and some finite message set $\mathcal{M}_{f_n} = \{1, 2, \ldots, |\mathcal{M}_{f_n}|\}$ is composed of a pair of mappings, encoder mapping $f_n : \mathcal{M}_{f_n} \to \mathcal{X}^n$ and decoder mapping $g_n : \mathcal{Y}^n \to \mathcal{M}_{f_n}$. We denote (f_n, g_n) by $\mathcal{C}^{(n)}$.

Given a message m, which is chosen from $\mathcal{M}_{f_n}$ according to the *uniform* distribution, the encoder maps it to a sequence $x^n(m) \in \mathcal{X}^n$ [2] and transmits this sequence through the channel, where we call $x^n(m)$ the *codeword* for message m and the entire set of codewords $\{x^n(m)\}$ a *codebook*. The decoder receives a sequence $y^n \in \mathcal{Y}^n$ and maps it to a message $\hat{m}$. The decoding error probability for message m is

$$P_{\mathrm{e}}(m) \triangleq \sum_{\hat{m} \neq m} W^n\left(g_n^{-1}(\hat{m}) | f_n(m)\right) = 1 - W^n\left(g_n^{-1}(m) | f_n(m)\right). \tag{2.7}$$

The *maximal* decoding error probability P_{e} and rate R of the code $\mathcal{C}^{(n)}$ is given by

$$P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right) \triangleq \max_{m} P_{\mathrm{e}}(m) \qquad \text{and} \qquad R(\mathcal{C}^{(n)}) \triangleq \frac{\log |\mathcal{M}_{f_n}|}{n}.$$

For derivations in this thesis, we also use the concept of average error probability, where the average is over all possible codewords and/or codebooks, depending on the context. We use an overline to denote the averaging operation, for example, $\overline{P_{\mathrm{e}}}$.

[1] Also known as Normal distribution.

[2] We use $x_i^j (i \leq j)$ to represent the sequence $x_i, x_{i+1}, \ldots, x_j$, and further use x^n to represent the sequence x_1^n.

■ 2.2.1 Subcode and extension

A code $\hat{\mathcal{C}}^{(n)} = (\hat{f}_n, \hat{g}_n)$ is called a *subcode* of $\mathcal{C}^{(n)} = (f_n, g_n)$ if $\mathcal{M}_{\hat{f}_n} \subset \mathcal{M}_{f_n}$ and $\hat{f}_n(m) = f_n(m)$ for any $m \in \mathcal{M}_{\hat{f}_n}$. Also, we call $\mathcal{C}^{(n)}$ an *extension* of $\hat{\mathcal{C}}^{(n)}$.

Note that there is no restriction on the relationship between $\hat{g}_n$ and g_n in the above definitions.

■ 2.2.2 Code sequences

A sequence of codes $\mathcal{Q} = \left\{\mathcal{C}^{(n)}, n \in \mathbb{N}\right\}$ indexed by their block-lengths n is called *reliable* if

$$\lim_{n\to\infty} P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right) = 0.$$

For any reliable code-sequence $\mathcal{Q}$, the rate $R_{\mathcal{Q}}$ is given by

$$R_{\mathcal{Q}} \triangleq \liminf_{n\to\infty} R\left(\mathcal{C}^{(n)}\right).$$

■ 2.3 Asynchronous DMC and Block Codes

In this section we extend the formal definition of (synchronous) DMC to asynchronous channels to accommodate the potential timing uncertainty in the communication.

> **Definition 2.1** (Asynchronous discrete memoryless channel [7]). An asynchronous discrete memoryless channel (Figure 2-1) $(\mathcal{X}, \star, W, \mathcal{Y})$ is a DMC with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, and transition probabilities $W(y|x)$, where the special symbol $\star \in \mathcal{X}$ is used to represent the channel input when the transmitter is silent.

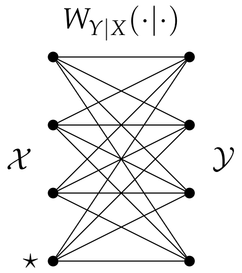


Figure 2-1: An asynchronous channel model with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$, special symbol $\star$ and transition probabilities $W(\cdot|\cdot)$.

For a length n block code for the slotted asynchronous channel, we have $g_n : \mathcal{A}_n \to \mathcal{M}_{f_n}$, where $\mathcal{A}_n \subset \mathcal{Y}^n$ is the acceptance region for codewords, i.e., if $y^n \in \mathcal{A}_n$, we consider the channel input as a certain codeword $x^n(m), m \in \mathcal{M}_{f_n}$, otherwise we consider the channel input as $\star^n$. We also use $\mathcal{B}_n = \mathcal{A}_n^c$ to denote the rejection region for codewords.

Remark:
Note that in the asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$, the domain of g_n is $\mathcal{A}_n$ rather than $\mathcal{Y}^n$, which is the domain of g_n in the synchronous DMC.

Furthermore, when the channel input is the noise symbol $\star$, we denote the output marginal distribution as Q_Y, i.e.,

$$Q_Y(\cdot) \triangleq W(\cdot|\star). \tag{2.8}$$

For asynchronous channels, in addition to $P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right)$, we have two error metrics corresponding to the detection problem, the probability of miss and the probability of false alarm, which depend on the acceptance region $\mathcal{A}_n$ of a code:

$$\begin{aligned} P_{\mathrm{m}}(m) &\triangleq W^n\left(\mathcal{A}_n{}^c | \, x^n(m)\right) \\ P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) &\triangleq \max_m P_{\mathrm{m}}(m) \\ P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) &\triangleq W^n\left(\mathcal{A}_n | \star^n\right). \end{aligned}$$

Note that by definition, $\mathcal{A}_n = \sqcup_{m \in \mathcal{M}_{f_n}} g_n^{-1}(m)$, and

$$P_{\mathrm{m}}(m) = W^n\left(\mathcal{A}_n{}^c | \, x^n(m)\right) \le W^n\left(g_n^{-1}(m)^c \middle| \, x^n(m)\right) = P_{\mathrm{e}}(m). \tag{2.9}$$

Thus

$$P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) \le P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right). \tag{2.10}$$

Moreover, we denote $P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right)$, $P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right)$ and $P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right)$ by $P_{\mathrm{e}}^{(n)}$, $P_{\mathrm{m}}^{(n)}$ and $P_{\mathrm{f}}^{(n)}$ when the code sequence is clear from context. Furthermore, we use the following shorthands

$$P_{\mathrm{e}} \triangleq \liminf_{n\to\infty} P_{\mathrm{e}}^{(n)} \tag{2.11}$$

$$P_{\mathrm{m}} \triangleq \liminf_{n\to\infty} P_{\mathrm{m}}^{(n)} \tag{2.12}$$

$$P_{\mathrm{f}} \triangleq \liminf_{n\to\infty} P_{\mathrm{f}}^{(n)}, \tag{2.13}$$

where the infimum is over all possible codes.

In this thesis, without loss of generality, we assume that for every $y \in \mathcal{Y}$, $\exists\, x$ such that $W(x|\,y) > 0$. Furthermore, we only consider the case $\mathrm{Support}(W(\cdot|\star)) = \mathcal{Y}$, which has the most interesting trade-off between the false alarm probability and the miss probability.

Remark:
When $\mathrm{Support}(W(\cdot|\star))$ is a proper subset of $\mathcal{Y}$, if we only require $P_{\mathrm{m}} \to 0$, it is not hard to see that we can achieve $P_{\mathrm{f}} = 0$ by designing the codebook and the detection rule properly. In this case, there is not much trade-off between the false alarm probability and the miss probability. If we require the miss probability to decay exponentially, certain trade-off exists but its analysis essentially reduces to the case that $\mathrm{Support}(W(\cdot|\star)) = \mathcal{Y}$.

■ 2.3.1 Error exponents for asynchronous DMC

In many regimes of interest, the error probabilities defined above decrease exponentially with the codeword block length n, and the corresponding exponents provide us with finer characterizations of the error performance. This motivates us to define the relevant error exponents for the asynchronous DMC.

Definition 2.2 (Achievable miss, false alarm, and decoding error exponents)**.** A number $e_{\mathrm{m}} \geq 0$ is called an *achievable miss error exponent* for an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ at rate R_{Q} if for any $\varepsilon > 0$, $\delta > 0$ and $\lambda > 0$, there exists a code $\mathcal{C}^{(n)}$ such that $R\left(\mathcal{C}^{(n)}\right) \geq R_{Q} - \delta$,

$$P_{\mathrm{m}}^{(n)} \leq \varepsilon \quad \text{and} \quad P_{\mathrm{m}}^{(n)} \leq e^{-n(e_{\mathrm{m}}-\lambda)},$$

when n sufficiently large.

Similarly, we can define the *achievable false alarm error exponent* e_{f} and *achievable decoding error exponent* e_{d} in terms of P_{f} and P_{e}.

Remark:
When $e_{\mathrm{m}} > 0$, $P_{\mathrm{m}}^{(n)} \leq e^{-n(e_{\mathrm{m}}-\lambda)}$ implies $P_{\mathrm{m}}^{(n)} \leq \varepsilon$, and hence the first condition is needed only when $e_{\mathrm{m}} = 0$.

In addition, since $P_{\mathrm{e}}^{(n)} \geq P_{\mathrm{m}}^{(n)}$, we have $e_{\mathrm{d}} \leq e_{\mathrm{m}}$.

Definition 2.3 (Achievable error exponent triplet)**.** A triplet of numbers $(e_{\mathrm{d}}, e_{\mathrm{m}}, e_{\mathrm{f}})$ is called *achievable for an asynchronous DMC* $(\mathcal{X}, \star, W, \mathcal{Y})$ *at rate* $R_{\mathcal{Q}}$ if they can be achieved simultaneously for this channel at rate $R_{\mathcal{Q}}$.

We denote the set of achievable error exponent triplets at rate $R_{\mathcal{Q}}$ by $\mathcal{E}(R_{\mathcal{Q}})$, i.e.,

$$\mathcal{E}(R_{\mathcal{Q}}) \triangleq \left\{(e_{\mathrm{m}}, e_{\mathrm{f}}, e_{\mathrm{d}}) : (e_{\mathrm{m}}, e_{\mathrm{f}}, e_{\mathrm{d}}) \text{ achievable at rate } R_{\mathcal{Q}}\right\}. \tag{2.14}$$

Definition 2.4 (Reliability functions)**.** For an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$, given rate $R_{\mathcal{Q}}$, we define the *false alarm reliability function* as

$$E_{\mathrm{f}}(R_{\mathcal{Q}}, e_{\mathrm{m}}, e_{\mathrm{d}}) \triangleq \sup_{(e_{\mathrm{m}}, e_{\mathrm{f}}, e_{\mathrm{d}}) \in \mathcal{E}(R_{\mathcal{Q}})} e_{\mathrm{f}} \tag{2.15}$$

and similarly, the *miss reliability function* as

$$E_{\mathrm{m}}(R_{\mathcal{Q}}, e_{\mathrm{f}}, e_{\mathrm{d}}) \triangleq \sup_{(e_{\mathrm{m}}, e_{\mathrm{f}}, e_{\mathrm{d}}) \in \mathcal{E}(R_{\mathcal{Q}})} e_{\mathrm{m}} \tag{2.16}$$

and the *decoding error reliability function* as

$$E_{\mathrm{d}}(R_{\mathcal{Q}}, e_{\mathrm{m}}, e_{\mathrm{f}}) \triangleq \sup_{(e_{\mathrm{m}}, e_{\mathrm{f}}, e_{\mathrm{d}}) \in \mathcal{E}(R_{\mathcal{Q}})} e_{\mathrm{d}}. \tag{2.17}$$

In this thesis, we analyze the optimal achievable e_{m} and e_{f} pairs without any constraints on e_{d}. It is not hard to see that $e_{\mathrm{d}} = 0$ gives us the best trade-off between e_{m} and e_{f}, so we let $e_{\mathrm{d}} = 0$, and adopt the following notations:

$$E_{\mathrm{f}}(R_{\mathcal{Q}}, e_{\mathrm{m}}) \triangleq E_{\mathrm{f}}(R_{\mathcal{Q}}, e_{\mathrm{m}}, e_{\mathrm{d}} = 0) \tag{2.18}$$

$$E_{\mathrm{m}}(R_{\mathcal{Q}}, e_{\mathrm{f}}) \triangleq E_{\mathrm{m}}(R_{\mathcal{Q}}, e_{\mathrm{f}}, e_{\mathrm{d}} = 0). \tag{2.19}$$

The reliability functions in (2.18) and (2.19) provide us with the general trade-off between miss and false alarm exponents. As mentioned in Chapter 1, we may also be interested in the case of maximizing the false alarm exponent, without any constraint for e_{m}. In this case, again it is not hard to see that setting $e_{\mathrm{m}} = 0$ allows us to achieve the maximum false alarm exponent. The objective of maximizing $E_{\mathrm{f}}(R_{\mathcal{Q}}, e_{\mathrm{m}})$ necessitates the following definition.

Definition 2.5. For an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$, given rate $R_{\mathcal{Q}}$, the *maximum false alarm reliability function* is

$$E_{\mathrm{f}}(R_{\mathcal{Q}}) \triangleq E_{\mathrm{f}}(R_{\mathcal{Q}}, e_{\mathrm{m}} = 0) \tag{2.20}$$

and similarly, the *maximum miss reliability function* is

$$E_{\mathrm{m}}(R_{\mathcal{Q}}) \triangleq E_{\mathrm{m}}(R_{\mathcal{Q}}, e_{\mathrm{f}} = 0). \tag{2.21}$$

Remark:

In this thesis, we characterize $E_{\mathrm{f}}(R_{\mathcal{Q}})$ exactly in Chapter 3 but not $E_{\mathrm{m}}(R_{\mathcal{Q}})$, as we only have partial results regarding $E_{\mathrm{m}}(R_{\mathcal{Q}})$, which can be obtained as a special case of the partial results in Chapter 4.

Sometimes it is useful to characterize a reliability function via its lower and upper bounds, where we use $\underline{E_{\mathrm{f}}}(R_{\mathcal{Q}}, \cdot)$ and $\overline{E_{\mathrm{f}}}(R_{\mathcal{Q}}, \cdot)$ to denote the *lower bound* and *upper bound* of the false alarm reliability function $E_{\mathrm{f}}(R_{\mathcal{Q}}, \cdot)$. Similarly, we denote the lower and upper bounds of the miss reliability function by $\underline{E_{\mathrm{m}}}(R_{\mathcal{Q}}, \cdot)$ and $\overline{E_{\mathrm{m}}}(R_{\mathcal{Q}}, \cdot)$.

Remark:

Definitions in this section can be generalized to channels with continuous input and output alphabets, such as the AWGN channel introduced in Section 2.4.2.

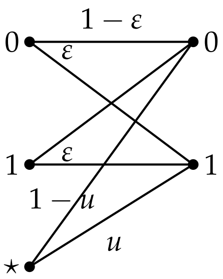


Figure 2-2: An asynchronous BSC with crossover probability ε and $\star$ output distribution $\mathsf{Bern}\left(u\right)$.

2.4 Important Asynchronous Channels

In this section we define the asynchronous versions of two useful channels in information theory: the binary symmetric channel (BSC) and the additive white Gaussian noise (AWGN) channel.

2.4.1 Binary symmetric channel

We define an asynchronous binary symmetric channel via its cross over probability ε and the $\star$ symbol output distribution $\mathsf{Bern}\left(u\right)$, as shown in Figure 2-2. Without loss of generality, we assume $\varepsilon < 1/2$ and $u \leq 1/2$.

A BSC with input distribution $\mathsf{Bern}\left(p\right)$ has output distribution $\mathsf{Bern}\left(s\right)$, where

$$s = p * \varepsilon \tag{2.22}$$

$$\triangleq p(1-\varepsilon) + (1-p)\varepsilon \tag{2.23}$$

and $*$ is called the "binary convolution".

2.4.2 Additive white Gaussian noise channel

We define a discrete-time additive white Gaussian noise (AWGN) channel with input X_i, output Y_i and noise Z_i as follows:

$$Y_i = X_i + Z_i, \quad Z_i \overset{i.i.d.}{\sim} \mathsf{N}\left(0,1\right), \tag{2.24}$$

where each codeword x^n transmitted over the channel satisfies the average power constraint:

$$\|x^n\|_2 \leq nP\,. \tag{2.25}$$

In addition, we have $Q_Y(\cdot) = W\left(\cdot|\star\right) = \mathsf{N}\left(0,1\right)$, i.e., when the $\star$ symbol is sent, the channel output distribution is the additive white Gaussian noise distribution.

Remark:

In this thesis, we normalize the noise power to 1, as in (2.24); hence P essentially denotes the SNR of the AWGN channel.

■ 2.5 Exponential Approximation

Since we are mainly concerned with error exponents in our performance evaluations, we define *equality in the exponential sense*, i.e., for a sequence a_n,

$$a_n \doteq e^{nF} \Leftrightarrow F = \lim_{n\to\infty} \frac{1}{n} \log a_n, \tag{2.26}$$

where the $\doteq$ sign denotes exponential equality, and where it is assumed that the limit in (2.26) exists.

■ 2.5.1 Algebras for exponential approximations

From the algebra of limits, we have for $a_n \doteq e^{nL_1}$ and $b_n \doteq e^{nL_2}$,

$$\lim_{n\to\infty} (a_n + b_n) \doteq e^{n \min\{L_1, L_2\}} \tag{2.27}$$

$$\lim_{n\to\infty} (a_n b_n) \doteq e^{n(L_1+L_2)}. \tag{2.28}$$

These rules are also valid for infinite limits using the rule

$$q + \infty = \infty \qquad\qquad \text{if } q \neq -\infty.$$

■ 2.5.2 Exponential inequalities

We extend the notion of exponential equalities to inequalities. If $a_n \doteq e^{nL}$ and $A \leq L \leq B$, we say

$$a_n \dot{\geq} e^{nA}$$

$$a_n \dot{\leq} e^{nB}.$$

■ 2.6 Quantities in Information Theory

We use the following notation for entropy, conditional entropy and mutual information:

$$\begin{aligned} H(P) &= \sum_{x\in\mathcal{X}} P(x)\log\frac{1}{P(x)} \\ H(W|P) &= \sum_{x\in\mathcal{X}} P(x)H(W(\cdot|x)) \\ I\left(P,W\right) &= \sum_{x\in\mathcal{X},y\in\mathcal{Y}} P(x)W(y|x)\log\frac{W(y|x)}{\sum_{x\in\mathcal{X}} W(y|x)P(x)}. \end{aligned}$$

For a Bernoulli distribution $\mathsf{Bern}\left(p\right)$, we use $H_b\left(p\right)$ to denote its entropy, where $H_b\left(\cdot\right)$ is called *binary entropy function*.

We denote the information divergence [3] between two distributions P and Q as $D\left(P\| Q\right)$, where

$$D\left(P\| Q\right) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}.$$

Also, for the information divergence between two Bernoulli distributions $\mathsf{Bern}\left(p\right)$ and $\mathsf{Bern}\left(q\right)$, we denote $D\left(\mathsf{Bern}\left(p\right)\| \mathsf{Bern}\left(q\right)\right)$ by $D\left(p\| q\right)$ for simplicity.

Similarly, the expectation of the conditional information divergence between $V(\cdot|\cdot)$ and $W(\cdot|\cdot)$ under $P(\cdot)$ is defined as

$$\begin{aligned} D\left(V\| W|P\right) &= \mathbb{E}_P\left[D\left(V\left(\cdot|P\right)\| W\left(\cdot|P\right)\right)\right] \\ &= \sum_{x\in\mathcal{X}} P(x)D\left(V\left(\cdot|x\right)\| W\left(\cdot|x\right)\right). \end{aligned}$$

Remark:

The above quantities can be generalized to continuous random variables by replacing summations with integrals.

■ 2.7 Method of Types

We use the method of types [12, 13] and the standard definitions of type, type set, typical shell, etc., following those in [13]. In addition, we use $\mathcal{P}_n\left(\mathcal{X}\right)$ to denote the set of possible types of sequences in $\mathcal{X}^n$ and adapt the following definition from [12].

[3] Also known as the Kullback-Leibler (KL) divergence.

Definition 2.6 (η-image and image size)**.** The η-image of a set $\mathcal{A} \subset \mathcal{X}^n$ over channel $W : \mathcal{X} \to \mathcal{Y}$ is the collection of sets

$$\mathsf{Img}_W(\mathcal{A}, \eta) \triangleq \{\mathcal{B} \subset \mathcal{Y}^n : W^n(\mathcal{B}| x^n) \geq \eta, \forall x^n \in \mathcal{A}\}. \tag{2.29}$$

We define the size of this image by the size of the element that has the minimum cardinality, i.e.,

$$|\mathsf{Img}_W(\mathcal{A}, \eta)| \triangleq \min_{\mathcal{B} \in \mathsf{Img}_W(\mathcal{A}, \eta)} |\mathcal{B}|. \tag{2.30}$$

We also adopt the "Delta convention" in [12], where the δ_n in typical set $\mathcal{T}^n_{[P]_{\delta_n}}$ is a sequence $\{\delta_n\}_{n=1}^{\infty}$ such that

$$\delta_n \to 0, \quad \sqrt{n}\delta_n \to \infty \quad \text{as } n \to \infty. \tag{2.31}$$

For simplicity, we frequently use $\mathcal{T}^n_{[P]_\delta}$ or $\mathcal{T}^n_{[P]}$ to denote the above sequence of typical sets.

■ 2.8 Important Codebooks

In this thesis, two types of codebooks are frequently used and they play important roles in our analysis.

The first one is the *independent and identically-distributed (i.i.d.) codebook*, where the symbols in each codeword of the codebook are generated independently according to the same probability distribution. We also call the corresponding code *i.i.d. random code*.

Remark:

Strictly speaking, the above codebook generation process corresponds to an ensemble of codebooks rather than one codebook. In this thesis, when we say an i.i.d. codebook achieves certain error probability or error exponent, we are referring to the ensemble average performance of these codebooks under detection/decoding rules that do not take the realizations of the random codebook generation process into account.

In addition, we call a codebook *constant composition codebook* if all codewords in this codebook have the same type. Also, we call the corresponding code *constant composition code*.

Chapter 3

Maximum False Alarm Exponent for DMC

As introduced in Chapter 1, this chapter investigates the codebook detection problem for sparse communication. This is equivalent to the characterization of the *maximum false alarm reliability function* $E_{\mathrm{f}}(R_Q)$, which is formally defined in Chapter 2. In this chapter, we provide a complete characterization of the maximum false alarm reliability function in Section 3.1, and prove the achievability and converse parts in Section 3.2 and Section 3.3 respectively. Our proofs show that an i.i.d. codebook is sufficient to achieve the maximum false alarm error exponent, and there is certain flexibility in decoder design, as discussed in Section 3.2.3. Based on these results, we calculate the maximum false alarm reliability function for various BSC and AWGN channels in Section 3.4. Furthermore, we compare the joint sync–coding approach to training in Section 3.5, and show that for most channel conditions, the joint sync–coding approach achieves significantly better detection performance at high rates, as shown in Figures 3-5 and 3-6. Finally, we discuss the implication of distinguishing codes from noise on the problem of communication via both codeword and timing in Section 3.6.

■ 3.1 Main Results

The main result of this chapter characterizes the maximum false alarm reliability function, as shown in Theorem 3.1.

Theorem 3.1 (Maximum false alarm reliability function)**.** An asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ has maximum false alarm reliability function

$$E_{\mathrm{f}}(R_Q) = \max_{P_X : I(P_X, W) = R_Q} D\left(P_Y \| \, Q_Y\right),$$

where $P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y(\cdot) = [P_X \cdot W]_Y = \sum_x W\left(\cdot|x\right) P_X(x)$.

This theorem indicates that the maximum false alarm exponent is the divergence between the channel output distribution of a codeword and that of the noise sequence. As discussed in

Section 3.2, it is sufficient to use an i.i.d. codebook with distribution P_X such that $I(P_X, W) = R$, though other codebook designs, combined with the proper decoding rule, may achieve the optimal performance as well.

The proof of the above theorem is established via the following key proposition.

Proposition 3.2. For an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$,

$$E_{\mathrm{f}}(R_Q) = \max_{P_X : I(P_X, W) \geq R_Q} D\left(P_Y \| \, Q_Y\right) + I(P_X, W) - R_Q, \tag{3.1}$$

where $P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y(\cdot) = [P_X \cdot W]_Y = \sum_x W(\cdot|x)\, P_X(x)$.

We delay the proof of Proposition 3.2 to Section 3.2 and Section 3.3, and first show that the maximization problem in (3.1) leads to Theorem 3.1.

Proof for Proposition 3.2 implies Theorem 3.1. Let $\mathcal{P}_{R_Q} \triangleq \{P_X : I(P_X, W) \geq R_Q\}$, then since $I(P, W)$ is concave in P for any fixed W, it follows that $\mathcal{P}_{R_Q}$ is a convex set. In addition, it is not hard to check that $\mathcal{P}_{R_Q}$ is compact.

Furthermore, note

$$\begin{aligned} D\left(P_Y \| \, Q_Y\right) + I\left(P_X, W\right) &= \left[-H(P_Y) - \sum_{y \in \mathcal{Y}} P_Y(y) \log Q_Y(y)\right] + \left[H(P_Y) - H(W|P_X)\right] && (3.2) \\ &= -\sum_{y \in \mathcal{Y}} P_Y(y) \log Q_Y(y) - H(W|P_X), && (3.3) \end{aligned}$$

where both $P_Y(\cdot) = \sum_{x \in \mathcal{X}} P_X(x) W(\cdot|x)$ and $H(W|P_X) = \sum_{x \in \mathcal{X}} H(W(\cdot|x)) P_X(x)$ are linear in P_X. Therefore, $D\left(P_Y \| \, Q_Y\right) + I\left(P_X, W\right)$ is linear in P_X and hence convex in P_X.

Then the function $g : \mathcal{P}_{R_Q} \to \mathbb{R}$ such that

$$g(P_X) \triangleq D\left(P_Y \| \, Q_Y\right) + I\left(P_X, W\right) - R_Q \tag{3.4}$$

is convex in P_X and has a convex and compact domain. Hence its maximum over $\mathcal{P}_{R_Q}$ can always be attained at a boundary point of its domain. Therefore,

$$\begin{aligned} \max_{P_X \in \mathcal{P}_{R_Q}} g(P_X) &= \max_{P_X : I(P_X, W) = R_Q} D\left(P_Y \| \, Q_Y\right) + I(P_X, W) - R_Q && (3.5) \\ &= \max_{P_X : I(P_X, W) = R_Q} D\left(P_Y \| \, Q_Y\right). && (3.6) \end{aligned}$$

□

Furthermore, we show a useful property about $E_{\mathrm{f}}(R_Q)$.

Corollary 3.3. $E_{\mathrm{f}}(R_Q)$ is concave in R_Q.

Proof. We show that for any R_1, R_2, $R_\lambda = \lambda R_1 + (1-\lambda)R_2$,

$$E_{\mathrm{f}}(R_\lambda) \geq \lambda E_{\mathrm{f}}(R_1) + (1-\lambda)E_{\mathrm{f}}(R_2). \tag{3.7}$$

Given R_1 and R_2, let P_1 and P_2 satisfy

$$P_1 \in \operatorname*{arg\,max}_{P_X : I(P_X, W) = R_1} D\left(P_Y \| Q_Y\right) = \operatorname*{arg\,max}_{P_X : I(P_X, W) = R_1} g(P_X) \tag{3.8}$$

$$P_2 \in \operatorname*{arg\,max}_{P_X : I(P_X, W) = R_2} D\left(P_Y \| Q_Y\right) = \operatorname*{arg\,max}_{P_X : I(P_X, W) = R_2} g(P_X) \tag{3.9}$$

where $g : \mathcal{P}(\mathcal{X}) \to \mathbb{R}$ is defined in (3.4). Now define $P_\lambda = \lambda P_1 + (1-\lambda)P_2$, then by the property of mutual information,

$$\begin{aligned} I\left(P_\lambda, W\right) &\geq \lambda I\left(P_1, W\right) + (1-\lambda) I\left(P_2, W\right) && (3.10)\\ &= \lambda R_1 + (1-\lambda)R_2 = R_\lambda. && (3.11) \end{aligned}$$

Therefore, $P_\lambda \in \mathcal{P}_{R_\lambda}$ and thus

$$\begin{aligned} E_{\mathrm{f}}(R_\lambda) &\geq g(P_\lambda) && (3.12)\\ &= \lambda g(P_1) + (1-\lambda) g(P_2) && (3.13)\\ &= \lambda E_{\mathrm{f}}(R_1) + (1-\lambda)E_{\mathrm{f}}(R_2), && (3.14) \end{aligned}$$

where (3.13) uses the property that $g(P)$ is linear in P. □

■ 3.2 Achievability

This section shows that an i.i.d. random code with typicality decoder achieves the performance in Proposition 3.2, when the random codes are generated by a distribution P_X such that $I\left(P_X, W\right) \geq R_Q$. Since both detection and decoding are carried out in the single operation of typicality checking, we call this an "one-stage" decoder. This result is summarized in Theorem 3.4.

Theorem 3.4. For an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ and a rate $R_{\mathcal{Q}}$, given a distribution P_X satisfying $I(P_X, W) \geq R_{\mathcal{Q}}$, for any $\lambda > 0$, $\varepsilon \in (0,1)$, there exists a code $\mathcal{C}^{(n)}$ such that

$$\begin{aligned}
P_{\mathrm{e}}^{(n)} &\leq \varepsilon, \\
P_{\mathrm{m}}^{(n)} &\leq \varepsilon, \\
P_{\mathrm{f}}^{(n)} &\leq \exp\left\{-n\left[E_a(P_X, R_{\mathcal{Q}}) - \lambda\right]\right\},
\end{aligned}$$

where

$$E_a(P_X, R_{\mathcal{Q}}) \triangleq D\left(P_Y \| \, Q_Y\right) + I(P_X, W) - R_{\mathcal{Q}}, \tag{3.15}$$

$P_X \in \mathcal{P}(\mathcal{X})$ and $P_Y(\cdot) = [P_X \cdot W]_Y = \sum_x W\left(\cdot|x\right) P_X(x)$.

Furthermore, we demonstrate that the optimal performance can also be achieved by using a two-stage decoder for an i.i.d. codebook, where we first detect whether a codeword is sent based on the empirical channel output distribution, then decode based on the regular channel decoding procedure. However, in this case, we require the codebook be generated by distribution P_X such that $I\left(P_X, W\right) = R_{\mathcal{Q}}$ to achieve the optimal performance, as discussed in Section 3.2.3.

The following two sections provide a proof for Theorem 3.4 by describing an encoding procedure and analyzing the corresponding decoding performance.

■ 3.2.1 Encoding

The standard random codebook generation is used, where we generate $e^{nR_{\mathcal{Q}}}$ codewords[1] randomly according to distribution $P_X(\cdot)$. Specifically, we independently generate codewords $\{x^n(m)$, $m = 1, 2, \cdots, M = e^{nR}\}$ according to the distribution $\prod_{i=1}^{n} P_X(x_i)$.

■ 3.2.2 Typicality (one-stage) decoding

We prove Theorem 3.4 by analyzing the performance of typicality decoding. The intuition behind using a typicality decoder is relatively straightforward: since we only require the miss and decoding error probability to vanish asymptotically, we can simply use a regular (synchronous) channel code, and take the union of all the typical shell of each codeword as the detection region $\mathcal{A}_n$, which achieves false alarm probability $P_{\mathrm{f}} = Q_Y(\mathcal{A}_n)$.

[1] Strictly speaking, we should use $\lceil e^{nR_{\mathcal{Q}}} \rceil$ rather than $e^{nR_{\mathcal{Q}}}$. But When n is large, the difference between the two is inconsequential, and we use $e^{nR_{\mathcal{Q}}}$ to simplify notation.

More formally, the typicality decoder declares a message m if there exist only one m such that $(x^n(m), y^n) \in \mathcal{T}^n_{[P_{XY}]_\delta}$, where $P_{XY} = P_X \cdot W$. If there are more than one codeword that is jointly typical with the received vector y^n, it declares a decoding error, otherwise the noise sequence $\star^n$. Therefore, the acceptance region is

$$\mathcal{A}_n = \bigsqcup \mathcal{D}_m, \tag{3.16}$$

where $\mathcal{D}_m \subset \mathcal{Y}^n$ is the typicality decoding region for message m, i.e., $\mathcal{D}_m = g_n^{-1}(m)$. This leads to the following the average (over codebook $\mathcal{C}$ and message m) error probability expressions:

$$\begin{aligned}
\overline{P_{\mathrm{e}}} &= \overline{P_{\mathrm{m}}} && (3.17)\\
&= \sum_{\mathcal{C}} \mathbb{P}\left[\mathcal{C}\right] \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\left[(x^n(m), W^n(\cdot|x^n(m))) \notin \mathcal{T}^n_{[P_{XY}]_\delta} | \mathcal{C}\right] && (3.18)\\
\overline{P_{\mathrm{f}}} &= \sum_{\mathcal{C}} \mathbb{P}\left[\mathcal{C}\right] \sum_{m=1}^{M} \mathbb{P}\left[(x^n(m), W^n(\cdot|\star^n)) \in \mathcal{T}^n_{[P_{XY}]_\delta} | \mathcal{C}\right]. && (3.19)
\end{aligned}$$

By the “symmetry” of the codebook,

$$\begin{aligned}
\overline{P_{\mathrm{m}}} &= \frac{1}{M} \sum_{m=1}^{M} \sum_{\mathcal{C}} \mathbb{P}\left[\mathcal{C}\right] \mathbb{P}\left[(x^n(m), W^n(\cdot|x^n(m))) \notin \mathcal{T}^n_{[P_{XY}]_\delta} | \mathcal{C}\right] && (3.20)\\
&= \mathbb{P}\left[(X^n(m), W^n(\cdot|X^n(m))) \in \mathcal{T}^n_{[P_{XY}]_\delta}, X^n(m) \overset{i.i.d.}{\sim} P_X\right]. && (3.21)
\end{aligned}$$

Based on the properties of strongly typical sets, when the δ in $\mathcal{T}^n_{[P_{XY}]_\delta}$ satisfies the “Delta convention” and n sufficiently large,

$$\overline{P_{\mathrm{e}}} = \overline{P_{\mathrm{m}}} \leq \varepsilon. \tag{3.22}$$

Similarly,

$$\begin{aligned}
\overline{P_{\mathrm{f}}} &= \sum_{m=1}^{M} \sum_{\mathcal{C}} \mathbb{P}\left[\mathcal{C}\right] \mathbb{P}\left[(x^n(m), W^n(\cdot|\star^n)) \in \mathcal{T}^n_{[P_{XY}]_\delta} | \mathcal{C}\right] && (3.23)\\
&= M \cdot \mathbb{P}\left[(X^n(m), W^n(\cdot|\star^n)) \in \mathcal{T}^n_{[P_{XY}]_\delta}, X^n(m) \overset{i.i.d.}{\sim} P_X\right]. && (3.24)
\end{aligned}$$

Here $(X^n(m), W^n(\cdot|\star^n)) \overset{i.i.d.}{\sim} [P_X \cdot Q_Y]$, and hence

$$\begin{aligned}
\overline{P_{\mathrm{f}}} &= e^{nR_{\mathcal{Q}}} [P_X \cdot Q_Y]^n (\mathcal{T}^n_{[P_{XY}]_\delta}) && (3.25)\\
&\doteq e^{nR_{\mathcal{Q}}} |\mathcal{T}^n_{[P_{XY}]_\delta}| \exp\left[-n(H(P_{XY}) + D\left(P_{XY} \| \left[P_X \cdot Q_Y\right]\right))\right] && \text{(Lemma A.2)} \quad (3.26)\\
&\doteq \exp\left\{-n\left[I(P_X; P_Y) + D\left(P_Y \| Q_Y\right) - R_{\mathcal{Q}}\right]\right\}, && \text{(Fact A.1)} \quad (3.27)
\end{aligned}$$

where we also use the standard fact that $|\mathcal{T}^n_{[P_{XY}]_\delta}| \doteq \exp\left[n\left(H(P_{XY})\right)\right]$ (Lemma 1.2.13 in [12]).

Hence, there exists at least one reliable sequence of codes $\mathcal{Q}$ with rate $R_{\mathcal{Q}}$ such that

$$P_{\mathrm{f}}^{(n)} \dot{\leq} \exp\left\{-n\left[I(P_X;P_Y) + D\left(P_Y \| Q_Y\right) - R_{\mathcal{Q}}\right]\right\}. \tag{3.28}$$

Therefore, for any $\lambda > 0$, when n sufficiently large,

$$P_{\mathrm{f}}^{(n)} \leq \exp\left\{-n\left[I(P_X;P_Y) + D\left(P_Y \| Q_Y\right) - R_{\mathcal{Q}} - \lambda\right]\right\}, \tag{3.29}$$

which proves Theorem 3.4.

■ 3.2.3 Two-stage decoding

We now investigate a decoder with two stages of operations. In the first stage the decoder detects the presence of a codeword based on the channel output distribution. If a codeword is detected, then it is decoded via the regular channel decoding procedure in the second stage. We show that two-stage decoding is optimal, which is somewhat surprising as the first detection stage does not take the codebook structure into account.

Note that the channel output distribution is either P_Y (when we use the i.i.d. codebook with input distribution P_X and thus $P_Y = [P_X \cdot W]_Y$) or Q_Y (when the $\star$ symbol is the channel input). Therefore we have a simple binary hypothesis testing problem between distribution P_Y and Q_Y. Stein's Lemma [14] indicates, for the detection step, given any input distribution P_X,

$$\overline{P_{\mathrm{f}}} \doteq \exp\left[-nD\left(P_Y \| Q_Y\right)\right]. \tag{3.30}$$

And to achieve rate R, we need $I\left(P_X, W\right) \geq R_{\mathcal{Q}}$. Therefore, by choosing a proper P_X and following the standard random coding argument, we can achieve the following false alarm error exponent at rate $R_{\mathcal{Q}}$,

$$E_{\mathrm{f}2}(R_{\mathcal{Q}}) = \max_{P_X : I(P_X,W) \geq R_{\mathcal{Q}}} D\left(P_Y \| Q_Y\right) \tag{3.31}$$

$$= \max_{P_X : I(P_X,W) = R_{\mathcal{Q}}} D\left(P_Y \| Q_Y\right) \tag{3.32}$$

Comparison of one-stage and two-stage decoding

The above analysis shows the false alarm reliability functions for both one-stage and two-stage decoding satisfy

$$E_{\mathrm{f1}}(R_Q) = E_{\mathrm{f2}}(R_Q) = \max_{P_X : I(P_X, W) = R_Q} D\left(P_Y \| \, Q_Y\right),$$

so they achieve the same optimal performance. However, because the two-stage decoder does not take the code rate R_Q into account, it imposes a stronger requirement on the codebook design—in general, it requires a codebook with input distribution P_X such that $I\left(P_X, W\right) = R$, while for the typicality decoder, the optimal performance can be achieved even when $I\left(P_X, W\right) > R$.

In other words, for a *given* P_X such that $I(P_X, W) > R_Q$, the one-stage scheme, in general, achieves a higher E_{f} than the two-stage scheme, as illustrated in Figure 3-1, where

$$\begin{aligned} E_{\mathrm{f1}}(R_Q, P_X) &\triangleq D\left(P_Y \| \, Q_Y\right) + I(P_X, W) - R_Q \\ &\in \left(D\left(P_Y \| \, Q_Y\right), E_{\mathrm{f}}(R_Q)\right] \\ E_{\mathrm{f2}}(R_Q, P_X) &\triangleq D\left(P_Y \| \, Q_Y\right) \\ &= E_{\mathrm{f}}(I\left(P_X, W\right)) \end{aligned}$$

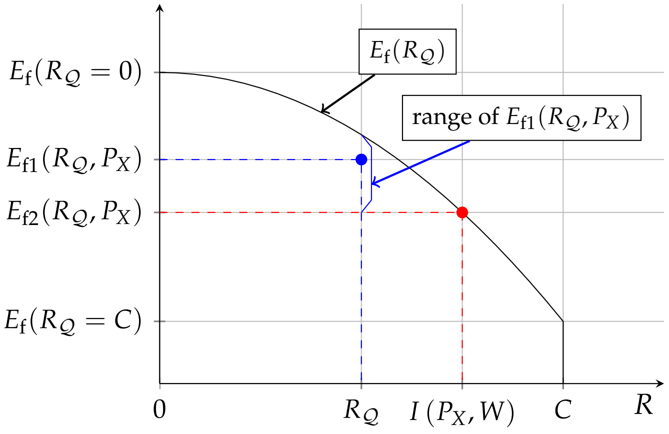


Figure 3-1: For a given P_X, one-stage decoding generally achieves a larger E_{f} than two-stage decoding at a given rate R_Q.

■ 3.3 Converse

In this section we derives an upper bound for $E_\mathrm{f}(R_\mathcal{Q})$, which agrees with the achievable $E_\mathrm{f}(R_\mathcal{Q})$ in Proposition 3.2. This completes the converse part of Theorem 3.1.

We start with a "sphere-packing"-like upper bound for constant composition codes, and later connect the result to general channel codes. The proof shows that, due to the requirement that P_e and P_m must be vanishing asymptotically, the codeword acceptance region $\mathcal{A}_n$ essentially needs to include the typical shells of all codewords, and hence cannot be too small. This leads to a lower bound for $P_\mathrm{f} = Q_Y(\mathcal{A}_n)$ and thus an upper bound for $E_\mathrm{f}(R_\mathcal{Q})$.

Lemma 3.5 (Upper bound for $E_\mathrm{f}(R_\mathcal{Q})$, constant composition codes)**.** Given an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$, for every $R > 0$, $\varepsilon > 0$, $\lambda > 0$, any constant composition code (f_n, g_n) with type P_X[2], rate

$$\frac{1}{n} \log \left|\mathcal{M}_{f_n}\right| \geq R - \lambda \tag{3.33}$$

and maximal probability of error

$$P_\mathrm{e}^{(n)} \leq \varepsilon \tag{3.34}$$

satisfies

$$P_\mathrm{f}^{(n)} \geq \mathrm{poly}(n) \exp\left[-n(E_a(P_X, R) + 2\lambda)\right], \tag{3.35}$$

when n sufficiently large and $E_a(\cdot,\cdot)$ is defined in (3.15).

Proof. Since $P_\mathrm{e}^{(n)} \leq \varepsilon$, when n sufficiently large, for all m,

$$W^n\left(\mathcal{D}_m \middle| x^n(m)\right) \geq 1 - \varepsilon. \tag{3.36}$$

In addition, when n sufficiently large,

$$W^n\left(\mathcal{T}^n_{[W]_\delta}(x^n(m)) \middle| x^n(m)\right) \geq 1 - \varepsilon. \tag{3.37}$$

[2] Strictly speaking, the type P_X depends on the block length n. However, for any distribution $P \in \mathcal{P}(\mathcal{X})$ and any $\varepsilon > 0$, we can find a type $P' \in \mathcal{P}_n(\mathcal{X})$ such that $\|P - P'\|_\infty < \varepsilon$ when n sufficiently large, and it is not hard to see that this dependence on n is inconsequential. Therefore, we avoid this technicality and simply treat P_X as a common type that can be achieved by a sequence of constant composition codebooks index by n, when n sufficiently large. This treatment is implicitly used in Theorem 1.5.3 of [12] as well and we follow this convention in the rest of this thesis.

Therefore, when n sufficiently large,

$$W^n\left(\mathcal{D}_m \cap \mathcal{T}^n_{[W]_\delta}(x^n(m))\middle|\, x^n(m)\right) \geq 1-2\varepsilon. \tag{3.38}$$

Let $\mathcal{F}_m \triangleq \mathcal{D}_m \cap \mathcal{T}^n_{[W]_\delta}(x^n(m))$, which are disjoint since $\mathcal{D}_m$ are disjoint. When n sufficiently large,

$$\frac{1}{n}\log|\mathcal{F}_m| \geq H(W|P_X) - \lambda/2. \tag{3.39}$$

And since $x^n(m) \in \mathcal{T}_{P_X}$,

$$y^n \in \mathcal{T}^n_{[W]_\delta}(x^n(m)) \quad \Rightarrow \quad y^n \in \mathcal{T}^n_{[P_Y]_{\delta'}}, \tag{3.40}$$

where $P_Y = [P_X \cdot W]_Y$ and both δ and δ' can be made to satisfy the "Delta Convention". Hence for any $y^n \in \mathcal{F}_m$, when n sufficiently large,

$$D\left(P_Y \| \, Q_Y\right) + H(P_Y) - \lambda/2 \leq -\frac{1}{n}\log Q_Y^n(y^n) \leq D\left(P_Y \| \, Q_Y\right) + H(P_Y) + \lambda/2. \tag{3.41}$$

Therefore,

$$\begin{aligned}
P_{\mathrm{f}}^{(n)} &= Q_Y(\mathcal{A}_n) && (3.42)\\
&\geq Q_Y\left(\bigcup_{m=1}^{|\mathcal{M}_{f_n}|} \mathcal{F}_m\right) = \sum_{m=1}^{|\mathcal{M}_{f_n}|} Q_Y(\mathcal{F}_m) = \sum_{m=1}^{|\mathcal{M}_{f_n}|}\sum_{y^n\in\mathcal{F}_m} Q_Y(y^n) && (3.43)\\
&\geq \mathrm{poly}(n) e^{n[R-\lambda]} e^{n[H(W|P_X)-\lambda/2]} e^{-n[D(P_Y\|Q_Y)+H(P_Y)+\lambda/2]} && (3.44)\\
&= \mathrm{poly}(n)\exp\left\{-n\left[D\left(P_Y\|\,Q_Y\right) + I(P_X, W) - R + 2\lambda\right]\right\} && (3.45)\\
&= \mathrm{poly}(n)\exp\left[-n(E_a(P_X, R) + 2\lambda)\right] && (3.46)
\end{aligned}$$

□

Now we connect the above result back to general channel codes via the following simple yet important fact about channel codes.

Lemma 3.6 (Every channel code contains a constant composition subcode of the same rate)**.** For any $\lambda > 0$, any code (f_n, g_n) satisfying

$$\left|\mathcal{M}_{f_n}\right| \geq e^{n(R-\lambda)} \tag{3.47}$$

has a constant composition subcode $(\hat{f}_n, \hat{g}_n)$ with

$$\left|\mathcal{M}_{\hat{f}_n}\right| \geq e^{n(R-2\lambda)}, \tag{3.48}$$

when n is sufficiently large.

Proof. Use the fact that the number of different types of sequences in $\mathcal{X}^n$ is polynomially many (Type Counting Lemma), or see P.171 in [12]. □

Now we are ready to prove the converse for general channel codes.

Theorem 3.7 (Upper bound for $E_{\mathrm{f}}(R_Q)$, general channel codes)**.** Given an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$, for every $R > 0$, $\varepsilon \in (0,1)$, $\lambda > 0$, every code $\mathcal{C} = (f_n, g_n)$ with rate

$$\frac{1}{n}\log\left|\mathcal{M}_{f_n}\right| \geq R - \lambda \tag{3.49}$$

and maximal probability of error

$$P_{\mathrm{e}}^{(n)} \leq \varepsilon \tag{3.50}$$

satisfies

$$P_{\mathrm{f}}(\mathcal{C}) \geq \mathrm{poly}(n)\exp\left[-n(E_b(R,\lambda) + 4\lambda)\right], \tag{3.51}$$

when n sufficiently large, where

$$E_b(R,\lambda) \triangleq \max_{P_X : I(P_X, W) \geq R - 3\lambda} E_a(P_X, R). \tag{3.52}$$

Proof. Lemma 3.6 shows that there exists a constant composition subcode $\hat{\mathcal{C}} = (\hat{f}_n, \hat{g}_n)$ with type P_X and rate

$$\frac{1}{n}\log\left|\mathcal{M}_{\hat{f}_n}\right| \geq R - 2\lambda. \tag{3.53}$$

Given $P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right) \leq \varepsilon$, it is not hard to show that there exists a $\hat{\mathcal{C}}^{(n)}$ such that $P_{\mathrm{e}}\left(\hat{\mathcal{C}}^{(n)}\right) \leq \varepsilon$ and $P_{\mathrm{f}}\left(\hat{\mathcal{C}}^{(n)}\right) \leq P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right)$. Then by Corollary 2.1.4 in [12], $P_{\mathrm{e}}\left(\hat{\mathcal{C}}^{(n)}\right) \leq \varepsilon$ indicates for any $\tau > 0$,

$$\frac{1}{n}\log\left|\mathcal{M}_{\hat{f}_n}\right| \leq I(P_X, W) + 2\tau, \tag{3.54}$$

when n sufficiently large. Let $\tau = \lambda/2$, then $I(P_X, W) \geq R - 3\lambda$.

From Lemma 3.5,

$$\begin{aligned} P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) &\geq P_{\mathrm{f}}(\hat{\mathcal{C}}^{(n)}) \\ &\geq \mathrm{poly}(n) \exp\left[-n(E_a(P_X, R) + 4\lambda)\right] \\ &\geq \mathrm{poly}(n) \exp\left[-n(E_b(R, \lambda) + 4\lambda)\right]. \end{aligned}$$

□

Combing the achievability and converse results, we can now prove Proposition 3.2.

Proof for Proposition 3.2. Pick a sequence of positive numbers $\{\lambda_n\}$ such that $\lambda_n \to 0$ as $n \to \infty$. Let

$$P_X^* \in \arg\max_{P_X} I(P_X; P_Y) + D\left(P_Y \| Q_Y\right) - R_Q \tag{3.55}$$

and apply Theorem 3.4, then

$$\begin{aligned} E_{\mathrm{f}}(R_Q) &\geq \liminf_{n\to\infty} -\frac{1}{n} \log P_{\mathrm{f}}^{(n)} && (3.56) \\ &\geq \lim_{n\to\infty} E_a(P_X^*, R_Q) - \lambda_n && (3.57) \\ &= E_a(P_X^*, R_Q). && (3.58) \end{aligned}$$

Then apply Theorem 3.7, we have

$$\begin{aligned} E_{\mathrm{f}}(R_Q) &\leq \liminf_{n\to\infty} -\frac{1}{n} \log P_{\mathrm{f}}^{(n)} && (3.59) \\ &\leq \liminf_{n\to\infty} \max_{P_X : I(P_X, W) \geq R_Q - 3\lambda_n} \left[E_a(P_X, R_Q) - 4\lambda_n\right] && (3.60) \\ &= \max_{P_X : I(P_X, W) \geq R_Q} E_a(P_X, R_Q). && (3.61) \end{aligned}$$

Therefore,

$$E_{\mathrm{f}}(R_Q) = \max_{P_X : I(P_X, W) \geq R_Q} E_a(P_X, R_Q) = \max_{P_X : I(P_X, W) \geq R_Q} D\left(P_Y \| Q_Y\right) + I(P_X, W) - R_Q \tag{3.62}$$

□

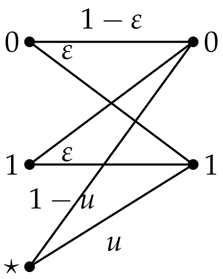


Figure 3-2: An asynchronous BSC with crossover probability ε and $\star$ output distribution $\mathsf{Bern}\,(u)$.

3.4 Examples

In this section, we specialize our results about $E_{\mathrm{f}}(R_Q)$ to both the AWGN channel[3] and the BSC to calculate the maximum false alarm reliability functions.

3.4.1 AWGN channel

For an asynchronous AWGN channel (defined in Section 2.4.2) with average codeword power constraint P, and noise $Z \sim \mathsf{N}(0,1)$, we use an i.i.d. non-zero mean Gaussian codebook with input distribution $P_X \sim \mathsf{N}(\mu, \sigma^2)$, where $\mu^2 + \sigma^2 = P$.

Let $k \triangleq \frac{\sigma^2}{P} \in [0,1]$, then

$$D\left(P_Y \| \, Q_Y\right) = \frac{1}{2}\Big[\mathsf{SNR} - \log(1 + k\mathsf{SNR})\Big] \tag{3.63}$$

$$I\left(P_X, W\right) = \frac{1}{2}\log(1 + k\mathsf{SNR}), \tag{3.64}$$

where $\mathsf{SNR} = P$. Hence

$$E_{\mathrm{f}}(R) = \mathsf{SNR}/2 - R, \tag{3.65}$$

which decreases linearly with R. Some sample $E_{\mathrm{f}}(R)$ plots are shown in Figure 3-5 and discussed later.

3.4.2 Binary symmetric channel

Recall that we define an asynchronous BSC in Section 2.4.1 via its crossover probability ε and $\star$ output distribution $\mathsf{Bern}\,(u)$ (Figure 3-2), where without loss of generality, we assume $\varepsilon < 1/2$ and $u \leq 1/2$.

[3] Strictly speaking, an AWGN channel does not have a discrete alphabet and hence is not a DMC, but results in this chapter can be extended to AWGN channels via standard arguments.

For a BSC, let the input distribution be $P_X \sim \text{Bern}\,(p)$, then the output distribution is $P_Y \sim \text{Bern}\,(s)$, where $s = p * \varepsilon$. We have

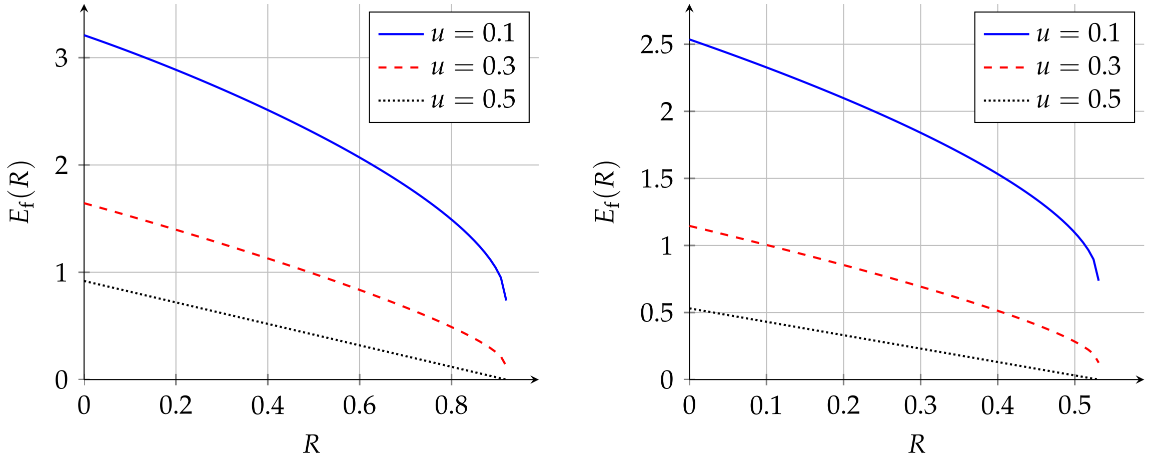
$$I\,(P_X, W) = H_b\,(s) - H_b\,(\varepsilon) \tag{3.66}$$

$$D\,(P_Y \| \, Q_Y) = D\,(s \| \, u) = s \log \frac{s}{u} + (1-s) \log \frac{1-s}{1-u}. \tag{3.67}$$

Given a rate R, choose $P_X^* \sim \text{Bern}\,(p^*)$ such that $I\,(P_X^*, W) = R$, where $p^* \geq 1/2$. Let $s^* = p^* * \varepsilon$, then $E_{\mathrm{f}}(R) = D\,(s^* \| \, u)$.

Figure 3-3 shows the $E_{\mathrm{f}}(R)$ of two BSCs under various u values. As we can see, at all rates below capacity, we achieve positive false alarm exponents. And even at capacity, the false alarm probability does not decrease exponentially only when $Q_Y\,(\cdot)$ is the same as the capacity-achieving distribution $\text{Bern}\left(\frac{1}{2}\right)$. In this case, the output distributions of the codebook and the noise sequence are identical and hence they "look the same" in the sense of empirical distribution, and no positive error exponent can be achieved.

In addition, it is obvious that the optimal false alarm exponent decreases as the rate R increases or the channel cross-over probability increases. Therefore, to achieve the same false alarm probability requirement, we need to use a longer code at higher rates or at higher channel crossover probabilities.



(a) $\varepsilon = 0.01$

(b) $\varepsilon = 0.1$

Figure 3-3: Maximum false alarm reliability function $E_{\mathrm{f}}(R)$ for BSCs with various crossover probability ε and $\star$ output distribution $\text{Bern}\,(u)$.

■ 3.5 Comparison with Separate Synchronization and Coding

In the separate synchronization–coding (training) approach, to transmit nR bits of information in n channel uses, we first use $(1-\frac{R}{C})n$ symbols for synchronization and then use a capacity-achieving code with block length $\frac{R}{C}n$, where the symbols x_c in the sync word satisfies

$$x_c = \arg\max_{x\in\mathcal{X}} D\left(W\left(\cdot|x\right)\| \, W\left(\cdot|\star\right)\right). \tag{3.68}$$

This is illustrated in Figure 3-4.

When the detection is based on the sync word only, the maximum achievable false alarm error exponent is

$$E_{\mathrm{t}}(R) = \left(1-\frac{R}{C}\right) D\left(W\left(\cdot|x_c\right)\| \, W\left(\cdot|\star\right)\right), \tag{3.69}$$

and it is strictly not better than the joint sync–coding approach, which is natural since the separate sync–coding approach can be viewed as a special case of the joint sync–coding approach.

More Specifically, we show that joint sync–coding is strictly better than separate sync–coding under a broad set of conditions.

Theorem 3.8 (Separated sync–coding is suboptimal)**.** For an asynchronous DMC $(\mathcal{X},\star,W,\mathcal{Y})$, if there exists a capacity-achieving distribution P_X^* such that $D\left(Q_Y\| \, P_Y^*\right) > 0$, then

$$E_{\mathrm{t}}(R) < E_{\mathrm{f}}(R)$$

for all $R > 0$, where $P_Y^* = [P_X^* \cdot W]_Y$ is the output distribution corresponding to the capacity-achieving input distribution.

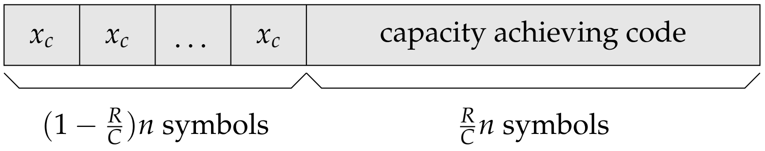


Figure 3-4: Separate sync–coding, where the first $(1-\frac{R}{C})n$ symbols are used for synchronization, and the next $\frac{R}{C}n$ symbols are coded as a capacity-achieving code for information transmission.

Proof. Note

$$E_{\mathrm{f}}(R=0) = E_{\mathrm{t}}(R=0),$$

$$E_{\mathrm{f}}(R=C) \geq D\left(Q_Y \| P_Y^*\right) > 0 = E_{\mathrm{t}}(R=C).$$

The statement now follows from the concavity of $E_{\mathrm{f}}(R)$ (Corollary 3.3). □

Figures 3-5 and 3-6 show the differences between the $E_{\mathrm{t}}(R)$ curves (separate approach) and $E_{\mathrm{f}}(R)$ curves (joint approach) on various AWGN and binary symmetric channels, with rate R in the unit of bits. The plots show that, in general, the loss of the separate sync–coding approach is more significant at high rates, because in this regime, the separate sync–coding approach needs to use most of the degrees of freedom for information transmission, rather than synchronization (detection), resulting poor performance.

Remark:

Note that under certain conditions, it is possible for separate sync–coding to achieve $E_{\mathrm{f}}(R)$ and thus to be optimal. For example, for the AWGN channel, separate sync–coding achieves $E_{\mathrm{f}}(R) = 0.5 - R$ (only) when SNR $= 1$, and for the BSC, separate sync–coding achieves $E_{\mathrm{f}}(R)$ (only) when $u = 1/2$. However, this is mainly due to our lenient requirement on P_{m}, which is merely $P_{\mathrm{m}} \to 0$. Chapter 4 shows that once we impose a stricter requirement on the miss error probability P_{m} by constraining its error exponent, separate sync–coding is suboptimal even under the above channel conditions.

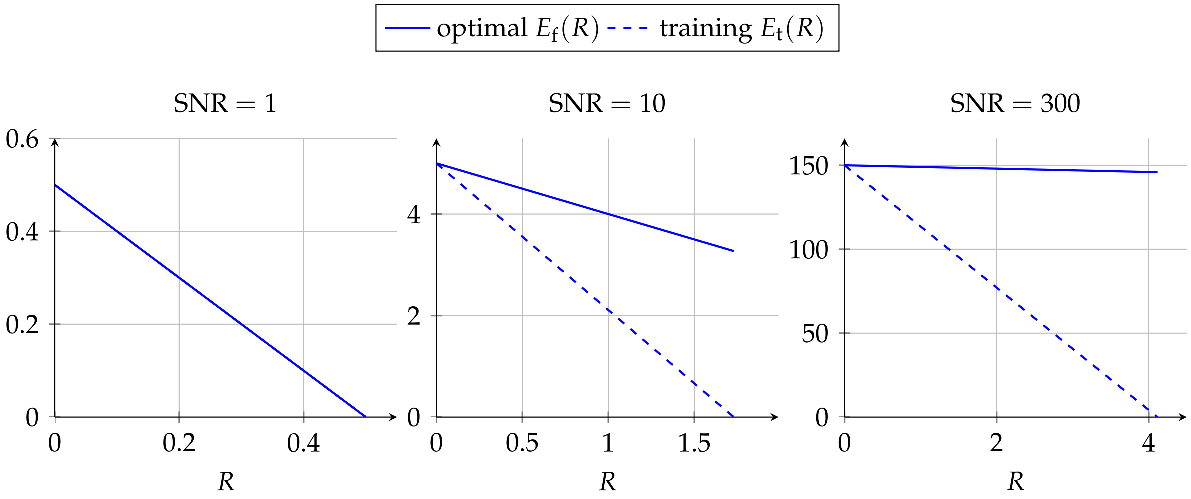


Figure 3-5: Maximum achievable false alarm exponent comparison of joint sync–coding and separate sync–coding (training) for AWGN channels with different SNRs.

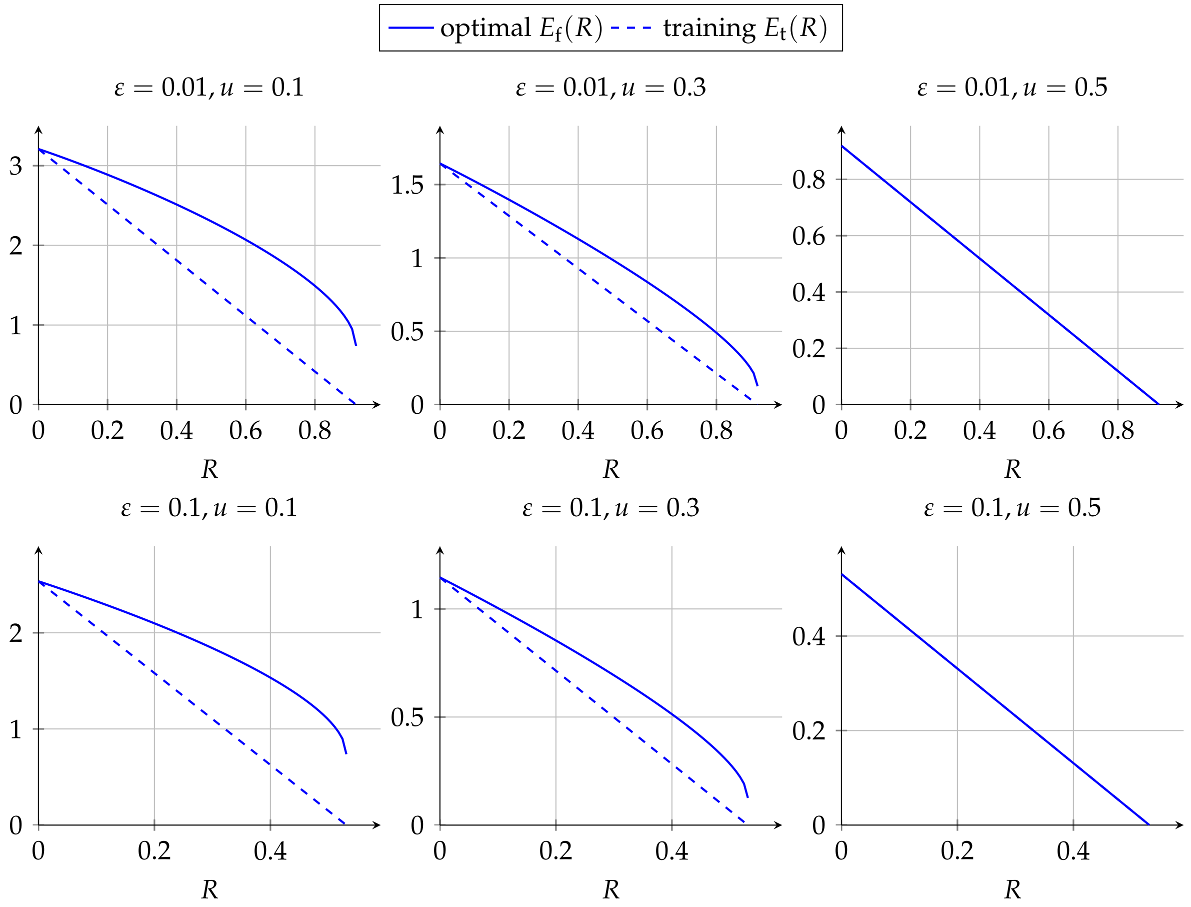


Figure 3-6: Maximum achievable false alarm exponent comparison of joint sync–coding and separate sync–coding (training) for BSCs with $\varepsilon = 0.01$, $\varepsilon = 0.1$ and different noise output distributions Bern (u).

■ 3.6 A Discussion on Communication by Timing

In this section we diverge from the problem of synchronization and discuss another implication of distinguishing codes from noise.

Suppose a transmitter and a receiver can establish a time window with L time slots, and they communicate by both the content of a codeword and the slot location (timing) of the codeword. This can be viewed as an extension of the pulse position modulation (PPM) [15]. In PPM, the pulse itself does not carry any information and all information is conveyed through the location of the pulse, while in this scheme, the "pulse" is a codeword, which carries certain information itself.

If the transmitter uses a code with rate R, then our analysis on maximum false alarm exponent indicates that $L \overset{\cdot}{\leq} \exp(nE_f(R))$, otherwise the total false alarm probability $(L-1)P_f$ is unbounded as n increases. Hence, the maximum number of bits we can communicate by both codeword and timing is

$$N(R) = nR + \max \log L = n(R + E_{\mathrm{f}}(R)). \tag{3.70}$$

Then by choosing a proper rate R^* that maximizes $N(R)$, we can obtain the optimal operating window length $L^* = \exp(nE_{\mathrm{f}}(R^*))$.

For the AWGN channel,

$$N_{\mathrm{AWGN}}(R) = n(R + E_{\mathrm{f}}(R)) = n\mathsf{SNR}/2. \tag{3.71}$$

Therefore, this communication scheme achieves the same throughput regardless of the rate or the window size we choose. Hence, it is sufficient to simply use a code with $R = 0$, which reduces the scheme to PPM.

For the BSC,

$$N_{\mathrm{BSC}}(R) = n(R + E_{\mathrm{f}}(R)) = n\left[H_b\left(s\right) - H_b\left(\varepsilon\right) + D\left(s\|\, u\right)\right], \tag{3.72}$$

where s satisfies $H_b\left(s\right) - H_b\left(\varepsilon\right) = R$. Assuming $u \leq 1/2$, then $1/2 \leq s \leq 1 - \varepsilon$. Thus

$$\begin{aligned} H_b\left(s\right) - D\left(s\|\, u\right) &= -s \log u - (1-s)\log(1-u) &(3.73)\\ &= s \log \frac{1-u}{u} - \log(1-u), &(3.74)\end{aligned}$$

which increases as s increases. Therefore $N_{\mathrm{BSC}}(R)$ is maximized at $s = 1 - \varepsilon$, which corresponds to $R = 0$. This indicates for this specific scheme that communicates via both codeword and timing, coding does not help, and sometimes it may reduce the throughput. Figure 3-7 shows an example that as the code rate R increases, the throughput decreases, because the reduction in $E_{\mathrm{f}}(R)$ outweighs the increase in R.

■ 3.7 Summary

The results in this chapter indicate that for sparse communication, it is beneficial to use the joint sync–coding scheme instead of training, especially in the high rate regime. In addition, the tradeoff between rate and maximum false alarm exponent indicates that, when we transmit at a rate R that is below capacity C, it is beneficial to use a code that just achieves rate R rather than that

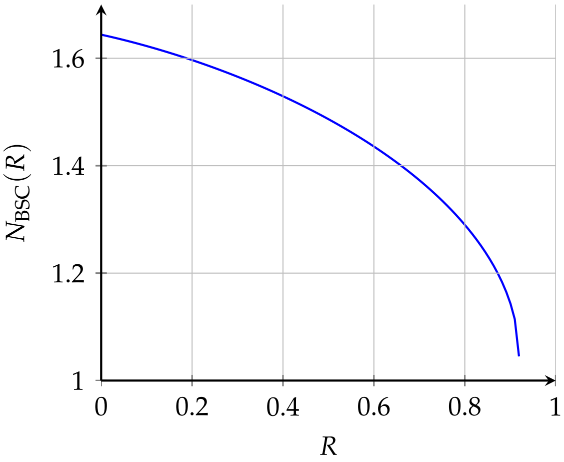


Figure 3-7: Throughput $N_{\text{BSC}}(R)$ for a BSC with $\varepsilon = 0.01$ and $u = 0.3$.

achieves the capacity C. By backing off from capacity, we can use a smaller acceptance region for the code, leading to a lower false alarm probability in detection.

Chapter 4

Optimal False Alarm Exponent with Miss Exponent Constraint: DMC Channels

In this chapter we investigate the codebook detection problem subject to a miss exponent constraint. By choosing different E_{m} values, the user can impose different requirements on the miss error probability. This problem is more general than the problem in Chapter 3, and leads to an investigation of the *false alarm reliability function* $E_{\mathrm{f}}(R_{\mathcal{Q}}, E_{\mathrm{m}})$, which has been defined in Chapter 2.

In addition, as mentioned in Chapter 2, finding the false alarm reliability function is equivalent to finding the achievable error exponent region $\mathcal{E}(R_{\mathcal{Q}})$ with $e_{\mathrm{d}} = 0$, and for certain analysis in this chapter, we adopt the error exponent region description, if it is more natural for the specific analysis[1] . Also note that this characterization about $\mathcal{E}(R_{\mathcal{Q}})$ leads to a characterization for $E_{\mathrm{m}}(R_{\mathcal{Q}})$ when we constrain $e_{\mathrm{f}} = 0$.

In our investigation, we first develop and analyze achievability schemes for the DMC based on the i.i.d. random codebook and the constant composition codebook in Section 4.1, which gives us several lower bounds to the false alarm reliability function. In addition, the performance of training is analyzed in Section 4.3. All these results are later specialized to BSC, and the comparisons in Section 4.4.4 demonstrate that the constant composition codebook achieves better performance than the i.i.d. codebook and training. Besides inner bounds, we establish a connection between the upper bound to the false alarm reliability function and the minimum output image size of a codebook for the DMC in Section 4.2. Then in Section 4.4.3, we apply this result to the BSC via an entropy inequality over the binary channel, and provide an upper bound to the false alarm reliability function that is asymptotically tight at low rate.

■ 4.1 Achievability schemes for DMC

■ 4.1.1 i.i.d. codebook with optimal detection

The optimal performance of an i.i.d. codebook is simple to analyze, based on the following result.

[1] In this case, the lower bound and upper bound for the false alarm reliability function become the *inner bound* and *outer bound* of the error exponent region.

Theorem 4.1. Given an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ W and a rate R_Q, a (e_m, e_f) pair is achievable by an i.i.d. codebook with distribution P_X if there exists a $P_\lambda = \zeta P_Y^\lambda Q_Y^{1-\lambda}$ with $0 \leq \lambda \leq 1$ and ζ being a normalization constant such that

$$e_{\mathrm{m}} < D\left(P_\lambda \| P_Y\right) \quad \text{and} \quad e_{\mathrm{f}} < D\left(P_\lambda \| Q_Y\right), \tag{4.1}$$

where $I\left(P_X, W\right) \geq R_Q$ and $P_Y = [P_X \cdot W]$.

Proof. The result follows directly from the properties of Neyman-Pearson test, Sanov's theorem (Section 11.7 and 11.9 in [14]) and the standard random coding argument. □

Remark:
The above result extends to channels with continuous alphabets as Sanov's theorem can be "extended to continuous distributions using quantization" (P.388 of [14]).

However, unlike the case in Chapter 3, i.i.d. codebooks are in general suboptimal in terms of false alarm error exponents when there is a miss error exponent constraint. A miss error exponent constraint implies a stronger requirement on the typicality of the codebook, because, although the atypical codewords produced during the i.i.d. codebook generation process only occupy an exponentially small fraction of the codebook, it has an impact on the miss error exponent. This motivates us to investigate a type of codebook that eliminates the atypicality in the codebook generation process, the constant composition codebook.

■ 4.1.2 Constant composition codebook

Since atypicality is eliminated in a constant composition codebook, we focus on the atypicality of the channel, and obtain the following achievable performance based on an argument that partitions the channel realizations $\{V\}$ into different types.

Theorem 4.2 (Achievability, constant composition codebook)**.** For an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$, given a rate R_Q and a miss error exponent constraint E_{m}, the following lower bound for the false alarm reliability function is achievable via a sequence of constant composition codebooks

$$\underline{E_{\mathrm{f}}}(R_Q, E_{\mathrm{m}}) = \max_{P_X : I(P_X, W) \geq R_Q} \min_{V : D(V \| W | P_X) \leq E_{\mathrm{m}}} \left[D\left(Q_V \| Q_Y\right) + \left\{ I\left(P_X, V\right) - R_Q \right\}^+ \right]. \tag{4.2}$$

Similarly, given a rate R_Q and a false alarm error exponent constraint E_{f}, the following lower bound for the miss reliability function is achievable via a sequence of constant composition codebooks

$$\underline{E_{\mathrm{m}}}(R_Q, E_{\mathrm{f}}) = \max_{P_X : I(P_X, W) \geq R_Q} \min_{V : D(Q_V \| Q_Y) \leq E_{\mathrm{f}}} D\left(V \| W | P_X\right). \tag{4.3}$$

Proof. We first prove the result regarding the lower bound for the false alarm reliability function.

Given a rate R_Q and a type P_X such that $I\left(P_X, W\right) \geq R_Q$, from the channel coding theorem (or see Corollary 2.1.3 and Exercise 2.1.17 in [12]), there exists a codebook $\mathcal{C}^{(n)}$ with type P_X such that for any $\tau > 0$, $\varepsilon \in (0, 1)$,

$$\frac{1}{n} \log \left| \mathcal{M}_{f_n} \right| \geq I\left(P_X, W\right) - 2\tau \geq R_Q - 2\tau \tag{4.4}$$

and $P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right) \leq \varepsilon$. In addition, we partition the channel realizations V into two sets

$$\mathcal{V}_1 \triangleq \left\{ V : D\left(V \| W | P_X\right) > E_{\mathrm{m}} \right\} \tag{4.5}$$

$$\mathcal{V}_2 \triangleq \left\{ V : D\left(V \| W | P_X\right) \leq E_{\mathrm{m}} \right\}, \tag{4.6}$$

and let the acceptance and rejection regions be

$$\begin{aligned} \mathcal{A}_n &= \bigcup_i \bigsqcup_{V \in \mathcal{V}_2} \mathcal{T}_V^n\left(x^n(i)\right) && (4.7) \\ &= \bigsqcup_{V \in \mathcal{V}_2} \bigcup_i \mathcal{T}_V^n\left(x^n(i)\right) && (4.8) \\ \mathcal{B}_n &= {\mathcal{A}_n}^c && (4.9) \\ &= \bigcap_i \bigsqcup_{V \in \mathcal{V}_1} \mathcal{T}_V^n\left(x^n(i)\right). && (4.10) \end{aligned}$$

We first show that this scheme satisfies the constraint that the miss probability has an exponent at least E_{m}.

For any given m,

$$\begin{aligned}
P_{\mathrm{m}}(x^n(m)) &= W^n\left(\mathcal{B}_n \middle| x^n(m)\right) && (4.11)\\
&= W^n\left(\bigcap_i \bigsqcup_{V\in\mathcal{V}_1} \mathcal{T}_V^n\left(x^n(i)\right) \middle| x^n(m)\right) && (4.12)\\
&\le W^n\left(\bigsqcup_{V\in\mathcal{V}_1} \mathcal{T}_V^n\left(x^n(m)\right) \middle| x^n(m)\right) && (4.13)\\
&= \sum_{V\in\mathcal{V}_1} W^n\left(\mathcal{T}_V^n\left(x^n(m)\right) \middle| x^n(m)\right) && (4.14)\\
&\le \sum_{V\in\mathcal{V}_1} \exp\left\{-n\left[D\left(V\| W|P_X\right)\right]\right\} && (4.15)\\
&< |\mathcal{V}_1|\, e^{-nE_{\mathrm{m}}} = e^{-n(E_{\mathrm{m}}-\lambda_n)}, && (4.16)
\end{aligned}$$

where $\lambda_n = \dfrac{1}{n}\log|\mathcal{V}_1| \to 0$ as $n \to \infty$. Hence

$$\begin{aligned}
\lim_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{m}}^{(n)} &\ge \lim_{n\to\infty} -\frac{1}{n}\log\left[\max_m P_{\mathrm{m}}(x^n(m))\right] && (4.17)\\
&\ge \lim_{n\to\infty}(E_{\mathrm{m}}-\lambda_n) = E_{\mathrm{m}}. && (4.18)
\end{aligned}$$

Given that the E_{m} constraint is satisfied, we now calculate the achievable false alarm error exponent.

$$\begin{aligned}
P_{\mathrm{f}}^{(n)} &= Q_Y(\mathcal{A}_n) && (4.19)\\
&= Q_Y\left(\bigsqcup_{V\in\mathcal{V}_2}\bigcup_i \mathcal{T}_V^n\left(x^n(i)\right)\right) = \sum_{V\in\mathcal{V}_2} Q_Y\left(\bigcup_i \mathcal{T}_V^n\left(x^n(i)\right)\right) && (4.20)\\
&\le \sum_{V\in\mathcal{V}_2}\sum_{i=1}^{|\mathcal{M}_{f_n}|} e^{-n[D(Q_V\|Q_Y)+H(Q_V)]}e^{nH(V|P_X)} && (4.21)\\
&= \sum_{V\in\mathcal{V}_2} \exp\left\{-n\left[D\left(Q_V\| Q_Y\right)+I\left(P_X,V\right)-(R_{\mathcal{Q}}-2\tau)\right]\right\}, && (4.22)
\end{aligned}$$

where $Q_V \triangleq [P_X\cdot V]_Y$. Therefore,

$$\lim_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{f}}^{(n)} \ge \min_{V\in\mathcal{V}_2}\left[D\left(Q_V\| Q_Y\right)+I\left(P_X,V\right)-R_{\mathcal{Q}}\right]. \qquad (4.23)$$

Since each $x^n(i)$ is constant composition, $\bigcup_i \mathcal{T}_V^n\left(x^n(i)\right) \subset \mathcal{T}_{Q_V}^n$. Therefore,

$$Q_Y\left(\bigcup_i \mathcal{T}_V^n\left(x^n(i)\right)\right) \leq \exp\left[-nD\left(Q_V\| Q_Y\right)\right]. \tag{4.24}$$

Hence

$$\lim_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{f}}^{(n)} \geq \min_{V\in\mathcal{V}_2} D\left(Q_V\| Q_Y\right). \tag{4.25}$$

Combing (4.23) and (4.25),

$$\lim_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{f}}^{(n)} \geq \min_{V\in\mathcal{V}_2}\left[D\left(Q_V\| Q_Y\right) + \left\{I\left(P_X, V\right) - R\right\}^+\right]. \tag{4.26}$$

Intuitively, the result in (4.23) corresponds to a bound that is tight when $\{\mathcal{T}_V^n\left(x^n(i)\right)\}$ are approximately disjoint, while the result in (4.25) corresponds to a bound that is tight when $\{\mathcal{T}_V^n\left(x^n(i)\right)\}$ are overlapping and almost fill the output type class $\mathcal{T}_{Q_V}^n$.

Then by choosing the P_X that maximizes

$$\min_{V:D(V\|W|P_X)\leq E_{\mathrm{m}}}\left[D\left(Q_V\| Q_Y\right) + \left\{I\left(P_X, V\right) - R_{\mathcal{Q}}\right\}^+\right],$$

we obtain the following lower bound to $E_{\mathrm{f}}(R_{\mathcal{Q}}, E_{\mathrm{m}})$,

$$\underline{E_{\mathrm{f}}}(R_{\mathcal{Q}}, E_{\mathrm{m}}) = \max_{P_X:I(P_X,W)\geq R_{\mathcal{Q}}}\ \min_{V:D(V\|W|P_X)\leq E_{\mathrm{m}}}\left[D\left(Q_V\| Q_Y\right) + \left\{I\left(P_X, V\right) - R_{\mathcal{Q}}\right\}^+\right]. \tag{4.27}$$

The result regarding the lower bound of miss reliability function can be established similarly. Let

$$\mathcal{V}_1' \triangleq \left\{V : D\left(Q_V\| Q_Y\right) \leq E_{\mathrm{f}}\right\} \tag{4.28}$$

$$\mathcal{V}_2' \triangleq \left\{V : D\left(Q_V\| Q_Y\right) > E_{\mathrm{f}}\right\}, \tag{4.29}$$

and define acceptance and rejection regions as

$$\mathcal{A}_n = \bigcup_i \bigsqcup_{V\in\mathcal{V}_2'} \mathcal{T}_V^n\left(x^n(i)\right) \tag{4.30}$$

$$\mathcal{B}_n = {\mathcal{A}_n}^c. \tag{4.31}$$

Then similarly, we can obtain

$$\lim_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{f}}^{(n)} \geq \min_{V\in\mathcal{V}_2'} D\left(Q_V \| \, Q_Y\right) \tag{4.32}$$

and

$$\lim_{n\to\infty} -\frac{1}{n}\log P_{\mathrm{m}}^{(n)} \geq \min_{V\in\mathcal{V}_1'} D\left(V \| \, W|P_X\right). \tag{4.33}$$

Hence the conclusion follows. □

■ 4.2 Upper bound for the false alarm reliability function of DMC

This section shows that the upper bound for $E_{\mathrm{f}}(R_{\mathcal{Q}}, E_{\mathrm{m}})$ is closely related to the output image size of a constant composition code, which can be characterized by the entropy of certain random vectors. These results do not lead to a computable upper bound for the DMC, but can be specialized to the BSC to obtain a computable upper bound that is asymptotically tight at low rate, as discussed in Section 4.4.3.

Intuitively, the size of $\mathcal{A}_n$ is related to $P_{\mathrm{f}}^{(n)} = Q_Y(\mathcal{A}_n)$, therefore, a lower bound of $|\mathcal{A}_n|$ is helpful in providing a lower bound for $P_{\mathrm{f}}^{(n)}$ and hence an upper bound for $E_{\mathrm{f}}(R_{\mathcal{Q}}, E_{\mathrm{m}})$. Also note that we would like the lower bound of $|\mathcal{A}_n|$ to be as large as possible to make the upper bound tight.

We first show that it is sufficient to consider a constant composition codebook.

Lemma 4.3. If a codebook $\mathcal{C}^{(n)}$ satisfies that, for any $\delta > 0$, $\varepsilon > 0$ and $\lambda > 0$,

$$\frac{1}{n}\log\left|\mathcal{M}_{f_n}\right| \geq R_{\mathcal{Q}} - \delta, \tag{4.34}$$

$$P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right) \leq \varepsilon \tag{4.35}$$

$$\text{and} \quad P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) \leq \exp[-n(E_{\mathrm{m}} - \lambda)], \tag{4.36}$$

then it contains a constant composition subcode $\hat{\mathcal{C}}^{(n)}$ such that

$$\frac{1}{n}\log\left|\mathcal{M}_{\hat{f}_n}\right| \geq R_{\mathcal{Q}} - 2\delta, \tag{4.37}$$

$$P_{\mathrm{e}}\left(\hat{\mathcal{C}}^{(n)}\right) \leq \varepsilon, \tag{4.38}$$

$$P_{\mathrm{m}}\left(\hat{\mathcal{C}}^{(n)}\right) \leq \exp[-n(E_{\mathrm{m}} - \lambda)], \tag{4.39}$$

and

$$P_{\mathrm{f}}\left(\hat{\mathcal{C}}^{(n)}\right) \le P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right). \tag{4.40}$$

Proof. First, we note that Lemma 3.6 shows that there is a subcode $\hat{\mathcal{C}}^{(n)}$ satisfying (4.37). Let $\hat{g}_n^{-1}(m) = g_n^{-1}(m)$ for any $m \in \mathcal{M}_{\hat{f}_n}$ and let $\hat{\mathcal{A}}_n = \mathcal{A}_n$, then

$$P_{\mathrm{e}}\left(\hat{\mathcal{C}}^{(n)}\right) \le P_{\mathrm{e}}\left(\mathcal{C}^{(n)}\right) \le \varepsilon \tag{4.41}$$

$$P_{\mathrm{m}}\left(\hat{\mathcal{C}}^{(n)}\right) \le P_{\mathrm{m}}\left(\mathcal{C}^{(n)}\right) \le \exp[-n(E_{\mathrm{m}} - \lambda)] \tag{4.42}$$

$$\text{and} \quad P_{\mathrm{f}}\left(\hat{\mathcal{C}}^{(n)}\right) = P_{\mathrm{f}}\left(\mathcal{C}^{(n)}\right) \tag{4.43}$$

□

Lemma 4.3 shows that we can lower bound the false alarm probability of a channel code by lower bounding the false alarm probability of its constant composition subcode with the same rate. Hence, it is sufficient to work with the constant composition codebook for the rest of this section.

We now show that the miss error exponent constraint E_{m} on W indicates that the probability of the channel output of a codeword falling into $\mathcal{A}_n$ needs to be large for a class of channels, which is provided by the following immediate corollary of Lemma A.3.

Corollary 4.4. Given an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ $W : \mathcal{X} \to \mathcal{Y}$ and a rate R_Q, if for any $\delta > 0$ and $\lambda > 0$, a constant composition code $\mathcal{C}^{(n)}$ with type P_X satisfies

$$\frac{1}{n} \log \left|\mathcal{M}_{f_n}\right| \ge R_{\mathcal{Q}} - \lambda \tag{4.44}$$

and

$$P_{\mathrm{m}}^{(n)} \le \exp\left[-n(E_{\mathrm{m}} - \delta)\right] \tag{4.45}$$

when n sufficiently large, then for any $V : \mathcal{X} \to \mathcal{Y}$ such that

$$D\left(V \| W | P_X\right) \le E_{\mathrm{m}} - 2\delta, \tag{4.46}$$

for any $\varepsilon > 0$ and any $x^n(m) \in \mathcal{C}^{(n)}$,

$$V^n\left(\mathcal{B}_n | x^n(m)\right) \leq \varepsilon, \text{ or equivalently, } V^n\left(\mathcal{A}_n | x^n(m)\right) \geq 1 - \varepsilon, \tag{4.47}$$

when n sufficiently large.

For a channel V, we have the following result regarding its image size based on Lemma A.6.

Lemma 4.5. For $\mathcal{C}^{(n)} \subset \mathcal{X}^n$, consider a random vector $\hat{X}^n = \hat{X}_1 \hat{X}_2 \cdots \hat{X}_n$ distributed over $\mathcal{C}^{(n)}$ with uniform distribution $P_{\hat{X}^n}$ and let the random vector $\hat{Z}^n = \hat{Z}_1 \hat{Z}_2 \cdots \hat{Z}_n$ be connected with $\hat{X}^n$ by the channel V, then for every $\tau > 0$,

$$\frac{1}{n} \log \left| \mathsf{Img}_V \left(\mathcal{C}^{(n)}, 1 - \varepsilon \right) \right| \geq S_V^{(n)} - \tau, \tag{4.48}$$

where $S_V^{(n)} \triangleq S_V\left(\mathcal{C}^{(n)}\right) = \dfrac{1}{n} H(\hat{Z}^n)$ and when n sufficiently large.

This leads to a useful lower bound on the size of $|\mathcal{A}_n|$.

Lemma 4.6. Given an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ $W : \mathcal{X} \to \mathcal{Y}$ and a rate $R_{\mathcal{Q}}$, if for any $\delta > 0$ and $\lambda > 0$, a constant composition code $\mathcal{C}^{(n)}$ with type P_X satisfies

$$\frac{1}{n} \log \left| \mathcal{M}_{f_n} \right| \geq R_{\mathcal{Q}} - \lambda \tag{4.49}$$

and

$$P_{\mathrm{m}}^{(n)} \leq \exp\left[-n(E_{\mathrm{m}} - \delta)\right] \tag{4.50}$$

when n sufficiently large, then for any $\tau > 0$,

$$\frac{1}{n} \log |\mathcal{A}_n| \geq S_*^{(n)}\left(E_{\mathrm{m}}\right) - \tau \tag{4.51}$$

when n sufficiently large, where

$$\mathcal{V}_{E_{\mathrm{m}}}^{(n)} \triangleq \left\{ V : D\left(V \| W | P_X\right) \leq E_{\mathrm{m}} - \delta \right\} \tag{4.52}$$

$$S_*^{(n)}\left(E_{\mathrm{m}}\right) \triangleq S_*\left(\mathcal{C}^{(n)}, E_{\mathrm{m}}^{(n)}\right) \triangleq \max_{V \in \mathcal{V}_{E_{\mathrm{m}}}^{(n)}} S_V^{(n)}. \tag{4.53}$$

Proof. For any $V \in \mathcal{V}_{E_\mathrm{m}}^{(n)}$, Corollary 4.4 indicates $\mathcal{A}_n$ must be an $(1-\varepsilon)$-image of $\mathcal{C}^{(n)}$ over V, thus

$$|\mathcal{A}_n| \geq \max_{V \in \mathcal{V}_{E_\mathrm{m}^{(n)}}} \left| \mathsf{Img}_V \left(\mathcal{C}^{(n)}, 1-\varepsilon \right) \right| . \tag{4.54}$$

Then the result is immediate from Lemma 4.5. □

Finally, we combine the above results to obtain the following upper bound to the false alarm reliability function.

Theorem 4.7. For a sequence of reliable codes $\mathcal{Q} = \left\{ \mathcal{C}^{(n)} \right\}$ with rate $R_\mathcal{Q}$ that achieves miss error exponent E_m over an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ $W : \mathcal{X} \to \mathcal{Y}$, the following function

$$\overline{E_\mathrm{f}}(R_\mathcal{Q}, E_\mathrm{m}) = \max_{P \in \mathcal{P}(\mathcal{Y}) : H(P) \geq S_*(R_\mathcal{Q}, E_\mathrm{m})} D\left(P \| Q_Y\right) \tag{4.55}$$

is an upper bound to the false alarm reliability function, where

$$S_*\left(R_\mathcal{Q}, E_\mathrm{m}\right) \triangleq \liminf_{n\to\infty} S_*^{(n)}\left(E_\mathrm{m}\right), \tag{4.56}$$

and the infimum is over all possible reliable codes with rate $R_\mathcal{Q}$.

Proof. Lemma 4.6 and Lemma A.5 indicate that for $\tau_n \to 0$,

$$\begin{aligned} P_\mathrm{f}^{(n)} &= Q_Y(\mathcal{A}_n) && (4.57) \\ &\geq \exp\left\{ -n \left[\max_{P' \in \mathcal{P}(\mathcal{Y}) : H(P') \geq S_*^{(n)}(E_\mathrm{m})} D\left(P' \| Q\right) + 2\tau_n \right] \right\} && (4.58) \end{aligned}$$

Thus

$$\begin{aligned} \liminf_{n\to\infty} -\frac{1}{n} \log P_\mathrm{f}^{(n)} &\leq \liminf_{n\to\infty} \max_{P' \in \mathcal{P}(\mathcal{Y}) : H(P') \geq S_*^{(n)}(E_\mathrm{m})} D\left(P' \| Q\right) + 2\tau_n && (4.59) \\ &\leq \max_{P \in \mathcal{P}(\mathcal{Y}) : H(P) \geq S_*(R_\mathcal{Q}, E_\mathrm{m})} D\left(P \| Q_Y\right) && (4.60) \end{aligned}$$

□

Theorem 4.7 shows the upper bound result is determined by $S_*\left(R_Q, E_{\mathrm{m}}\right)$, which is in turn determined by $S_V^{(n)} = \frac{1}{n}H(\hat{Z}^n)$. Thus the key is to obtain the single-letter characterization of $\frac{1}{n}H(\hat{Z}^n)$. However, this kind of characterizations for general DMCs has not been found yet and we only give an upper bound for BSC in Section 4.4.3.

■ 4.3 Performance of Separate Synchronization and Coding

In the separate synchronization–coding (training) approach, we have the same codeword structure shown in Figure 3-4. However, to satisfy the miss error exponent E_{m} requirement, the detection rule needs to be changed accordingly. Theorem 4.1 indicates that the following $(e_{\mathrm{m}}, e_{\mathrm{f}})$ are achievable at rate R_Q for an asynchronous DMC $(\mathcal{X}, \star, W, \mathcal{Y})$ W:

$$e_{\mathrm{m}} < \left(1 - \frac{R}{C}\right) D\left(P_\lambda \| P_c\right) \quad \text{and} \quad e_{\mathrm{f}} < \left(1 - \frac{R}{C}\right) D\left(P_\lambda \| Q_Y\right), \tag{4.61}$$

where $P_c\left(\cdot\right) = W\left(\cdot|x_c\right)$ and $Q_Y\left(\cdot\right) = W\left(\cdot|\star\right)$, and $P_\lambda = \zeta P_c^\lambda Q_Y^{1-\lambda}$, $0 \le \lambda \le 1$.

In Section 4.4.4, the performance of the separate sync–coding approach is compared to the performance of the joint sync–coding approach, in the context of BSC.

■ 4.4 BSC Channel

This section applies our results regarding the DMC to a simple class of channels, the BSC. We first analyze the achievable performances of both the i.i.d. codebook and the constant composition codebook. Then we show an upper bound to the false alarm reliability function, which is asymptotically tight at low rate. A tight upper bound for all rates is unknown.

Recall that we define an asynchronous BSC in Section 2.4.1 via its cross over probability ε and $\star$ output distribution $\mathsf{Bern}\left(u\right)$ (Figure 4-1). Also, without loss of generality, we assume $\varepsilon < 1/2$ and $u \le 1/2$.

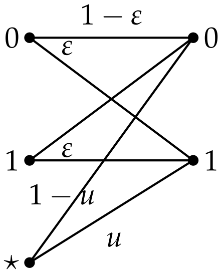


Figure 4-1: An asynchronous BSC with crossover probability ε and $\star$ output distribution $\mathsf{Bern}\left(u\right)$.

■ 4.4.1 i.i.d. codebook with optimal detection

Given a rate R, let the input distribution $\mathsf{Bern}\left(p\right)$ satisfy

$$I\left(P_X, W\right) = H_b\left(p * \varepsilon\right) - H_b\left(\varepsilon\right) = R, \tag{4.62}$$

where $p \geq 1/2$. Then $s = p * \varepsilon \geq u$, and Theorem 4.1 indicates the following $\left(e_m, e_f\right)$ values are achievable at rate R_Q for any $\gamma \in \left(u, s\right)$,

$$e_m < D\left(\gamma \| \, s\right) \text{ and } e_f < D\left(\gamma \| \, u\right). \tag{4.63}$$

■ 4.4.2 Constant composition codebook

In this section we analyze a straightforward constant composition codebook design. We first describe the codebook via its encoding strategy, then provide a simple decoding rule, and finally evaluate its performance by analyzing the detection error exponents. Numerical results show that this codebook is equivalent to the codebook proposed in Theorem 4.2.

Encoding

Given a rate R on BSC, we choose $s \geq 1/2$ and $p \geq 1/2$ such that

$$R = H_b\left(s\right) - H_b\left(\varepsilon\right) \tag{4.64}$$

$$p = \frac{s - \varepsilon}{1 - 2\varepsilon}. \tag{4.65}$$

Then following the same argument in Theorem 4.2, there exists a sequence of constant composition codebooks with rate R and type $\mathsf{Bern}\left(p\right)$ that achieves vanishing decoding error probability asymptotically. Hence, we only need to analyze the detection performance.

Detection

When we receive a sequence y^n, we use the following detection rule, which use the Hamming weight of y^n, $|y^n|_H$, as a statistic:

$$\mathcal{A}_n = \left\{y^n : |y^n|_H \geq \eta\right\} \qquad \text{(declare a codeword)} \tag{4.66}$$

$$\mathcal{B}_n = \left\{y^n : |y^n|_H < \eta\right\}, \qquad \text{(declare noise)} \tag{4.67}$$

where $\eta = \delta n$ is a properly chosen threshold. Obviously, we should use $u < \delta < s < p$.

Remark:

Strictly speaking, we need to impose the constraint that $\eta = \delta n \in \mathbb{Z}$. However, this constraint becomes inconsequential as $n \to \infty$. Hence, we ignore this constraint throughout the thesis and treat the thresholds as if they were real valued.

Error probability analysis Note that for a Binomial random variable $X \sim \mathsf{Binom}\left(n, p\right)$, we have the following exponential approximations

$$\mathbb{P}\left[X \leq n\delta\right] \doteq \exp\left[-nE_B(\delta, p)\right] \tag{4.68}$$

$$\mathbb{P}\left[X = n\delta\right] \doteq \exp\left[-nD\left(\delta\| \, p\right)\right], \tag{4.69}$$

where

$$E_B(\delta, p) = \begin{cases} 0 & \text{when } \delta \geq p \\ D\left(\delta \| \, p\right) & \text{when } \delta < p \end{cases}. \tag{4.70}$$

Based on the above results, the false alarm error probability calculation is straightforward.

$$P_{\mathrm{f}} = \mathbb{P}\left[|y^n|_H > \eta, y^n \in T_{\mathsf{Bern}(u)}\right] \tag{4.71}$$

$$\doteq \exp\left(-nD\left(\delta \| \, u\right)\right). \tag{4.72}$$

Therefore

$$E_{\mathrm{f}}(\delta) \triangleq D\left(\delta \| \, u\right). \tag{4.73}$$

Now we calculate the miss probability. Since we use a constant composition codebook, we can partition each codeword into two subsequences, each with 0s and 1s only, where they corresponds to two output subsequence $\mathbf{y}_0$ and $\mathbf{y}_1$, with

$$L_0 \triangleq |\mathbf{y}_0|_H \sim \mathsf{Binom}\left(n(1-p), \varepsilon\right) \tag{4.74}$$

$$L_1 \triangleq |\mathbf{y}_1|_H \sim \mathsf{Binom}\left(np, 1-\varepsilon\right) \tag{4.75}$$

$$|y^n|_H = L_0 + L_1 \tag{4.76}$$

$$\mathbb{E}\left[|y^n|_H\right] = n\varepsilon(1-p) + n(1-\varepsilon)p \tag{4.77}$$

$$= n(\varepsilon + p - 2\varepsilon p). \tag{4.78}$$

Then

$$
\begin{aligned}
P_{\mathrm{m}} &= \mathbb{P}\left[|y^n|_H \leq \eta\right] = \mathbb{P}\left[L_0 + L_1 \leq \eta\right] && (4.79)\\
&= \sum_{k=0}^{\eta} \mathbb{P}\left[L_0 \leq \eta - k\right] \mathbb{P}\left[L_1 = k\right] && (4.80)\\
&\doteq \sum_{k=0}^{\eta} \exp\left[-n(1-p)E_B\left(\frac{\eta - k}{(1-p)n}, \varepsilon\right)\right] \exp\left[-npD\left(\frac{k}{pn}\middle\| 1-\varepsilon\right)\right] && (4.81)\\
&\doteq \exp\left[-n \min_{\kappa\in[0,\delta]} \left((1-p)E_B\left(\frac{\delta-\kappa}{1-p}, \varepsilon\right) + pD\left(\frac{\kappa}{p}\middle\| 1-\varepsilon\right)\right)\right]. && (4.82)
\end{aligned}
$$

Note that $(1-p)E_B\left(\frac{\delta-\kappa}{1-p}, \varepsilon\right) + pD\left(\frac{\kappa}{p}\middle\| 1-\varepsilon\right)$ is convex in κ and hence the minimization can be solved numerically in a straightforward way, and we let

$$
E_{\mathrm{m}}(\delta) \triangleq \min_{\kappa\in[0,\delta]} \left((1-p)E_B\left(\frac{\delta-\kappa}{1-p}, \varepsilon\right) + pD\left(\frac{\kappa}{p}\middle\| 1-\varepsilon\right)\right), \qquad (4.83)
$$

where by convexity

$$
E_{\mathrm{m}}(\delta) \geq D\left(\delta \| s\right). \qquad (4.84)
$$

Note

$$
\begin{aligned}
\frac{\delta-\kappa}{(1-p)} < \varepsilon &\iff \kappa > \delta - (1-p)\varepsilon && (4.85)\\
\delta < s &\iff \delta - (1-p)\varepsilon < p(1-\varepsilon), && (4.86)
\end{aligned}
$$

and $D\left(\frac{\kappa}{p}\middle\| 1-\varepsilon\right)$ changes from decreasing to increasing at $\kappa = p(1-\varepsilon)$. Some simplifications show

$$
E_{\mathrm{m}}(\delta) = \min_{\kappa\in[\delta-(1-p)\varepsilon,\kappa^*]} \left[(1-p)D\left(\frac{\delta-\kappa}{1-p}\middle\| \varepsilon\right) + pD\left(\frac{\kappa}{p}\middle\| 1-\varepsilon\right)\right], \qquad (4.87)
$$

where $\kappa^* = \min\{\delta, p(1-\varepsilon)\}$.

■ 4.4.3 An upper bound for the false alarm reliability function of BSC

An upper bound for the constant composition codebook

In this section we provide an upper bound for the false alarm reliability function achieved by the constant composition codebook $\mathcal{C}$.

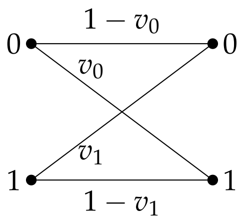


Figure 4-2: A binary channel with cross over probabilities v_0 and v_1.

Note that given a BSC channel W and miss error exponent requirement E_{m}, we have the corresponding set

$$\mathcal{V}_{E_{\mathrm{m}}} = \left\{ V : D\left(V \| W | P_X\right) \leq E_{\mathrm{m}} \right\}, \tag{4.88}$$

where V is an binary channel with cross-over probabilities $V(1|0) = v_1$ and $V(0|1) = v_0$, as shown in Figure 4-2.

Let the output of the channel V be $\hat{Z}^n$. Theorem 4.7 indicates that the key to finding a good upper bound is a good bound for $S_*\left(R_Q, E_{\mathrm{m}}\right)$, which lower bounds $\dfrac{1}{n}\log H(\hat{Z}^n)$. Below we obtain one lower bound for $\dfrac{1}{n}\log H(\hat{Z}^n)$ via a channel decomposition technique and a binary entropy inequality.

Note that for a channel V with $v_0 \geq \varepsilon$ and $v_1 \geq \varepsilon$, we can decompose it as the cascade of a binary symmetric channel W and a binary channel U, as shown in Figure 4-3, where

$$\begin{cases} v_0 &= (1-u_1)\varepsilon + u_0(1-\varepsilon) \\ v_1 &= u_1(1-\varepsilon) + (1-u_0)\varepsilon \end{cases} \Longleftrightarrow \begin{cases} u_0 &= \dfrac{(1-\varepsilon)v_1 + \varepsilon v_0 - \varepsilon}{1-2\varepsilon} \\ u_1 &= \dfrac{(1-\varepsilon)v_0 + \varepsilon v_1 - \varepsilon}{1-2\varepsilon} \end{cases}. \tag{4.89}$$

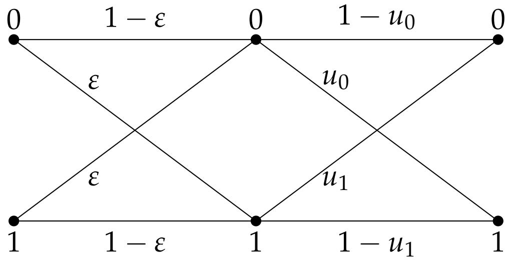


Figure 4-3: Decomposing a binary channel into the cascade of a BSC and a binary channel.

$$\begin{tikzcd}
\hat{X}^n \arrow[r, "W^n"] \arrow[rr, bend left, "V^n"] & \hat{Y}^n \arrow[r, "U^n"] & \hat{Z}^n
\end{tikzcd}$$

Figure 4-4: Input and output random vectors for a cascaded channel, where $\hat{Y}^n$ connects to $\hat{X}^n$ via channel W^n, $\hat{Z}^n$ connects to $\hat{Y}^n$ via channel U^n, and $\hat{Z}^n$ connects to $\hat{X}^n$ via channel V^n.

Let $h(\cdot)$ be the function that maps V to the corresponding U in the above decomposition, and define

$$\mathcal{V}_{E_{\mathrm{m}},+} \triangleq \{V : V \in \mathcal{V}_{E_{\mathrm{m}}}, v_0 \geq \varepsilon, v_1 \geq \varepsilon\} \tag{4.90}$$

$$\mathcal{U}_{E_{\mathrm{m}},+} \triangleq \{U : U = h(V), V \in \mathcal{V}_{E_{\mathrm{m}},+}\}\,. \tag{4.91}$$

Furthermore, we let $\hat{X}^n$ be uniformly distributed over $\mathcal{C}$, and $\hat{Y}^n$ be connected with $\hat{X}^n$ by the channel $W^n : \{0,1\}^n \to \{0,1\}^n$, $\hat{Z}^n$ be connected with $\hat{X}^n$ by the channel $V^n : \{0,1\}^n \to \{0,1\}^n$, or equivalently, $\hat{Z}^n$ be connected with $\hat{Y}^n$ by the channel $U^n : \{0,1\}^n \to \{0,1\}^n$, as shown in Figure 4-4.

We first show the well-known fact that we can characterize $H(\hat{Y}^n)$ when $\mathcal{C}$ is a channel node.

Proposition 4.8. For any (n,ε) channel code $\mathcal{C}$ on channel W with rate $R_{\mathcal{Q}}$, i.e., for any $\delta > 0$, $\frac{1}{n}\log|\mathcal{M}_{f_n}| \geq R_{\mathcal{Q}} - \delta$, then for any $\lambda > 0$,

$$\frac{1}{n}H(\hat{Y}^n) \geq R_{\mathcal{Q}} + H_b(\varepsilon) - \lambda, \tag{4.92}$$

when n sufficiently large.

Proof. Note $H(\hat{X}^n) = \frac{1}{n}\log|\mathcal{M}_{f_n}| \geq R_{\mathcal{Q}} - \delta$, and for any (n,ε) channel code $\mathcal{C}$, Lemma A.6 shows

$$\frac{1}{n}\log|\mathsf{Img}_W(\mathcal{C}, 1-\varepsilon)| - \delta \leq \frac{1}{n}H(\hat{Y}^n) \leq \frac{1}{n}\log|\mathsf{Img}_W(\mathcal{C}, 1-\varepsilon)| + \delta, \tag{4.93}$$

when n sufficiently large and for any $\tau > 0$

$$\frac{1}{n}\log|\mathsf{Img}_W(\mathcal{C}, 1-\varepsilon)| > \frac{1}{n}\log|\mathcal{M}_{f_n}| + H(W|P) - \tau \tag{4.94}$$

(from Lemma 2.1.4 in [12]).

Thus,

$$\frac{1}{n}H(\hat{Y}^n) \geq R_{\mathcal{Q}} + H_b(\varepsilon) - \delta - \tau. \tag{4.95}$$

Letting $\delta = \tau = \lambda/2$ gives (4.92). □

Given that $\frac{1}{n}H(\hat{Y}^n)$ is lower bounded, we can show that $\frac{1}{n}H(\hat{Z}^n)$ cannot be too small either, based on the following entropy inequality.

Theorem 4.9 (Entropy inequality for the binary channel [16])**.** For a binary channel U with cross over probabilities u_0 and u_1, input vector $\hat{Y}^n$, output vector $\hat{Z}^n$, and $\frac{1}{n}H(\hat{Y}^n) \geq x$,

$$\frac{1}{n}H(\hat{Z}^n) \geq \min_{p:H_b(p)=x} H_b(p * U), \tag{4.96}$$

where $p * U \triangleq p * (1-u_1) + (1-p) * u_0$.

Remark:

The lower bound in Theorem 4.9 is tight in the sense that there exists a $\hat{Y}^n$ such that $\frac{1}{n}H(\hat{Y}^n) = x$ and its corresponding output vector $\hat{Z}^n$ has entropy $\frac{1}{n}H(\hat{Z}^n) = \min_{p:H_b(p)=x} H_b(p * U)$. However, given extra information about $\hat{Y}^n$, such as which set it is distributed over, the lower bound in (4.96) may no longer be tight.

Proposition 4.8 and Theorem 4.9 give us the following upper bound for the BSC.

Theorem 4.10. For a sequence of reliable codes $\mathcal{Q} = \left\{\mathcal{C}^{(n)}\right\}$ with rate $R_{\mathcal{Q}}$ that achieves miss error exponent E_{m} over a BSC W, the following function

$$\overline{E_{\mathrm{f}}}(R_{\mathcal{Q}}, E_{\mathrm{m}}) = \max_{s:H_b(s)\geq x} D(s \| u) \tag{4.97}$$

is an upper bound for the false alarm reliability function $E_{\mathrm{f}}(R_{\mathcal{Q}}, E_{\mathrm{m}})$, where

$$x = \max_{U\in\mathcal{U}_{E_{\mathrm{m}},+}} \min_{q:H_b(q)=R+H_b(\varepsilon)} H(q * U). \tag{4.98}$$

Proof. For any given $V \in \mathcal{V}_{E_{\mathrm{m}},+}$ and its corresponding $U = h(V)$, let $\lambda_n \to 0$, we can obtain

$$\frac{1}{n} H(\hat{Z}^n) \geq l(U) \triangleq \min_{q_n : H_b(q_n) = R + H_b(\varepsilon) - \lambda_n} H(q_n * U). \tag{4.99}$$

Thus

$$\begin{aligned} S_*^{(n)}\left(E_{\mathrm{m}}\right) &= \max_{U \in \mathcal{U}_{E_{\mathrm{m}}}} S_U^{(n)} && (4.100) \\ &\geq \max_{U \in \mathcal{U}_{E_{\mathrm{m}},+}} S_U^{(n)} && (4.101) \\ &\geq \max_{U \in \mathcal{U}_{E_{\mathrm{m}},+}} l(U) && (4.102) \\ &\geq \max_{U \in \mathcal{U}_{E_{\mathrm{m}},+}} \min_{q_n : H_b(q_n) = R + H_b(\varepsilon) - \lambda_n} H(q_n * U), && (4.103) \end{aligned}$$

when n sufficiently large. Therefore,

$$S_*\left(R_{\mathcal{Q}}, E_{\mathrm{m}}\right) = \liminf_{n \to \infty} S_*^{(n)}\left(E_{\mathrm{m}}\right) \geq \max_{U \in \mathcal{U}_{E_{\mathrm{m}},+}} \min_{q : H_b(q) = R + H_b(\varepsilon)} H(q * U). \tag{4.104}$$

Combine this with Theorem 4.7, and let $P = \mathsf{Bern}\left(s\right)$, then

$$\begin{aligned} E_{\mathrm{f}}(R_{\mathcal{Q}}, E_{\mathrm{m}}) &\leq \max_{s : H_b(s) \geq S_*(R_{\mathcal{Q}}, E_{\mathrm{m}})} D\left(s \| u\right) && (4.105) \\ &\leq \max_{s : H_b(s) \geq x} D\left(s \| u\right), && (4.106) \end{aligned}$$

where

$$x = \max_{U \in \mathcal{U}_{E_{\mathrm{m}},+}} \min_{q : H_b(q) = R + H_b(\varepsilon)} H(q * U) \quad \leq S_*\left(R_{\mathcal{Q}}, E_{\mathrm{m}}\right). \tag{4.107}$$

□

However, this upper bound is not tight, especially at high rates and large E_{m} requirements, as shown in Figure 4-5. The reason is explained in the remark of Theorem 4.9.

■ 4.4.4 Performance comparisons

We compare the performance of the i.i.d. codebook, constant composition codebook, and training for BSC in Figure 4-5, with the following conclusions.

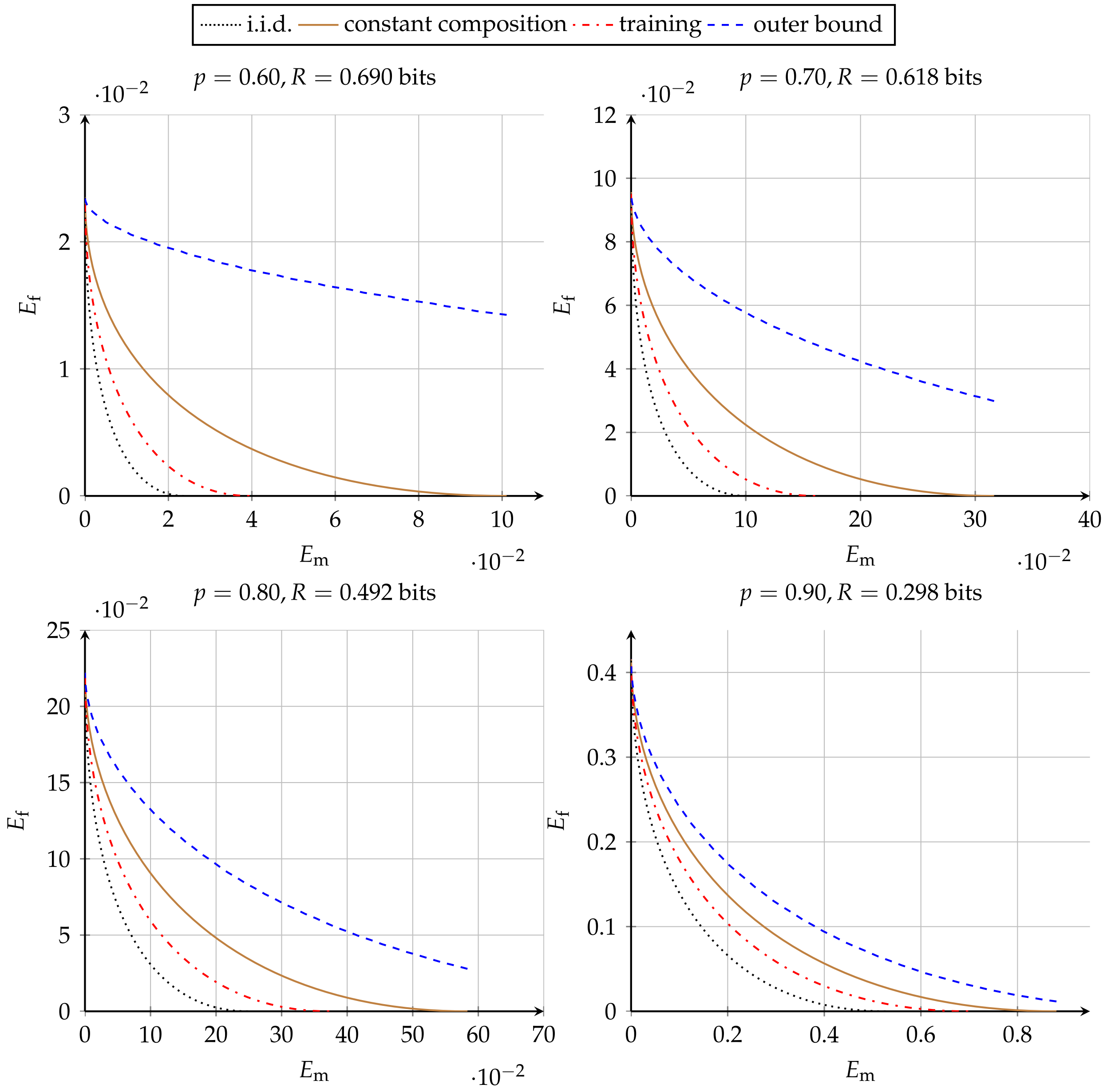


Figure 4-5: Performance comparison between i.i.d. codebook, constant composition codebook, and training for a BSC with $\varepsilon = 0.05$ and $u = 0.5$.

First, the i.i.d. codebook achieves good false alarm exponent only at very low E_m requirements. Therefore, whenever we have a strong requirement on the miss probability and thus the miss error exponent E_m, we should not use an i.i.d. codebook.

Second, the constant composition codebook achieves much larger false alarm error exponent than training, especially at high rates. Therefore, when transmitting at rates close to capacity, it is very beneficial to use a constant composition codebook instead of training. On the other hand, when transmitting at low rate, we may use training without much performance penalty, gaining the benefit of faster detection.

■ 4.5 Summary

In this chapter we establish various lower and upper bounds for the optimal false alarm reliability function with miss error exponent constraint, and these bounds indicate that it is beneficial to use the joint sync–coding scheme instead of the separate sync–coding scheme (training), especially in the high rate regime.

Unlike Chapter 3, we do not have a complete converse result and it remains an interesting open problem. Our conjecture is that for a given rate R, a constant composition codebook with type P_X such that $I\left(P_X, W\right) = R$ achieves the optimal false alarm reliability function.

One would expect that the analysis of constant composition codes for the DMC can be easily extended to the AWGN channel, as in Chapter 3. However, this is not the case. Hence, in the next chapter we characterize the optimal false alarm reliability function given a miss exponent constraint for the AWGN channel.

Chapter 5

Optimal False Alarm Exponent with Miss Exponent Constraint: AWGN Channels

As mentioned at the end of Chapter 4, we now analyze the optimal false alarm exponent with miss exponent constraint for the AWGN channel, which requires a different set of techniques that are mostly from the theory of large deviations.

In this chapter, we mainly focus on two-stage decoding strategies, i.e., strategies that first detect the presence of a codeword, then decode it according to the usual channel coding procedure. These strategies are usually simpler to analyze and often admit simpler detection rules. We formulate all detection problems as finding the large deviation exponents of certain random functions, where these functions are related to random variables that generate the codebook. Then via the standard random coding argument, achieving these large deviation exponents implies the existence of a codebook sequence that achieves the same detection error exponents asymptotically.

Similar to Chapter 4, we adopt the error exponent region description when it is more natural than the false alarm reliability function description.

Recall that the AWGN channel is of the following form:

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathsf{N}\left(0, 1\right) \ . \tag{5.1}$$

where X_i is the channel input, Y_i is the channel output and Z_i is the additive white Gaussian noise.

■ 5.1 i.i.d. Codebook with Optimal Detection

Given a rate R, following the standard random coding argument, we can generate an i.i.d. channel code according to the distribution $P_X \sim \mathsf{N}\left(\mu, \sigma^2\right)$, where μ and σ^2 satisfies $R = \frac{1}{2}\log\left(1+\sigma^2\right)$ and $\mu^2 + \sigma^2 = P$. Then at the channel output, we have a simple binary hypothesis testing problem between P_Y and Q_Y, where

$$P_Y \sim X + Z \sim \mathsf{N}(\mu, \sigma^2 + 1) \tag{5.2}$$

$$Q_Y \sim Z \sim \mathsf{N}(0, 1). \tag{5.3}$$

Then based on Theorem 4.1, let $P_\lambda = \zeta P_Y^\lambda Q_Y^{1-\lambda} \sim \mathsf{N}\left(\mu_\lambda, \sigma_\lambda^2\right)$ (see Section B.1.2 for details), then the following $(e_\mathrm{m}, e_\mathrm{f})$ values are achievable for any $\lambda \in (0,1)$,

$$e_\mathrm{m} < D\left(P_\lambda \| \, P_Y\right) \text{ and } e_\mathrm{f} < D\left(P_\lambda \| \, Q_Y\right). \tag{5.4}$$

■ 5.2 Spherical Codebook with Energy Detection

The i.i.d. Gaussian codebook in Section 5.1 corresponds to the case that codewords are uniformly distributed *within* the n-dimensional sphere. In this section, we investigate a codebook design that has all codewords uniformly distributed *on the surface* of the n-dimensional sphere.

■ 5.2.1 Encoding and decoding

We uniformly pick e^{nR} points out of the surface of a n-dimensional sphere with radius $\sqrt{nP}$, and let each point be a codeword. It is not hard to show this can generate a codebook that achieves rate R with $P_\mathrm{e} \to 0$.

At the decoder, we use the following energy detection rule:

$$\mathcal{A}_n = \left\{y^n : \|y^n\|^2 > n\eta\right\} \qquad \text{(declare a codeword)} \tag{5.5}$$

$$\mathcal{B}_n = \left\{y^n : \|y^n\|^2 \leq n\eta\right\}, \qquad \text{(declare noise)} \tag{5.6}$$

where $1 < \eta < P + 1$.

■ 5.2.2 Performance analysis

Let X^n be uniformly distributed on a n-dimensional sphere with radius $\sqrt{nP}$, and $Z^n \sim \mathsf{N}(\mathbf{0}, \mathbf{I}_k)$, where X^n and Z^n are statistically independent. Then

$$\begin{aligned} P_\mathrm{m} &= \mathbb{P}\left[\|X^n + Z^n\|^2 \leq n\eta\right] \\ &\doteq \exp[-nI_{\mathsf{SG},\leq}\left(P, \eta\right)], \end{aligned}$$

where $I_{\mathsf{SG},\leq}\left(\cdot\right)$ is given in (B.34), and

$$\begin{aligned} P_\mathrm{f} &= \mathbb{P}\left[\|Z^n\|^2 > n\eta\right] \\ &\doteq \exp[-nI_{\chi_1^2,\geq}\left(\eta\right)], \end{aligned}$$

where $I_{\chi_1^2,\geq}(\eta)$ is given in (B.16).

Therefore, given η, we can achieve the following miss and false alarm error exponent pair

$$e_{\mathrm{m}} = I_{\mathsf{SG},\leq}(P,\eta) \tag{5.7}$$

$$e_{\mathrm{f}} = I_{\chi_1^2,\geq}(\eta)\,. \tag{5.8}$$

Remark:

This energy detection rule does not take the rate of the code into account, therefore its performance is expected to be close to optimal only at high rates.

■ 5.3 Clustered Spherical Codebook

Intuitively, to reduce the false alarm error probability, we want the codeword detection region $\mathcal{A}_n$ to be small. Therefore, we would like to have codewords be close to each other, while maintaining the channel code requirements. This leads to the "clustered spherical codebook" design analyzed in this section.

■ 5.3.1 Codebook design

Given a rate R, define P_c and P_s such that (recall that noise power $N = 1$)

$$R = \log(1 + P_c) \tag{5.9}$$

$$P_s = 1 - P_c. \tag{5.10}$$

where we called P_c the "communication power" and P_s the "synchronization power".

Then we can generate a codebook as follows: choose e^{nR} points uniformly from the surface of a $(n-1)$-dimensional sphere with radius $\sqrt{nP_c}$, where the points are $\hat{X}^{n-1}(1), \hat{X}^{n-1}(2), \cdots$. Then let

$$X^n(i) = \left(\sqrt{nP_s}, \hat{X}_1(i), \hat{X}_2(i), \cdots, \hat{X}_{n-1}(i)\right) \quad i = 1, 2, \cdots, e^{nR}, \tag{5.11}$$

and use $\{X^n(i), i = 1, 2, \cdots, e^{nR}\}$ as a codebook. It is not hard to show that this can indeed generate a codebook that achieves rate R with vanishing decoding error probability.

Remark:

One drawback for the codebook design in (5.11) is the high peak power due to first symbol $\sqrt{nP_s}$, which is unbounded as n grows. However, since $\hat{X}^n$ is spherically symmetric, we can transform the codebook in (5.11) via an orthogonal transform (let **U** be an orthogonal matrix) to the following codebook,

$$X^n(i) = (\sqrt{P_s}, \dots, \sqrt{P_s}) + \mathbf{U}(0, \hat{X}_1(i), \hat{X}_2(i), \cdots, \hat{X}_{n-1}(i)), \tag{5.12}$$

which has a much lower peak power.

Therefore, when we need to impose the more practical peak power constraint rather than the average power constraint, we can use the codebook constructed in (5.12) and it achieves exactly the same performance as the one in (5.11), and the analysis in this section still holds.

■ 5.3.2 Optimal detection rule

Given the above encoding strategy, it is possible to use the log-likelihood ratio test, which is known to be optimal. And since the first component and the next $(n-1)$ components of a codeword are independent, and the $(n-1)$ components are spherically symmetric, the acceptance region can be parameterized by y_1 and $\|y_2^n\|$. However, the performance of this detector is difficult to analyze so we use the simpler detection rule below for performance evaluation.

■ 5.3.3 Heuristic detection rules

In this section we develop a heuristic detection rule that allows asymptotic performance analysis, where the following detection rule are used:

$$\mathcal{A}_n = \{y^n : ay_1 + b\|y_2^n\| \geq \sqrt{n}\eta\} \qquad \text{(declare a codeword)} \tag{5.13}$$

$$\mathcal{B}_n = \{y^n : ay_1 + b\|y_2^n\| < \sqrt{n}\eta\}\,, \qquad \text{(declare noise)} \tag{5.14}$$

where $a \in [0,1]$ and $b \in [0,1]$ are weights to be selected. Here we use y_1 and $\|y_2^n\|$ as the elements of the linear combination, because their values are also the building blocks for the optimal detection rule. This gives us a "cone shape" detection region in the n-dimension space.

Remark:

Some reasonable criteria for choosing a and b are:

- When $P_s = P$, we should have $a = 1$ and $b = 0$, and vice versa.
- When $P_s \geq P_c$, we should have $a \geq b$, and vice versa.

Given a and b, the above detection rule achieves the following false alarm and miss error exponents (see calculation details in Section B.5.1).

$$e_{\mathrm{f}}(\eta, a, b) = \begin{cases} 0 & \eta \leq b \\ \min\limits_{0 \leq r \leq \eta - b} \frac{r^2}{2a^2} + I_{\chi_1^2}\left(\frac{(\eta - r)^2}{b^2}\right) & \eta > b \end{cases} \tag{5.15}$$

$$e_{\mathrm{m}}(\eta, a, b) = \begin{cases} +\infty & \eta \leq 0 \\ \min\limits_{\eta - b\sqrt{P_c + 1} \leq r \leq \eta} \frac{(r - a\sqrt{P_s})^2}{2a^2} + I_{\mathsf{SG}}\left(P_c, \frac{(\eta - r)^2}{b^2}\right) & 0 < \eta < a\sqrt{P_s} + b\sqrt{P_c + 1} \\ 0 & \eta \geq a\sqrt{P_s} + b\sqrt{P_c + 1} \end{cases} . \tag{5.16}$$

Hence the achievable error exponents of this scheme are

$$e_{\mathrm{f}}(\eta) = \max_{(a,b) \in [0,1] \times [0,1]} e_{\mathrm{f}}(\eta, a, b) \tag{5.17}$$

$$e_{\mathrm{m}}(\eta) = \max_{(a,b) \in [0,1] \times [0,1]} e_{\mathrm{m}}(\eta, a, b). \tag{5.18}$$

Remark:

The optimization problems in (5.15) and (5.16) in general do not admit analytical solutions, but they are easy to solve numerically.

Alternatively, we can imagine using the following detection rule, which also seems quite natural:

$$\mathcal{A}_n = \left\{y^n : a\|y_1\|^2 + b\|y_2^n\|^2 \geq n\eta\right\} \qquad \text{(declare a codeword)} \tag{5.19}$$

$$\mathcal{B}_n = \left\{y^n : a\|y_1\|^2 + b\|y_2^n\|^2 < n\eta\right\}, \qquad \text{(declare noise)} \tag{5.20}$$

This detection rule achieves the following error exponents (see detailed calculations in Section B.5.2):

$$e_f(\eta, a, b) = \begin{cases} 0 & \eta \leq b \\ \min\limits_{0 \leq r \leq \eta - b} \dfrac{r}{2a} + I_{\chi_1^2}\left(\dfrac{\eta - r}{b}\right) & \eta > b \end{cases} \tag{5.21}$$

$$e_m(\eta, a, b) = \begin{cases} +\infty & \eta \leq 0 \\ \min\limits_{r_l \leq r \leq r_u} \dfrac{(\sqrt{r/a} - \sqrt{P_s})^2}{2} + I_{\mathsf{SG}}\left(P_c, \dfrac{\eta - r}{b}\right) & 0 < \eta < aP_s + b(P_c + 1)\,. \\ 0 & \eta \geq aP_s + b(P_c + 1) \end{cases} \tag{5.22}$$

However, the scheme based on (5.20) in general performs slightly worse than the scheme based on (5.14). This is due to the fact that we use y_1^2 in (5.20), where the sign of y_1 contains some useful information for detection.

■ 5.3.4 Training: detection based on synchronization power only

For the AWGN channel, separate sync–coding (training) has the same codebook structure as the clustered spherical codebook, but the detection is simply based on the synchronization power $\sqrt{nP_s}$. Thus, given threshold $n\eta$,

$$P_{\mathrm{m}} = \mathbb{P}\left[\mathsf{N}\left(\sqrt{nP_s}, 0\right) + \mathsf{N}(0,1) < n\eta\right] \qquad P_{\mathrm{f}} = \mathbb{P}\left[\mathsf{N}(0,1) \geq n\eta\right] \tag{5.23}$$

$$\doteq \exp\left[-n\left(\frac{(\sqrt{P_s} - \eta)^2}{2}\right)\right], \qquad \doteq \exp\left[-n\left(\frac{\eta^2}{2}\right)\right]. \tag{5.24}$$

Hence we can achieve error exponents $e_{\mathrm{m}}(\eta) = \left(\dfrac{(\sqrt{P_s} - \eta)^2}{2}\right)$ and $e_{\mathrm{f}}(\eta) = \left(\dfrac{\eta^2}{2}\right)$.

Since this detection relies only on the synchronization power P_s, which is large only at low rate. Therefore, the detection performance of training degrades as the rate increases.

■ 5.4 Performance Comparisons

In this section, we compute the achievable error exponents of the schemes analyzed in this chapter and compare their performances. Four schemes are considered: i.i.d. codebook with optimal detection, spherical codebook with energy detection, training, and clustered spherical codebook with the heuristic detection rule specified in (5.14), where we choose $a = \dfrac{\sqrt{P_s}}{\sqrt{P_s} + \sqrt{P_c}}$ and $b = \dfrac{\sqrt{P_c}}{\sqrt{P_s} + \sqrt{P_c}}$ as the weighting parameters. This set of parameters is shown to be optimal or near-optimal for most channels and rates in numerical calculations.

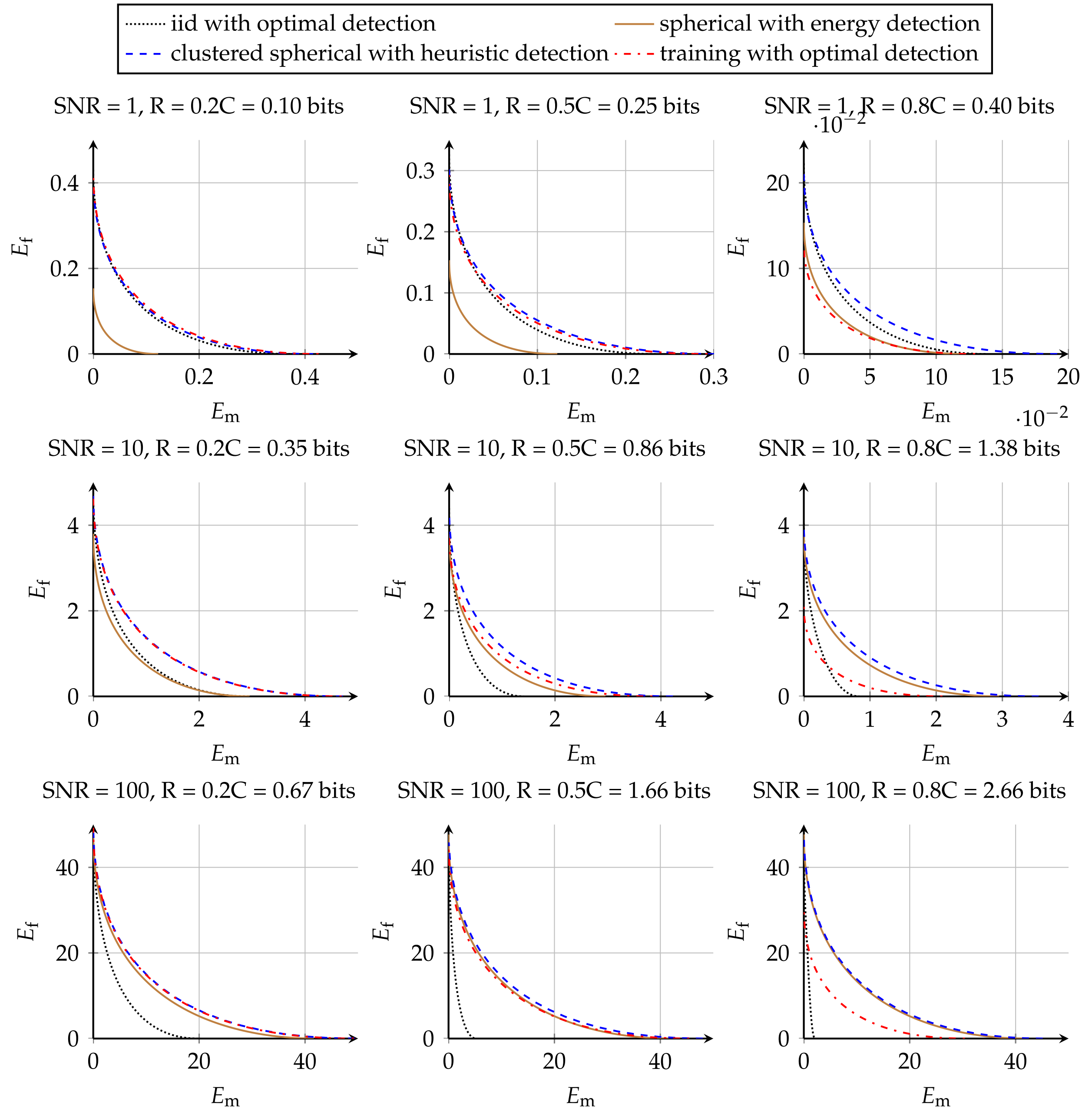


Figure 5-1: Performance comparison between i.i.d. codebook, spherical codebook, clustered spherical codebook, and training for AWGN channels at different SNRs and communication rates.

Figure 5-1 shows the achievable error exponents of these four schemes for AWGN channels with different SNRs and communication rates. As we can see, in different regimes, it may be beneficial to use different coding and detection schemes.

In the low SNR regime, when we communicate at low rates, i.i.d. codebook, clustered spherical codebook and training perform almost equally well and hence any of these schemes can be selected. When we communicate at high rates, the clustered spherical codebook is more suitable because it achieves better E_{m}–E_{f} trade-off than all other three schemes.

In the high SNR regime, the i.i.d. codebook performs poorly, because in this regime, the channel is well-behaved so that the atypicality introduced during the codebook generation affects the performance significantly. Also, the spherical codebook with energy detection and the clustered spherical codebook perform almost equally well. Furthermore, training performs well only at low rates and suffers significant performance degradation at high rates. Hence, in the high SNR regime, we can use a spherical codebook without much loss of performance, gaining the benefit of having a simpler detection rule.

As the design of the clustered spherical codebook combines the advantages of training and the spherical codebook, its performance is robust with respect to SNR and rate changes. This demonstrates that better codebook design and better detection strategy combined can achieve significant performance improvements.

■ 5.5 Summary

This chapter investigates the achievable false alarm error exponent (given miss error exponent constraint) for a variety of schemes at different channel conditions and rates for the AWGN channel. The results indicate that, if the system is designed to operate under low rates, then there is no need to modify the existing separate sync–coding architecture. However, if the system may operate in the high SNR or high rate regime, it is more beneficial to use the spherical codebook with energy detection or the clustered spherical codebook with heuristic detection, respectively.

Among all schemes investigated, the clustered spherical codebook with heuristic detection rule performs the best at essentially all SNRs and all rates; hence it is a robust choice when there is uncertainty about the regime that the system operates at.

Chapter 6

Conclusions

This thesis shows in sparse communication, for many regimes of interest, it is beneficial to treat the problem of synchronization and coding jointly, instead of the traditional approach of separating synchronization and coding. By designing codebooks that can be easily distinguished from noise, the joint approach could result significant performance gain in terms of detection error exponents.

In Chapter 3, we show that for the DMC, if we only require $P_{\mathrm{e}} \to 0$ and $P_{\mathrm{m}} \to 0$, then an i.i.d. codebook is sufficient to achieve the optimal error exponent. And training is in general suboptimal, especially at high rates.

In Chapter 4, we show, when there is a requirements on the error exponent of P_{m}, a constant composition codebook achieves the best known performance on the DMC.

The performance of various coding schemes on the DMC is summarized in Table 6.1.

	constant composition codebook	**i.i.d. codebook**	**training**
$E_{\mathrm{m}} = 0$	optimal	optimal	suboptimal
$E_{\mathrm{m}} \geq 0$	best known achievable	suboptimal	suboptimal

Table 6.1: Comparisons for coding schemes on DMCs.

In Chapter 5, we turn to the AWGN channel and analyze its false alarm reliability function. We show that the clustered spherical codebook with heuristic decoding performs uniformly well in all SNR and rate regimes, while other schemes may be better or worse depending on the channel SNR and communication rate. The results are summarized in Table 6.2, where the schemes are ordered by performance (A $<$ B means the performance of A is worse than the performance of B).

	Low SNR	**High SNR**
Low rate	spherical $<$ i.i.d. $\approx$ training $\approx$ clustered spherical	i.i.d. $<$ spherical $<$ training $\approx$ clustered spherical
High rate	training $<$ spherical $<$ i.i.d. $<$ clustered spherical	i.i.d. $<$ training $<$ spherical $\approx$ clustered spherical

Table 6.2: Comparisons for coding schemes on AWGN channels.

Remark: analogy between BSC and AWGN channels

The BSC and the AWGN channel are popular channel models for information-theoretic analysis. Results for these two channels can often be connected via a nice analogy, since both channels are additive and there is a geometric similarity between the Hamming space and the Euclidean space.

In our results, we show that for BSC channels, the constant composition codebook with type P_X such that $I(P_X, W) = R$ achieves the best known performance, whose analog for the AWGN channel is the spherical codebook. However, on the AWGN channel, we show that the clustered spherical codebook performs better than the spherical codebook in general. This distinction between the BSC and the AWGN channel is mainly due to the power constraint we impose on coding for the AWGN channel. For the BSC, if $R = H_b(p * \varepsilon) - H_b(\varepsilon)$ and we constrain the Hamming weight of each codeword to be no more than αn, where $\alpha < p$, then we can no longer use the constant composition codebook with type $\mathsf{Bern}(p)$, and the resulting codebook would be something analogous to the clustered spherical codebook for the AWGN channel.

Future Work

Several interesting issues are still open for future work.

Regarding our analysis for the DMC, the upper bound in Chapter 4 is not tight, and it may be of interest to further enhance the upper bound results. For example, it will be of much practical interest to see if a constant composition codebook with type P_X such that $I(P_X, W) = R_Q$ is optimal. Similarly, it will be interesting to establish upper bounds for our analysis on the AWGN channel in Chapter 5 as well.

In addition, this work can be extended to accommodate more performance metrics. One possibility is to include the decoding error exponent in the analysis, and analyze the trade-offs between the three error exponents E_m, E_f and E_d, instead of two of them. Another possible extension is to characterize error bounds with respect to finite block length rather than infinite block length, as in channel coding [17].

Finally, there is a connection between the problem of distinguishing codes from noise and the problem of unequal error protection of one special message (which corresponds to the noise sequence in this thesis), where results at capacity are given in [18]. The main difference is, in the unequal error protection setting, we can design the special message to make it “different” from other messages. This actually corresponds to a more active transmission strategy in sparse communication: the transmitter transmits even in the time slots that we have no message to sent, and try to induce channel outputs that are different from the outputs of the codewords in the codebook. It will be of interest to draw connection between the two. One possibility is to extend the results in [18] to rates below capacity, which provides us with the performance limit in the scenario of using an active transmitter.

Appendix A
Useful Results in Information Theory

This appendix lists certain results in information theory that are used in the thesis.

■ A.1 Inequalities

Fact A.1.

$$D\left(P_{XY}\|\left[P_X\cdot Q_Y\right]\right)=I(P_X;P_Y)+D\left(P_Y\|\,Q_Y\right). \tag{A.1}$$

Proof.

$$\begin{aligned}D\left(P_{XY}\|\left[P_X\cdot Q_Y\right]\right)&=\sum_{x,y}P_{XY}(x,y)\log\frac{P_{XY}(x,y)}{P_X(x)Q_Y(y)} &\text{(A.2)}\\&=\sum_{x,y}P_{XY}(x,y)\log\frac{P_{X|Y}(y|x)}{P_X(x)}+\sum_{x,y}P_{XY}(x,y)\log\frac{P_Y(y)}{Q_Y(y)} &\text{(A.3)}\\&=I(P_X;P_Y)+D\left(P_Y\|\,Q_Y\right) &\text{(A.4)}\end{aligned}$$

□

■ A.2 Lemmas and Propositions

Proposition A.1. Define $f:\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{X})\to\mathbb{R}$ that

$$f(P,Q)\triangleq D\left(P\|\,Q\right)+H(P)=-\sum_{x\in\mathcal{X}}P(x)\log Q(x). \tag{A.5}$$

Given probability distributions P, P' and Q over finite alphabet $\mathcal{X}$, if Support$(P,P')\subset$ Support(Q) and

$$\|P'-P\|_\infty<\delta, \tag{A.6}$$

then $f(\cdot,Q)$ is Lipschitz continuous with constant $M|\mathcal{X}|$, where $M=\max_{x\in\text{Support}(P,P')}|\log Q(x)|$.

Proof. Support$(P,P')\subset$ Support(Q) indicates that for any $x\in$ Support(P,P'), $0<Q(x)\leq 1$, therefore, the set $\{|\log Q(x)|: x\in\text{Support}(P,P')\}$ is finite and bounded, and hence M indeed exists.

Then for any P' and P such that

$$\|P' - P\|_\infty < \delta, \tag{A.7}$$

we have

$$\begin{aligned}\left|f(P,Q) - f(P',Q)\right| &= \left|\sum_{x\in\mathcal{X}} (P'(x) - P(x)) \log Q(x)\right| && \text{(A.8)}\\ &\le \sum_{x\in\mathcal{X}} \left|P'(x) - P(x)\right| \left|\log Q(x)\right| && \text{(A.9)}\\ &\le M|\mathcal{X}|\delta && \text{(A.10)}\end{aligned}$$

□

Lemma A.2. Given two probability distribution P and Q over alphabet $\mathcal{X}$, and Support$(Q) = \mathcal{X}$, for any $x^n \in \mathcal{T}^n_{[P]_{\delta_n}}$,

$$Q(x^n) \doteq \exp\left[-n(H(P) + D\left(P\|\, Q\right))\right]. \tag{A.11}$$

Proof. Note

$$\mathcal{T}^n_{[P]_{\delta_n}} = \bigcup_{P':\|P'-P\|_\infty<\delta_n} \mathcal{T}^n_{P'} \tag{A.12}$$

For any $x^n \in \mathcal{T}_{P'}$,

$$Q(x^n) = \text{poly}(n) \exp\left[-n(H(P') + D\left(P'\|\, Q\right))\right] \tag{A.13}$$

then based on Proposition A.1,

$$\begin{aligned}\text{poly}(n) \exp\left[-n\left(H(P) + D\left(P\|\, Q\right) + M|\mathcal{X}|\delta_n\right)\right] &\le Q(x^n) && \text{(A.14)}\\ &\le \text{poly}(n) \exp\left[-n\left(H(P) + D\left(P\|\, Q\right) - M|\mathcal{X}|\delta_n\right)\right] && \text{(A.15)}\end{aligned}$$

Since $\delta_n \to 0$ as $n \to \infty$ (Delta Convention), and $\lim_{n\to\infty} \frac{1}{n} \log \text{poly}(n) = 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log Q(x^n) = H(P) + D\left(P\|\, Q\right) \tag{A.16}$$

Hence,

$$Q(x^n) \doteq \exp\left[-n\left(H(P) + D\left(P\| Q\right)\right)\right] \tag{A.17}$$

□

Lemma A.3. For a sequence of sets $\{\mathcal{B}_n : \mathcal{B}_n \subset \mathcal{X}^n\}$ and a distribution $Q \in \mathcal{P}(\mathcal{X})$ such that for any $\delta > 0$,

$$Q^n(\mathcal{B}_n) \le e^{-n(E-\delta)},$$

when n sufficiently large, then for any $\delta' > \delta$ and any $P \in \mathcal{P}(\mathcal{X})$ with $D\left(P\| Q\right) \le E - \delta'$, for any $\varepsilon > 0$,

$$P^n(\mathcal{B}_n) \le \varepsilon, \tag{A.18}$$

when n sufficiently large.

Proof. Let $\mathcal{D}_1 \triangleq \mathcal{B}_n \cap \mathcal{T}^n_{[P]}$ and $\mathcal{D}_2 \triangleq \mathcal{B}_n \cap \mathcal{T}^{n\ c}_{[P]}$, then $\mathcal{B}_n = \mathcal{D}_1 \bigsqcup \mathcal{D}_2$ and

$$\begin{aligned} e^{-n(E-\delta)} &\ge Q^n(\mathcal{B}_n) \\ &\ge Q^n(\mathcal{D}_1) = \sum_{x^n \in \mathcal{D}_1} Q^n(x^n). \end{aligned}$$

Note that for any $x^n \in \mathcal{T}^n_{[P]}$ and $\lambda = (\delta - \delta')/4$, when n sufficiently large,

$$\begin{aligned} \mathrm{poly}(n) \exp\left[-n\left(H(P) + D\left(P\| Q\right) + \lambda\right)\right] &\le Q(x^n) \qquad \text{(A.19)} \\ &\le \mathrm{poly}(n) \exp\left[-n\left(H(P) + D\left(P\| Q\right) - \lambda\right)\right]. \qquad \text{(A.20)} \end{aligned}$$

Hence

$$\begin{aligned} e^{-n(E-\delta)} &\ge \sum_{x^n \in \mathcal{D}_1} Q^n(x^n) \qquad \text{(A.21)} \\ &\ge |\mathcal{D}_1| \,\mathrm{poly}(n) \exp\left[-n\left(H(P) + D\left(P\| Q\right) + \lambda\right)\right]. \qquad \text{(A.22)} \end{aligned}$$

Therefore,

$$\begin{aligned} |\mathcal{D}_1| &\le \mathrm{poly}(n) \exp\left[n\left(H(P) + D\left(P\| Q\right) + \lambda - E + \delta\right)\right] \qquad \text{(A.23)} \\ &\le \mathrm{poly}(n) \exp\left[n\left(H(P) - \delta' + \delta + \lambda\right)\right]. \qquad \text{(A.24)} \end{aligned}$$

Note that

$$P^n(\mathcal{B}_n) = P^n(\mathcal{D}_1) + P^n(\mathcal{D}_2). \tag{A.25}$$

And when n sufficiently large,

$$\begin{aligned}
P^n(\mathcal{D}_2) &\leq P^n({\mathcal{T}^n_{[P]}}^c) && \text{(A.26)}\\
&= 1 - P^n(\mathcal{T}^n_{[P]}) && \text{(A.27)}\\
&\leq 1 - (1 - \varepsilon/2) && \text{(A.28)}\\
&\leq \varepsilon/2, && \text{(A.29)}\\
P^n(\mathcal{D}_1) &= \sum_{x^n \in \mathcal{D}_1} P^n(x^n) && \text{(A.30)}\\
&\leq |\mathcal{D}_1| \operatorname{poly}(n) \exp\left[-n\left(H(P) - \lambda\right)\right] && \text{(A.31)}\\
&\leq \operatorname{poly}(n) \exp\left[n\left(H(P) - \delta' + \delta + \lambda\right)\right] \exp\left[-n\left(H(P) - \lambda\right)\right] && \text{(A.32)}\\
&= \operatorname{poly}(n) \exp\left[-n(\delta' - \delta - 2\lambda)\right] && \text{(A.33)}\\
&= \operatorname{poly}(n) \exp\left[-n(2\lambda)\right] && \text{(A.34)}\\
&\leq \varepsilon/2. && \text{(A.35)}
\end{aligned}$$

Hence $P^n(\mathcal{B}_n) \leq \varepsilon$ when n sufficiently large. □

Lemma A.4. For the set $\mathcal{A}_n \subset \mathcal{X}^n$ and $P \in \mathcal{P}(\mathcal{X})$ such that for any $\delta > 0$,

$$|\mathcal{A}_n| \geq \exp[n(H(P) - \delta)]. \tag{A.36}$$

Then there exists $P' \in \mathcal{P}(\mathcal{X})$ such that $H(P') \geq H(P)$ and

$$|\mathcal{A}_n \cap \mathcal{T}^n_{P'}| \geq \exp[n(H(P') - 2\delta)], \tag{A.37}$$

when n sufficiently large.

Proof. This directly follows from the Type Counting Lemma (Lemma 1.2.2 in [12]). □

Lemma A.5. For the set $\mathcal{A}_n \subset \mathcal{X}^n$ and $P \in \mathcal{P}(\mathcal{X})$ such that for any $\delta > 0$,

$$|\mathcal{A}_n| \geq \exp[n(H(P) - \delta)]. \tag{A.38}$$

We have for any $Q \in \mathcal{P}(\mathcal{X})$,

$$Q^n(\mathcal{A}_n) \geq \exp\left\{-n\left[\max_{P' \in \mathcal{P}(\mathcal{X}): H(P') \geq H(P)} D\left(P' \| Q\right) + 2\delta\right]\right\}, \tag{A.39}$$

when n sufficiently large.

Proof. Lemma A.4 indicates there exist at least one $P' \in \mathcal{P}(\mathcal{X})$ such that $H(P') \geq H(P)$ and

$$|\mathcal{A}_n \cap \mathcal{T}^n_{P'}| \geq \exp[n(H(P') - 2\delta)]. \tag{A.40}$$

Thus

$$\begin{aligned} Q(\mathcal{A}_n) \geq Q(\mathcal{A}_n \cap \mathcal{T}^n_{P'}) \quad &= |\mathcal{A}_n \cap \mathcal{T}^n_{P'}| \exp\left\{-n\left[D\left(P' \| Q\right) + H(P')\right]\right\} \\ &\geq \exp[n(H(P') - 2\delta)] \exp\left\{-n\left[D\left(P' \| Q\right) + H(P')\right]\right\} \\ &\geq \exp\left\{-n\left[D\left(P' \| Q\right) + 2\delta\right]\right\}. \end{aligned}$$

So,

$$Q(\mathcal{A}_n) \geq \exp\left\{-n\left[\max_{P': H(P') \geq H(P)} D\left(P' \| Q\right) + 2\delta\right]\right\}. \tag{A.41}$$

□

A.2.1 Entropy and image size characterization

Lemma A.6 (Relationship between image size and entropy [12])**.** For any set $\mathcal{A} \subset \mathcal{X}^n$, consider random vector X^n distributed over $\mathcal{A}$ and let random vector $Y^n = Y^n$ be connected with X^n by the channel $W^n : \mathcal{X}^n \to \mathcal{Y}^n$. Then for every $\delta > 0$, $0 < \varepsilon < 1$

$$\frac{1}{n} H(Y^n) - \delta \leq \frac{1}{n} \log \left|\mathsf{Img}_W(\mathcal{A}, \eta)\right|, \tag{A.42}$$

when $n \geq n_0\left(|\mathcal{X}|, |\mathcal{Y}|, \delta, \eta\right)$.

Moreover, if $\mathcal{A} \subset \mathcal{T}^n_{[X]}$ is the codeword set of an (n, ε)-code for the DMC W, and X^n has uniform distribution over $\mathcal{A}$, then one also has

$$\frac{1}{n} H(Y^n) + \delta + \varepsilon \log |\mathcal{Y}| \geq \frac{1}{n} \log \left|\mathsf{Img}_W(\mathcal{A}, \eta)\right|, \tag{A.43}$$

provided $n \geq n_0\left(|\mathcal{X}|, |\mathcal{Y}|, \delta, \eta\right)$.

Appendix B
Large Deviation Results for the AWGN Channel

In this appendix, we derive certain exponential approximation results that are useful for calculating the performance bounds in Chapter 5. Most results are based on the theory of large deviations.

Given $X_i \sim P_X, i = 1, \cdots, n$, define $I_{X,\geq}(x)$ and $I_{X,\leq}(x)$ as follows:

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \geq x\right] \doteq \exp\left(-nI_{X,\geq}(x)\right) \tag{B.1}$$

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n} X_i \leq x\right] \doteq \exp\left(-nI_{X,\leq}(x)\right). \tag{B.2}$$

We derive $I_{X,\geq}(x)$, $I_{X,\leq}(x)$ for various distributions P_X that are related to the AWGN channel in the following sections.

■ B.1 Gaussian Distribution

For $P_X \sim \mathsf{N}(\mu, \sigma^2)$, then it is well known that

$$I_{\mathsf{N},\geq}(x) = \begin{cases} \dfrac{(\mu - x)^2}{2\sigma^2} & x > \mu \\ 0 & x \leq \mu \end{cases} \tag{B.3}$$

$$I_{\mathsf{N},\leq}(x) = \begin{cases} \dfrac{(\mu - x)^2}{2\sigma^2} & x < \mu \\ 0 & x \geq \mu \end{cases}. \tag{B.4}$$

In addition to the above rate functions, we also derive the expressions for divergence between two Gaussians and the geometric mean of two Gaussians, which is useful in the application of Chernoff bound.

■ B.1.1 Divergences between two Gaussians

If $P_1 \sim \mathsf{N}(\mu_1, \sigma_1^2), P_2 \sim \mathsf{N}(\mu_2, \sigma_2^2)$, then

$$D(P_1 \| P_2) = \frac{1}{2}\Big[\log\frac{\sigma_2^2}{\sigma_1^2} + \frac{(\mu_1 - \mu_2)^2 + (\sigma_1^2 - \sigma_2^2)}{\sigma_2^2}\Big]. \tag{B.5}$$

When $\sigma_1 = \sigma_2 = \sigma$,

$$D(P_1 \| P_2) = \frac{(\mu_1 - \mu_2)^2}{2\sigma^2} = D\left(P_2 \| \, P_1\right). \tag{B.6}$$

When $\mu_1 = \mu_2 = \mu$,

$$D(P_1 \| P_2) = \frac{1}{2}\Big[\log\frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} - 1\Big]. \tag{B.7}$$

B.1.2 Geometric mean of two Gaussians

Given two Gaussian distributions $P_1 \sim \mathsf{N}\left(\mu_1, \sigma_1^2\right)$ and $P_2 \sim \mathsf{N}\left(\mu_2, \sigma_2^2\right)$, their geometric mean is $P_\lambda \sim \mathsf{N}\left(\mu_\lambda, \sigma_\lambda^2\right)$, where

$$\mu_\lambda = \frac{\lambda\mu_1\sigma_2^2 + (1-\lambda)\mu_2\sigma_1^2}{\lambda\sigma_2^2 + (1-\lambda)\sigma_1^2} \tag{B.8}$$

$$\sigma_\lambda^2 = \frac{\sigma_1^2\sigma_2^2}{\lambda\sigma_2^2 + (1-\lambda)\sigma_1^2}. \tag{B.9}$$

When $\mu_1 = \mu_2 = 0$, $P_\lambda \sim \mathsf{N}(0, \sigma_\lambda^2)$, and

$$D(P_\lambda \| P_1) = \frac{1}{2}\Big[\log\frac{\sigma_1^2}{\sigma_\lambda^2} + \frac{\sigma_\lambda^2}{\sigma_1^2} - 1\Big] \tag{B.10}$$

$$D(P_\lambda \| P_2) = \frac{1}{2}\Big[\log\frac{\sigma_2^2}{\sigma_\lambda^2} + \frac{\sigma_\lambda^2}{\sigma_2^2} - 1\Big]. \tag{B.11}$$

When $\sigma_1^2 = \sigma_2^2 = \sigma^2$, $P_\lambda \sim \mathsf{N}(\lambda\mu_1 + (1-\lambda)\mu_2, \sigma^2)$, and

$$D(P_\lambda \| P_1) = (1-\lambda)^2\frac{P}{2N} \tag{B.12}$$

$$D(P_\lambda \| P_2) = \lambda^2\frac{P}{2N}. \tag{B.13}$$

B.2 χ^2-distribution with Degree of Freedom k

Let

$$I_{\chi_k^2}(x) \triangleq \frac{x}{2} - \frac{k}{2} + \frac{k}{2}(\ln k - \ln x), \tag{B.14}$$

it is well known that

$$I_{\chi_k^2,\geq}(x) = \begin{cases} I_{\chi_k^2}(x) & x > k \\ 0 & x \leq k \end{cases} \tag{B.15}$$

$$I_{\chi_k^2,\leq}(x) = \begin{cases} +\infty & x \leq 0 \\ I_{\chi_k^2}(x) & 0 < x < k \\ 0 & x \geq k \end{cases} \tag{B.16}$$

We frequently refer to the special case $k = 1$, which has a simpler form:

$$I_{\chi_1^2}(x) = \frac{1}{2}(x - \ln x - 1). \tag{B.17}$$

■ B.3 Sum of Spherically Uniform Distribution and Gaussian Distribution

Let S^n be uniformly distributed on a n-dimensional sphere with radius $\sqrt{nP}$ and $Z^n \sim \mathsf{N}(\mathbf{0}, \mathbf{I}_k)$, where S^n and Z^n are statistically independent. In this section, we calculate the exponents $I_{\mathsf{SG},\geq}(P,\eta)$ and $I_{\mathsf{SG},\leq}(P,\eta)$, which are defined as

$$\mathbb{P}\left[\frac{1}{n}\|S^n + Z^n\|^2 \geq \eta\right] \doteq \exp\left[-nI_{\mathsf{SG},\geq}(P,\eta)\right] \tag{B.18}$$

$$\mathbb{P}\left[\frac{1}{n}\|S^n + Z^n\|^2 \leq \eta\right] \doteq \exp\left[-nI_{\mathsf{SG},\leq}(P,\eta)\right], \tag{B.19}$$

and the expressions for $I_{\mathsf{SG},\geq}(P,\eta)$ and $I_{\mathsf{SG},\leq}(P,\eta)$ are shown in (B.27) and (B.34).

■ B.3.1 Derivations for $I_{\mathsf{SG},\geq}(P,\eta)$

Let $s^n = (\sqrt{nP}, 0, 0, \cdots, 0)$. Note that $S^n + Z^n$ is spherically symmetric, we have

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{n}\|S^n + Z^n\|^2 \geq \eta\right] &= \mathbb{P}\left[\frac{1}{n}\|s^n + Z^n\|^2 \geq \eta\right] && \text{(B.20)}\\
&= \mathbb{P}\left[(\sqrt{nP} + Z_1)^2 + \sum_{i=2}^{n} Z_i^2 \geq n\eta\right] && \text{(B.21)}\\
&= \int_{-\infty}^{+\infty} f_{Z_1}(z_1)\mathbb{P}\left[\sum_{i=2}^{n} Z_i^2 \geq n\eta - (\sqrt{nP} + z_1)^2\right] dz_1 && \text{(B.22)}\\
(\text{let } t = z_1/\sqrt{n}) \quad &\doteq \int_{-\infty}^{+\infty} \exp\left(-n\frac{t^2}{2}\right) \mathbb{P}\left[\sum_{i=2}^{n} Z_i^2 \geq n\left[\eta - (\sqrt{P} + t)^2\right]\right] dt && \text{(B.23)}
\end{aligned}$$

Then by *Laplace's Principle*,

$$I_{\mathsf{SG},\geq}(P,\eta) = \min_t \frac{t^2}{2} + I_{\chi_1^2,\geq}\left(\eta - (\sqrt{P}+t)^2\right) \tag{B.24}$$

Some simplifications show that

$$I_{\mathsf{SG},\geq}(P,\eta) = \begin{cases} 0 & \eta \leq P+1 \\ \min\limits_{t\in[-\sqrt{P}-\sqrt{\eta-1}, -\sqrt{P}+\sqrt{\eta-1}]} \left[\frac{t^2}{2} + I_{\chi_1^2}\left(\eta - (\sqrt{P}+t)^2\right)\right] & \eta > P+1 \end{cases} \tag{B.25}$$

Solving the minimization problem (see Section B.3.3 for details) in (B.25) gives us the minimum value

$$I_{SG}(P,\eta) \triangleq \frac{1}{2}\left(P + \eta - \sqrt{1+4P\eta} - \log\left[\frac{\sqrt{1+4P\eta}-1}{2P}\right]\right) \tag{B.26}$$

then

$$I_{\mathsf{SG},\geq}(P,\eta) = \begin{cases} 0 & \eta \leq P+1 \\ I_{SG}(P,\eta) & \eta > P+1 \end{cases} \tag{B.27}$$

■ B.3.2 Derivations for $I_{\mathsf{SG},\le}(P,\eta)$

Again, using the spherical symmetry property of S^n+Z^n, let $s^n=(\sqrt{nP},0,\dots,0)$, we have

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{n}\|S^n+Z^n\|^2\le\eta\right] &= \mathbb{P}\left[\frac{1}{n}\|s^n+Z^n\|^2\le\eta\right] && \text{(B.28)}\\
&= \mathbb{P}\left[(\sqrt{nP}+Z_1)^2+\sum_{i=2}^{n}Z_i^2\le n\eta\right] && \text{(B.29)}\\
&= \int_{-\infty}^{+\infty} f_{Z_1}(z_1)\mathbb{P}\left[\sum_{i=2}^{n}Z_i^2\le n\eta-(\sqrt{nP}+z_1)^2\right]dz_1 && \text{(B.30)}\\
(\text{let } t=z_1/\sqrt{n}) \quad &\doteq \int_{-\infty}^{+\infty}\exp\left(-n\frac{t^2}{2}\right)\mathbb{P}\left[\sum_{i=2}^{n}Z_i^2\le n\left[\eta-(\sqrt{P}+t)^2\right]\right]dt && \text{(B.31)}
\end{aligned}$$

Then by *Laplace's Principle*, we have

$$I_{\mathsf{SG},\le}(P,\eta)=\min_t\left[\frac{t^2}{2}+I_{\chi_1^2,\le}\left(\eta-(\sqrt{P}+t)^2\right)\right] \tag{B.32}$$

Some simplifications show that

$$I_{\mathsf{SG},\le}(P,\eta)=\begin{cases}+\infty & \eta\le 0\\ \min\limits_{t\in[-\sqrt{P}-\sqrt{\eta},-\sqrt{P}+\sqrt{\eta}]}\left[\frac{t^2}{2}+I_{\chi_1^2}\left(\eta-(\sqrt{P}+t)^2\right)\right] & 0<\eta\le 1\\ \min\limits_{t\in[-\sqrt{P}+\sqrt{\eta-1},-\sqrt{P}+\sqrt{\eta}]}\left[\frac{t^2}{2}+I_{\chi_1^2}\left(\eta-(\sqrt{P}+t)^2\right)\right] & 1\le\eta<P+1\\ 0 & \eta\ge P+1\end{cases} \tag{B.33}$$

Solve the minimization problem (see Section B.3.3 for details) in (B.33) gives that when $0<\eta<P+1$, the minimum value is also $I_{SG}(P,\eta)$ and we have

$$I_{\mathsf{SG},\le}(P,\eta)=\begin{cases}+\infty & \eta\le 0\\ I_{SG}(P,\eta) & 0<\eta<P+1\\ 0 & \eta\ge P+1\end{cases} \tag{B.34}$$

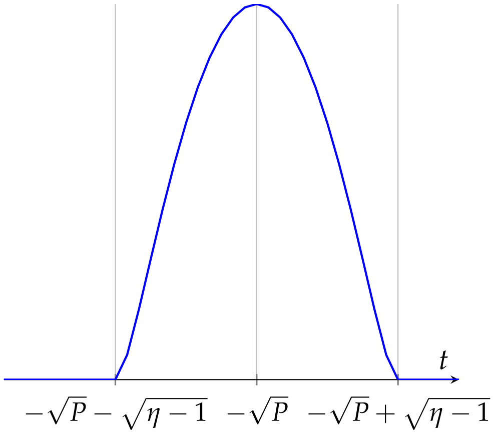


Figure B-1: $I_{\chi_1^2,\geq}\left(\eta-(\sqrt{P}+t)^2\right)$ with $\eta > P+1$.

■ B.3.3 Calculation details

Note when $\eta > 0$,

$$\eta-(\sqrt{P}+t)^2 > 0 \Longleftrightarrow t \in [-\sqrt{P}-\sqrt{\eta}, -\sqrt{P}+\sqrt{\eta}] \tag{B.35}$$

and when $\eta > 1$,

$$\eta-(\sqrt{P}+t)^2 > 1 \Longleftrightarrow t \in [-\sqrt{P}-\sqrt{\eta-1}, -\sqrt{P}+\sqrt{\eta-1}] \tag{B.36}$$

The simplifications that lead to (B.25) and (B.33) can be seen after observing Figure B-1 and Figure B-2.

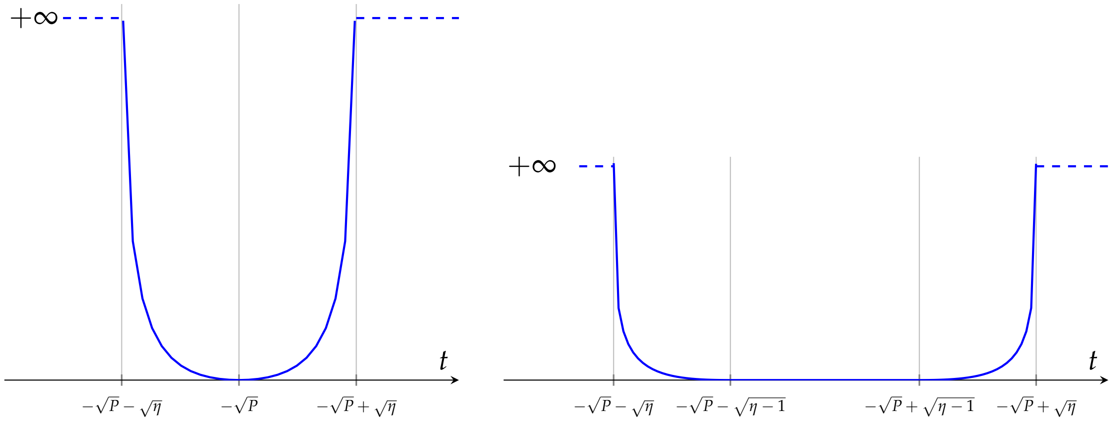
For the minimization problem in both (B.25) and (B.33), setting the derivative of the objective function to zero, we solve can for the roots and have

$$t_1 = -\frac{1+2P-\sqrt{1+4P\eta}}{2\sqrt{P}} \tag{B.37}$$

$$t_2 = -\frac{1+2P+\sqrt{1+4P\eta}}{2\sqrt{P}} \tag{B.38}$$

and we can verify that only t_1 is a true solution, because

$$\begin{cases} t_1 \in (-\sqrt{P}+\sqrt{\eta-1}, \sqrt{P}+\sqrt{\eta}) & \eta > 1 \\ t_1 \in (-\sqrt{P}-\sqrt{\eta}, \sqrt{P}+\sqrt{\eta}) & 0 < \eta \leq 1 \end{cases} \tag{B.39}$$



(a) $0 < \eta < 1$

(b) $1 < \eta < 1 + P$

Figure B-2: $I_{\chi_1^2,\le}\left(\eta - (\sqrt{P}+t)^2\right)$.

Therefore we can substitute t_1 and obtain (B.26).

■ B.4 Exponential Approximation for the Square of Gaussian Distributions

Let $X \sim Z^2$, where $Z \sim \mathsf{N}\left(\sqrt{n\lambda}, 1\right)$, then

$$\mathbb{P}\left[X = nt\right] = \mathbb{P}\left[Z = -\sqrt{nt} \text{ or } Z = \sqrt{nt}\right] \tag{B.40}$$

$$\doteq \mathbb{P}\left[Z = \sqrt{nt}\right] \tag{B.41}$$

$$\doteq \exp\left(-n\frac{(\sqrt{t}-\sqrt{\lambda})^2}{2}\right). \tag{B.42}$$

■ B.5 Calculation Details for Heuristic Detection Rules

In this section we document the calculation details for the results in Section 5.3.3. We let $Z_i \overset{i.i.d.}{\sim} \mathsf{N}(0,1)$ in this section.

■ B.5.1 Detection rule #1

$$\begin{aligned} P_{\mathrm{f}} &= \mathbb{P}\left[aZ_1 + b\|Z_2^n\| \geq \sqrt{n}\eta\right] && \text{(B.43)} \\ &\doteq \int_{-\infty}^{+\infty} \mathbb{P}\left[Z_1 = \sqrt{n}\frac{r}{a}\right] \mathbb{P}\left[\|Z_2^n\|^2 \geq n\left(\frac{\eta - r}{b}\right)^2\right] dr && \text{(B.44)} \\ &\doteq \int_{-\infty}^{+\infty} \exp\left[-n(\frac{r^2}{2a^2})\right] \exp\left[-n\left(I_{\chi_1^2,\geq}\left(\frac{(\eta - r)^2}{b^2}\right)\right)\right] dr && \text{(B.45)} \\ &\doteq \exp\left[-ne_f(\eta, a, b)\right] && \text{(B.46)} \end{aligned}$$

where

$$e_f(\eta, a, b) = \min_{r \geq 0} \frac{r^2}{2a^2} + I_{\chi_1^2,\geq}\left(\frac{(\eta - r)^2}{b^2}\right) \tag{B.47}$$

Some further simplifications show

$$e_f(\eta, a, b) = \begin{cases} 0 & \eta \leq b \\ \min\limits_{0 \leq r \leq \eta - b} \dfrac{r^2}{2a^2} + I_{\chi_1^2}\left(\dfrac{(\eta - r)^2}{b^2}\right) & \eta > b \end{cases} \tag{B.48}$$

$$\begin{aligned} P_{\mathrm{m}} &= \mathbb{P}\left[a(\sqrt{nP_s} + Z_1) + b\|S_2^n + Z_2^n\| \leq \sqrt{n}\eta\right] && \text{(B.49)} \\ &\doteq \int_{-\infty}^{+\infty} \mathbb{P}\left[\mathsf{N}\left(\sqrt{nP_s}, 1\right) = \sqrt{n}\frac{r}{a}\right] \mathbb{P}\left[\|S_2^n + Z_2^n\| \leq \sqrt{n}\frac{\eta - r}{b}\right] dr && \text{(B.50)} \\ &\doteq \int_{-\infty}^{+\infty} \exp\left[-n\frac{(r - a\sqrt{P_s})^2}{2a^2}\right] \exp\left[-n\left(I_{\mathsf{SG},\leq}\left(P_c, \frac{(\eta - r)^2}{b^2}\right)\right)\right] dr && \text{(B.51)} \\ &\doteq \exp\left[-ne_m(\eta, a, b)\right] && \text{(B.52)} \end{aligned}$$

where

$$e_m(\eta, a, b) = \min_{r > 0} \frac{(r - a\sqrt{P_s})^2}{2a^2} + I_{\mathsf{SG},\leq}\left(P_c, \frac{(\eta - r)^2}{b^2}\right) \tag{B.53}$$

Some further simplifications show

$$e_m(\eta, a, b) = \begin{cases} +\infty & \eta \leq 0 \\ \min\limits_{\eta - b\sqrt{P_c+1} \leq r \leq \eta} \dfrac{(r - a\sqrt{P_s})^2}{2a^2} + I_{\mathsf{SG}}\left(P_c, \dfrac{(\eta - r)^2}{b^2}\right) & 0 < \eta < a\sqrt{P_s} + b\sqrt{P_c+1} \\ 0 & \eta \geq a\sqrt{P_s} + b\sqrt{P_c+1} \end{cases} \tag{B.54}$$

■ B.5.2 Detection rule #2

$$P_\mathrm{f} = \mathbb{P}\left[a\|Z_1\|^2 + b\|Z_2^n\|^2 \geq n\eta\right] \tag{B.55}$$

$$\doteq \int_0^{+\infty} \mathbb{P}\left[\|Z_1\|^2 = n\frac{r}{a}\right] \mathbb{P}\left[\|Z_2^n\|^2 \geq n\frac{\eta - r}{b}\right] dr \tag{B.56}$$

$$\doteq \int_0^{+\infty} \exp\left[-n(\frac{r}{2a})\right] \exp\left[-n\left(I_{\chi_1^2,\geq}\left(\frac{\eta - r}{b}\right)\right)\right] dr \tag{B.57}$$

$$\doteq \exp\left[-ne_f(\eta, a, b)\right] \tag{B.58}$$

where

$$e_f(\eta, a, b) = \min_{r \geq 0} \frac{r}{2a} + I_{\chi_1^2,\geq}\left(\frac{\eta - r}{b}\right) \tag{B.59}$$

Some further simplifications show

$$e_f(\eta, a, b) = \begin{cases} 0 & \eta \leq b \\ \min\limits_{0 \leq r \leq \eta - b} \dfrac{r}{2a} + I_{\chi_1^2}\left(\dfrac{\eta - r}{b}\right) & \eta > b \end{cases} \tag{B.60}$$

$$P_\mathrm{m} = \mathbb{P}\left[a\|\sqrt{nP_s} + Z_1\|^2 + b\|S_2^n + Z_2^n\|^2 \leq n\eta\right] \tag{B.61}$$

$$\doteq \int_0^{+\infty} \mathbb{P}\left[\|\mathsf{N}\left(\sqrt{nP_s}, 1\right)\|^2 = n\frac{r}{a}\right] \mathbb{P}\left[\|S_2^n + Z_2^n\|^2 \leq n\frac{\eta - r}{b}\right] dr \tag{B.62}$$

$$\doteq \int_0^{+\infty} \exp\left[-n\frac{(\sqrt{r/a} - \sqrt{P_s})^2}{2}\right] \exp\left[-n\left(I_{\mathsf{SG},\leq}\left(P_c, \frac{\eta - r}{b}\right)\right)\right] dr \tag{B.63}$$

$$\doteq \exp\left[-ne_m(\eta, a, b)\right] \tag{B.64}$$

where

$$e_m(\eta, a, b) = \min_{r>0} \frac{(\sqrt{r/a} - \sqrt{P_s})^2}{2} + I_{\mathsf{SG},\leq}\left(P_c, \frac{\eta - r}{b}\right) \tag{B.65}$$

Some further simplifications show

$$e_m(\eta, a, b) = \begin{cases} +\infty & \eta \leq 0 \\ \min\limits_{r_l \leq r \leq r_u} \dfrac{(\sqrt{r/a} - \sqrt{P_s})^2}{2} + I_{\mathsf{SG}}\left(P_c, \dfrac{\eta - r}{b}\right) & 0 < \eta < aP_s + b(P_c + 1) \\ 0 & \eta \geq aP_s + b(P_c + 1) \end{cases} \tag{B.66}$$

where $r_l = \max\{0, \eta - b(P_c + 1)\}$ and $r_u = \min\{\eta, aP_s\}$.

Bibliography

[1] R. H. Barker, "Group synchronization of binary digital systems," *Communication Theory*, pp. 273–287, 1953.

[2] S. Golomb, J. Davey, I. Reed, H. V. Trees, and J. Stiffler, "Synchronization," *Communications Systems, IEEE transactions on*, vol. 11, no. 4, pp. 481–491, 1963.

[3] L. Franks, "Carrier and bit synchronization in data Communication–A tutorial review," *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 28, no. 8, pp. 1107–1121, 1980.

[4] R. Scholtz, "Frame synchronization techniques," *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 28, no. 8, pp. 1204–1213, 1980.

[5] J. Massey, "Optimum frame synchronization," *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 20, no. 2, pp. 115–119, 1972.

[6] G. Lui and H. Tan, "Frame synchronization for gaussian channels," *Communications, IEEE Transactions on [legacy, pre - 1988]*, vol. 35, no. 8, pp. 818–829, 1987.

[7] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4508–4528, 2009.

[8] ——, "On the capacity region of asynchronous channels," in *Proceedings of IEEE International Symposium on Information Theory*, Toronto, ON, Canada, 2008, pp. 1213–1217.

[9] C. E. Shannon and W. Weaver, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, p. 379423, 1948.

[10] C. E. Shannon, "Probability of error for optimal codes in a gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, May 1959.

[11] V. Chandar, A. Tchamkerten, and G. W. Wornell, “Training-based schemes are suboptimal for high rate asynchronous communication,” in *Proceedings of IEEE Information Theory Workshop*, Taormina, Italy, 2009, pp. 389 –393.

[12] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. New York: Academic Press, 1981.

[13] I. Csiszár, “The method of types,” *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2505–2523, 1998.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, Jul. 2006.

[15] M. J. E. Golay, “Note on the theoretical efficiency of information reception with PPM,” *Proc. IRE*, vol. 37, p. 1031, 1949.

[16] H. Witsenhausen, “Entropy inequalities for discrete channels,” *Information Theory, IEEE Transactions on*, vol. 20, no. 5, pp. 610–616, 1974.

[17] Y. Polyanskiy, H. V. Poor, and S. Verdú, “New channel coding achievability bounds,” in *2008 IEEE International Symposium on Information Theory*, Toronto, ON, Canada, 2008, pp. 1763–1767.

[18] S. Borade, B. Nakiboğlu, and L. Zheng, “Unequal error protection: An Information-Theoretic perspective,” *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5511–5539, 2009.