On the Optimality of the Hamming Metric for Decoding Block Codes over Binary Additive Noise Channels

by

Ghady Azar

A thesis submitted to the
Department of Mathematics and Statistics
in conformity with the requirements for
the degree of Master of Applied Science

Queen's University
Kingston, Ontario, Canada
July 2013

Abstract

Most of the basic concepts of algebraic coding theory are derived for the memoryless binary symmetric channel. These concepts do not necessarily hold for time-varying channels or for channels with memory. However, errors in real-life channels seem to occur in bursts rather than independently, suggesting that these channels exhibit some statistical dependence or memory. Nonetheless, the same algebraic codes are still commonly used in current communication systems that employ interleaving to spread channel error bursts over the set of received codewords to make the channel appear memoryless to the block decoder. This method suffers from immediate shortcomings as it fails to exploit the channel's memory while adding delay to the system.

We study optimal maximum likelihood block decoding of binary codes sent over several binary additive channels with infinite and finite memory. We derive conditions on general binary codes and channels parameters under which maximum likelihood and minimum distance decoding are equivalent. The channels considered in this work are the infinite and finite memory Polya contagion channels [1], the queue-based channel [29], and the Gilbert-Elliott channel [9, 12]. We also present results on the optimality of classical perfect and quasi-perfect codes when used over the aforementioned channels under maximum likelihood decoding.

Acknowledgments

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr. Fady Alajaji, whose expertise, encouragement and understanding considerably enriched my graduate experience. I honestly could not have hoped for a better supervisor. I highly appreciate his vast knowledge, his attention to detail and his contribution to this work. I also cannot thank him enough for being very patient and for giving me the moral support and the freedom I needed to figure out my future plans. I am also grateful to my family and friends for their unconditional love and invaluable support throughout my studies.

Contents

Abstract **i**

Acknowledgments **ii**

List of Figures **v**

List of Tables **v**

List of Acronyms **x**

1 Introduction **1**

1.1 Problem Description and Thesis Contribution 1

1.2 Literature Review 4

1.3 Thesis Overview 5

2 Communication Channel Models **7**

2.1 Binary Memoryless Channels 8

2.2 Binary Channels with Memory 9

2.2.1 Binary First-Order Markov Noise Channel 9

2.2.2 Polya Contagion Channels 11

2.2.3 Queue-Based Channel 15

2.2.4 Gilbert-Elliott Channel 21

3 Channel Coding 24

3.1 Binary Block Codes over Binary Additive Noise Channels 26

3.2 Binary Linear Codes over Binary Additive Noise Channels 30

3.3 Probabilistic Decoding over Binary Additive Noise Channels . . . 32

3.3.1 Maximum A-Posteriori (MAP) Decoder 32

3.3.2 Maximum Likelihood (ML) Decoder 32

3.3.3 Minimum Hamming Distance (MD) Decoder 33

3.4 Hamming Codes 34

3.5 The Binary Golay Code 35

3.6 Reed-Muller Codes 35

3.7 Construction of Nonlinear Perfect Codes 37

3.8 Generalized Concepts of Coding Theory 39

4 Optimality of MD Decoding over Infinite Memory Channels 48

4.1 Properties of ML Decoding over the IMCC 49

4.2 On the Equivalence of MD and ML Decoding over the IMCC . . . 52

4.3 Numerical Results 58

4.4 On the Equivalence of MD and ML decoding over the GEC 60

4.4.1 Case 1: State Vector is Unknown at the Decoder 61

4.4.2 Case 2: State Vector is Known at the Decoder 65

5 Optimality of MD decoding over the BFMNC 69

5.1 Block Transition Probability for the BFMNC 70

5.2 On the Equivalence of SMD and ML decoding over the BFMNC . 71

5.3 Numerical Results 78

5.4 A Closer Look at Theorems 5.1 and 5.2 80

6 Optimality of MD Decoding over the QBC 84

6.1 On the equivalence of MD and ML Decoding of **n**-block codes over the QBC with $\mathbf{M} \geq \mathbf{n}$ 85

6.2 Block Transition Probability for the QBC with $\mathbf{M} = \mathbf{2}$ 87

6.3 On the Equivalence of SMD and ML Decoding over the QBC with $\mathbf{M} = \mathbf{2}$ 91

6.4 Numerical Results 105

7 Conclusion and Future Work 108

7.1 Concluding Remarks 108

7.2 Application to Syndrome Source Coding 109

7.3 Future Work 110

List of Tables

4.1 Verifying Theorem 4.3 and Lemma 4.3 for the $[7,4,3]$ Hamming code over the IMCC with parameters $p = 0.1$ and ϵ. 59

4.2 Verifying Theorem 4.3 and Lemma 4.3 for the $[8,4,4]$ Reed-Muller code $\mathcal{RM}(1,3)$ over the IMCC with parameters $p = 0.1$ and ϵ. . . 59

4.3 Verifying Theorem 4.3 and Lemma 4.4 for the $(15, 2^{11}, 3)$ Vasil'ev nonlinear code over the IMCC with parameters $p = 0.1$ and ϵ. . . 60

5.1 Verifying Theorem 5.2 for the $[7,4,3]$ Hamming code over the BFMNC with parameters $p = 0.1$ and ϵ. 79

5.2 Verifying Theorem 5.2 for the $(15, 2^{11}, 3)$ Vasil'ev nonlinear perfect code over the BFMNC with parameters $p = 0.1$ and ϵ. 79

5.3 Verifying Theorem 5.2 for the $[8,4,4]$ Reed Muller code $\mathcal{RM}(1,3)$ over the BFMNC with parameters $p = 0.1$ and ϵ. 80

5.4 Verifying Theorem 5.2 for the $[24,12,8]$ extended Golay code over the BFMNC with parameters $p = 0.1$ and ϵ. 80

6.1 Verifying Theorem 6.5 for the $[7,4,3]$ Hamming code over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ. 106

6.2 Verifying Theorem 6.5 for the $(15, 2^{11}, 3)$ Vasil'ev nonlinear perfect code over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ. 107

6.3 Verifying Theorem 6.5 for the $[8, 4, 4]$ Reed Muller code $\mathcal{RM}(1, 3)$ over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ. 107

6.4 Verifying Theorem 6.5 for the $[24, 12, 8]$ extended Golay code over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ. 107

List of Figures

2.1 The binary symmetric channel with cross over probability p (BSC(p)). 9

2.2 A queue of length M. 15

2.3 The Gilbert-Elliott channel model. 21

4.1 Plot of the condition in Theorem 4.2: maximum allowable r_{cov} over the IMCC with respect to the correlation coefficient ϵ and for different values of the BER p. 53

4.2 Plot of $\hat{m}(n, P_G, P_B)$ for the GEC with respect to P_B for different values of n and for $P_G = 0.001$. 65

5.1 Plot of $m^*(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.001$. 80

5.2 Plot of $m^*(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.01$. 81

5.3 Plot of $m^*(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.1$. 81

6.1 Plot of $\tilde{m}(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.001$, FMCC with $M = 2$. 102

6.2 Plot of $\tilde{m}(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.01$, FMCC with $M = 2$. 103

6.3 Plot of $\tilde{m}(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.1$, FMCC with $M = 2$. 103

6.4 Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of α, for $p = 0.001$ and $n = 10$, QBC with $M = 2$. 104

6.5 Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of α, for $p = 0.01$ and $n = 10$, QBC with $M = 2$. 105

6.6 Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of α, for $p = 0.1$ and $n = 10$, QBC with $M = 2$. 105

7.1 Syndrome source coding method of using error correcting codes for data compression 109

7.2 Source encoder with side information at the receiver 110

List of Acronyms

BER	Bit Error Rate
BFMNC	Binary First-Order Markov Noise Channel
BSC	Binary Symmetric Channel
FMCC	Finite Memory Contagion Channel
GEC	Gilbert-Elliott Channel
IMCC	Infinite Memory Contagion Channel
MAP	Maximum A-Posteriori
MD	Minimum Distance
ML	Maximum Likelihood
PCE	Probability of Codeword Error
QBC	Queue-Based Channel
SMD	Strict Minimum Distance

Chapter 1

Introduction

1.1 Problem Description and Thesis Contribution

Before Shannon's 1948 paper "A Mathematical Theory of Communication" [19], communication was strictly an engineering discipline with little mathematical theory to support it. It was also widely believed that the only way to reliably transmit information over a noisy medium was by reducing the data rate (e.g., by retransmitting the message) until the error probability becomes "small enough". Hence, sending information over a noisy channel with a negligible probability of error and at a positive rate was thought to be impossible to achieve. In his paper [19], Shannon showed that this belief is incorrect by proving that every channel has a maximum rate for transmitting data reliably known as the channel capacity. He also proved that this theoretical limit can be achievable by a "more intelligent" coding of the information without proposing an explicit method to construct such codes, hence opening the field of coding theory which aims to discover efficient codes capable of achieving the channel capacity.

The fundamental results in coding theory are derived under the assumption that the communication channel is a binary symmetric channel (BSC). In fact, for this memoryless channel model, coding theorists were able to benefit from the extensive literature of abstract algebra to develop good codes with rich algebraic structures. However, these codes are not necessarily good for other channel models. Moreover, most real life channels have statistical memory which cannot be modeled by the BSC. Memory has not been efficiently exploited in current communication systems despite proving that it increases the channel capacity in several cases. Instead, interleaving is most commonly used to spread channel error bursts over the set of received codewords so that block decoding can overcome most of the corrupted codewords (if the number of channel errors within a codeword is within the code's error correcting capability). In other words, the use of interleaving makes the channel appear memoryless to the block decoder. This method has immediate shortcomings as it fails to exploit the channel memory while adding delay to the system.

It is well known that the maximum likelihood (ML) decoding of binary codes over the memoryless binary symmetric channel (BSC) with crossover probability $p < \frac{1}{2}$ is equivalent to minimum Hamming distance decoding. When the communication channel has memory, the above equivalence does not necessarily hold. Hence, it is natural to investigate whether a similar relation exists for channels with memory. In this work, we derive conditions on codes and on the channel characteristics, under which the equivalence holds. The channel models considered in this work are the infinite and finite memory Polya-contagion channels introduced in [1], the Gilbert-Elliott channel (GEC) introduced in [9, 12] and the queue-based channel

(QBC) introduced in [29].

The infinite memory contagion channel (IMCC) is a non-ergodic binary additive channel which can be used in modeling non-ergodic fading channels [8, 21]. It has a closed-form expression for its epsilon-capacity and admits a simple ML decoding rule. For this channel model, we show both necessary and sufficient conditions for which minimum distance (MD) and ML decoding are equivalent. We also give sufficient conditions under which classical perfect and quasi-perfect codes are optimal under ML decoding over the IMCC.

Alternatively, the finite memory contagion channel (FMCC) and the QBC, its queue-based extension, both feature stationary and ergodic Markov noise processes of order M, and they were shown to accurately model ergodic correlated Rayleigh and Rician fading channels [16, 29, 30]. Furthermore, it has been recently observed that, in the context of LDPC coding, iterative decoders designed for these channels can outperform the theoretical limit that is achievable on the equivalent BSC (realized via ideal interleaving) [17] (see also [7] and [11] for decoders designed for Gilbert-Elliott and finite-state Markov channels). Since the n-fold block transition probability for the QBC admits two different expressions depending on whether $M \geq n$ or $M < n$, we treat these two cases separately. For the case when $M \geq n$, we show both necessary and sufficient conditions for which minimum distance and ML decoding are equivalent. We also give sufficient conditions under which classical perfect and quasi-perfect codes are optimal under ML decoding over the QBC. For the case when $n > M$, we restrict our study to the QBC with $M = 1$ (or equivalently the binary first-order Markov noise channel or the FMCC with $M = 1$) and $M = 2$. For both cases, we determine sufficient conditions on any binary code under which strict minimum Hamming distance decoding is equivalent to strict ML decoding. We also present sufficient conditions under which classical perfect codes are optimal under ML decoding over the QBC. We specialize the results derived in the latter case for the FMCC with $M = 2$ which is a special case of the QBC with $M = 2$.

Finally, the GEC is one of the most widely used binary channel models in the literature (belonging to the class of finite-state Markov channels [10]) for describing burst error patterns in real communication channels. This channel is governed by an underlying two-state hidden Markov model where one state, denoted by G, represents the "good" state of the channel and the other state, denoted by B, represents its bad state. The GEC noise process is a stationary ergodic hidden Markov source (of infinite memory). We study separately the case when the state vector is unknown and known at the decoder. In the former case, we present sufficient conditions on binary codes under which strict MD decoding and ML decoding are equivalent. We also determine sufficient conditions under which classical perfect codes are optimal under ML decoding over the GEC. In the second case (when the state vector is available at the decoder), we present partial results pertaining to the equivalence between the Hamming weight and the likelihood of error patterns.

1.2 Literature Review

In related works [**?**], it was proven that strict minimum Hamming distance decoding is equivalent to strict ML decoding for perfect codes of minimum distance 3 over the first-order Markov channel (finite memory contagion channel with $M = 1$) with a positive correlation coefficient. In [2], sufficient conditions, under which strict minimum Hamming distance decoding of binary linear perfect codes becomes equivalent to strict ML decoding, are derived for the same channel. A near equivalence relationship between strict MD and strict ML decoding is also obtained for binary linear quasi-perfect codes for a range of channel parameters and the codes' minimum distance. We extend the provided conditions to obtain even tighter sufficient conditions that apply for any binary code (linear or nonlinear). We also provide similar results for the finite memory contagion channel with $M = 2$. In an another work [5], a sufficient condition on the infinite memory contagion channel is provided, under which ML block decoding is equivalent to minimum Hamming distance block decoding for linear codes containing the all-one codeword. We also improve these results by obtaining necessary and sufficient conditions for any binary codes used over the same channel.

1.3 Thesis Overview

In Chapter 2, we give a brief introduction of the communication channel models considered in this work. Specifically, we discuss the binary symmetric channel (BSC), the binary first-order Markov noise channel (BFMNC), the infinite and finite memory contagion channel (IMCC and FMCC), and the Gilbert-Elliott channel (GEC).

In Chapter 3, we provide an overview of the basic concepts in coding theory. We then specialize these concepts to linear block codes before briefly introducing some binary codes such as the family of Hamming codes, the Golay code and the Reed-Muller codes. We also explain a method of generating nonlinear perfect codes from Hamming codes. Finally, we introduce a generalized likelihood distance that is associated with a wide class of channels that includes all additive noise channels. We generalize the basic notions in coding theory to this new distance and we present two lemmas that determine sufficient conditions on error patterns under which classical perfect and quasi-perfect codes are also generalized perfect and quasi-perfect codes.

In Chapter 4, we study ML decoding of binary block codes over the IMCC and the GEC. We derive useful conditions on binary codes under which we obtain some equivalence between ML and MD decoding.

In Chapter 5, we study ML decoding of binary codes over the BFMNC. We also derive sufficient conditions on these codes under which SMD and ML decoding are equivalent.

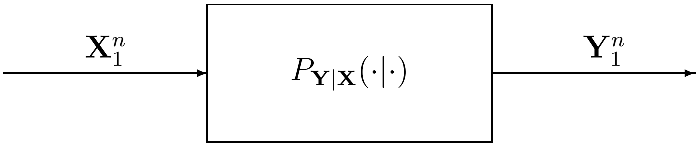
In Chapter 6, we study separately ML decoding of length-n binary codes over the QBC with memory $M \geq n$ and $M < n$. For the latter case, we restrict our treatment to the QBC with $M = 2$. For both cases, we derive useful conditions on binary codes under which we obtain some equivalent between MD and ML decoding.

Finally, in Chapter 7, we summarize our results and suggest some possible future work.

Chapter 2

Communication Channel Models

By definition, the communication channel is the connection between the transmitter and the receiver in a given communication system. It may take the form of an optical fiber, data storage device, free space, etc... With all this variety, a common problem in any communication system is the noise interference which is usually modeled as an additive process. The noise may arise from internal components at the receiver as well as from outside interference from other users of the channels. Most communication systems use quantizers at the receiver to make the number of possible received signals finite. In this case, the noise is described by a probability transition matrix $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_1^n = \mathbf{y}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n)$ which characterizes the conditional distribution of the output Y given the input X. Formally, a communication channel can be defined by the sequence $\left\{\mathcal{X}^n, P_{\mathbf{Y}|\mathbf{X}}(.|.), \mathcal{Y}^n\right\}_{n=1}^{\infty}$, where $\mathcal{X}, \mathcal{Y}$ are the sets of possible channel input and output symbols, respectively.



In other words, given the n-tuples $\mathbf{x}_1^n = (x_1, ...x_n) \in \mathcal{X}^n$ and $\mathbf{y}_1^n = (y_1, ..., y_n) \in \mathcal{Y}^n$, then $P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{Y}_1^n = \mathbf{y}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n\right)$ denotes the probability that $\mathbf{y}_1^n$ will be received given that $\mathbf{x}_1^n$ was transmitted. An additive noise channel is a channel where the output at time k is given by: $y_k = x_k + z_k$, where $x_k \in \mathcal{X}, z_k \in \mathcal{Z}$ are the input and the noise symbols at time k, respectively. The noise process is assumed to be independent of the input of the channel. Hence, $P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{Y}_1^n = \mathbf{y}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n\right) = P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{X}_1^n + \mathbf{Z}_1^n = \mathbf{y}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n\right) = P\left(\mathbf{Z}_1^n = \mathbf{y}_1^n - \mathbf{x}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n\right) = P(\mathbf{Z}_1^n = \mathbf{y}_1^n - \mathbf{x}_1^n)$, where the addition and the subtraction of the vectors are respectively the component-wise addition and subtraction defined for the sets $\mathcal{X}, \mathcal{Y}$. This work considers only binary additive channels where $\mathcal{X} = \mathcal{Y} = \mathcal{Z} =$GF(2) $= \{0, 1\}$ and addition is modulo-2. For simplicity, we will reserve the notation $\mathbf{F} =$ GF(2) throughout.

2.1 Binary Memoryless Channels

We first consider the simplest mathematical model for communication channels. A binary additive noise channel is called memoryless iff, for any two vectors $\mathbf{y}_1^n$ and $\mathbf{x}_1^n \in \mathbf{F}^n$, $P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_1^n = \mathbf{y}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n) = \prod_{i=1}^n P_{Y|X}(Y_i = y_i | X_i = x_i) = \prod_{i=1}^n P(Z_i = y_i \oplus x_i)$, where $\oplus$ denotes the modulo-2 addition. Let $P(Z_i = 1) = p = 1 - P(Z_i = 0)$, this channel is known as the binary symmetric channel with crossover probability p or BSC(p) shown in Fig. 2.1. In that case, the conditional probability of the output given the input can be further reduced to:

$$\begin{aligned} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}_1^n = \mathbf{y}_1^n | \mathbf{X}_1^n = \mathbf{x}_1^n) &= p^{w_H(\mathbf{y}_1^n \oplus \mathbf{x}_1^n)} (1-p)^{n - w_H(\mathbf{y}_1^n \oplus \mathbf{x}_1^n)} \\ &= (1-p)^n \left(\frac{p}{1-p}\right)^{w_H(\mathbf{y}_1^n \oplus \mathbf{x}_1^n)}, \end{aligned} \tag{2.1}$$

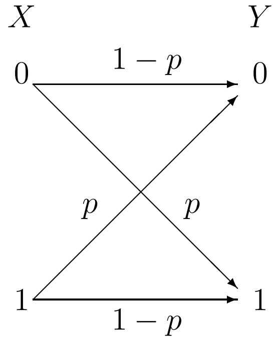


Figure 2.1: The binary symmetric channel with cross over probability p (BSC(p)).

where $w_H(\mathbf{z}_1^n)$ denotes the Hamming weight of the binary vector $\mathbf{z}_1^n$. If we make the unrestrictive assumption that $p < 1/2$, we can see that the likelihood of an error word is inversely proportional to its Hamming weight. This is one of the simplest and most studied channel models. In fact, most of the basic concepts of coding theory are derived for this particular communication channel.

2.2 Binary Channels with Memory

For the memoryless channel discussed in Section 2.1, the likelihood of a future error event is completely independent of the previous outcomes. Most real-life channels of interest are far from being memoryless. Memory in channel models is a result of multipath propagation, intersymbol interference, fading, etc... Errors in these channels seem to come in bursts rather than independently. We will next discuss several channel models with memory.

2.2.1 Binary First-Order Markov Noise Channel

The binary first-order Markov noise channel (BFMNC) is one of the simplest binary additive noise channels with memory. The noise process $\{Z_i\}_{i\in\mathbb{N}^*}$ generated in this channel forms a Markov chain, where $\mathbb{N}^*$ is the set of all natural numbers excluding 0. In other words, for any $k > 1$, $P(Z_k = z_k | Z_{k-1} = z_{k-1}, ..., Z_1 = z_1) = P(Z_k = z_k | Z_{k-1} = z_{k-1})$, for all $\mathbf{z}_1^k \in \mathbf{F}_2^k$. Hence, the probability distribution of the error word $\mathbf{Z}_1^n$ is given by:

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= P(Z_1 = z_1) \prod_{i=2}^{n} P(Z_i = z_i | Z_{i-1} = z_{i-1}, ..., Z_1 = z_1) \\ &= P(Z_1 = z_1) \prod_{i=2}^{n} P(Z_i = z_i | Z_{i-1} = z_{i-1}). \end{aligned}$$

If the channel noise is stationary, then it can be fully characterized by two parameters λ and q. We can associate with the channel the following transition matrix:

$$\mathbf{P} = \begin{bmatrix} 1-\lambda & \lambda \\ 1-q & q \end{bmatrix}. \tag{2.2}$$

The $(i,j)^{\text{th}}$ entry in $\mathbf{P}$ represents the probability that the current noise bit is j given that the previous noise bit was i, where $(i,j) \in \{0,1\}^2$. If $(\lambda, q) \in (0,1)^2$, then this Markov chain is irreducible and admits a unique stationary distribution $\boldsymbol{\pi} = [1-p, p] = \left[\frac{1-q}{1-q+\lambda}, \frac{\lambda}{1-q+\lambda}\right]$, where $p = P(Z_k = 1) \in (0,1)$. We also define the noise correlation coefficient $\epsilon = \frac{\text{Cov}(Z_k, Z_{k-1})}{\text{Var}(Z_k)} \in (-1,1)$. We can re-write the transition matrix $\mathbf{P}$, defined in (2.2), as follows:

$$\mathbf{P} = \begin{bmatrix} \epsilon + (1-\epsilon)(1-p) & (1-\epsilon)p \\ (1-\epsilon)(1-p) & \epsilon + (1-\epsilon)p \end{bmatrix}. \tag{2.3}$$

The range of the bit error rate (BER) p can be reduced to $(0, 1/2)$ without loss of generality. As we mentioned earlier, the channel models with memory attempt to generate errors in bursts to replicate the trend observed in real-life channels. Hence, the occurrence of an error should increase the likelihood of future errors. To capture this effect, we need to have $q \geq \lambda$ or $\epsilon \geq 0$. Hence, we will assume that $\epsilon \in [0, 1)$.

2.2.2 Polya Contagion Channels

The binary contagion channel model is introduced in [1]. The errors propagate in the channel in a way similar to the spread of a contagious disease through a population in the sense that the event of a bit error increases the probability of a future bit error. It has been shown in [20] that the distribution of defects in semiconductor memory is well modeled by the Polya-Eggenberger distribution which is a distribution realized by Polya's contagion model. This family of channels presents an alternative to the Gilbert-Elliott channel [9, 12] which belongs to the class of finite-state Markov channels [10] and has some attractive properties. Two different contagion channel models are proposed in [1], an infinite-memory model and a finite-memory model. The latter is obtained via a modification of the first. We study both models separately.

Infinite-Memory Contagion Channel

The infinite-memory contagion channel (IMCC) is a communication channel with stationary non-ergodic additive noise. The corresponding noise process $\{Z_i\}_{i=1}^{\infty}$ is generated by Polya's contagion urn scheme as follows: consider an urn that contains $R > 0$ red balls and $S > 0$ black balls. We denote by p the proportion of red balls, i.e., $p = \frac{R}{R+S}$. We assume without loss of generality that $p < 1/2$. We make successive draws with replacement to the urn, adding an additional $\Delta > 0$ balls of the same color just drawn. When we set $\Delta = 0$, the draws are independent and the resulting channel is nothing but the familiar memoryless BSC. We define $\delta = \frac{\Delta}{R+S}$. We associate with this scheme the noise process $\{Z_i\}_{i=1}^{\infty}$ as follows:

$$Z_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ ball drawn is red} \\ 0, & \text{otherwise.} \end{cases}$$

The probability of an n-bit error pattern $\mathbf{z}_1^n$ can be written as follows:

$$\begin{aligned} \mathrm{P}(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= \frac{p(p+\delta)...(p+(d-1)\delta)(1-p)(1-p+\delta)...(1-p+(n-d-1)\delta)}{(1+\delta)(1+2\delta)...(1+(n-1)\delta)} \\ &= \frac{\Gamma(\frac{1}{\delta})\Gamma(\frac{p}{\delta}+d)\Gamma(\frac{1-p}{\delta}+n-d)}{\Gamma(\frac{p}{\delta})\Gamma(\frac{1-p}{\delta})\Gamma(\frac{1}{\delta}+n)} \end{aligned} \tag{2.4}$$

where d is the Hamming weight of the error pattern, and $\Gamma(\cdot)$ is the Gamma function given by:

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt, \qquad x > 0.$$

The correlation between any two distinct noise bits is given by:

$$\epsilon = \mathrm{Cor}(Z_i, Z_j) = \frac{\mathrm{Cov}(Z_i, Z_j)}{\mathrm{Var}(Z_i)} = \frac{\delta}{1+\delta} \quad \forall i \neq j.$$

It is shown in [1] that, for this channel, the maximum likelihood (ML) decoding reduces to either minimum Hamming distance (MD) decoding or maximum Hamming distance decoding. We will discuss this in more details in later chapters. It is also proven in [1] that the all-zero error word $\mathbf{0}^n$ is the most likely among all error words of length n generated by the IMCC.

When we set $\delta = 0$, the IMCC reduces to a BSC with BER p. It is also proven that the IMCC (for $\delta > 0$) belongs to the class of averaged channels with memory, has zero capacity and admits a closed-form expression for its epsilon-capacity. We can see that the probability distribution of an error pattern depends only on its Hamming weight, and does not depend on how the errors are clustered. The fact that the first noise sample have the same effect as "more recent" noise samples on future outcomes poses limitations for modeling real-life communication systems where errors seem to be more temporally localised.

Finite-Memory Contagion Channel

The finite-memory contagion channel (FMCC) is a communication channel model derived from the IMCC. As previously mentioned, the identical contribution of the "old" noise samples and the "more recent" ones to future samples renders the IMCC inadequate for modeling real-life channels that behave ergodically. The FMCC addresses this problem by making the current noise sample independent of "older" samples given the last M error bits. In fact, the FMCC noise process is a binary additive M^{th}-order Markov chain that can be generated using a slightly modified version of the Polya contagion urn scheme. The only difference is that the added $\Delta \geq 0$ balls at the i^{th} draw are removed from the urn at the $(i + M)^{th}$ draw. Everything else remains the same. It was shown in [1, Section VI] that the generated noise process is both stationary and ergodic yielding a positive channel capacity that increases with the memory M. The FMCC and its queue-based generalization, which will be discussed later, were also shown in [30] to model

Rician fading channels more accurately than the Gilbert-Elliott channel.

Note that for a block length $n \leq M$, the FMCC becomes analytically equivalent to the IMCC and the probability of an n-bit error word $\mathbf{z}_1^n$ is given by (2.4). If $n > M$, the probability of the error word is given by:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = L^{(M)} \prod_{i=M+1}^{n} \left[\frac{p + s_{i-1}\delta}{1 + M\delta}\right]^{z_i} \left[\frac{1 - p + (M - s_{i-1})\delta}{1 + M\delta}\right]^{1-z_i} \tag{2.5}$$

where

$$L^{(M)} = \frac{\prod_{i=0}^{s_M-1}(p + i\delta) \prod_{j=0}^{M-s_M-1}(1 - p + j\delta)}{\prod_{l=1}^{M-1}(1 + l\delta)}$$

and for $k \geq M$,

$$s_k = \sum_{i=k-M+1}^{k} z_i.$$

The correlation coefficient of the noise process is given by:

$$\begin{aligned} \epsilon &= \frac{\mathbb{E}[Z_i Z_{i+1}] - \mathbb{E}[Z_i]^2}{\mathrm{Var}(Z_i)} \\ &= \frac{\delta}{\delta + 1} \geq 0. \end{aligned} \tag{2.6}$$

As the IMCC, when $\delta = 0$ (or equivalently $\epsilon = 0$), the FMCC reduces to the BSC(p).

2.2.3 Queue-Based Channel

Introduced in [29], the queue-based channel (QBC) is another communication channel model that belongs to the class of channels with binary additive M^{th}-order Markov noise. It is a generalization of the FMCC while remaining mathematically tractable. The corresponding noise process can be generated by a slightly more complicated scheme than the Polya contagion urn scheme.

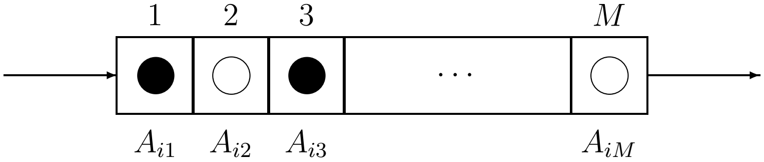


Figure 2.2: A queue of length M.

Consider two parcels:

- **Parcel 1** is a queue of length M (see Fig. 2.2) that contains M balls, either red or black. The random variables A_{ij} (i is a time index, $i > 0$; j is the cell index in the queue, $1 \leq j \leq M$) are defined as follows:

$$A_{ij} \;=\; \begin{cases} 1, & \text{if the } j^{\text{th}} \text{ cell contains a red ball at time } i \\ 0, & \text{otherwise.} \end{cases}$$

- **Parcel 2** is an urn containing a very large number of balls where the proportion of red balls is $p \in (0, 1/2)$, without loss of generality.

For the i^{th} draw, a biased coin with $P(H) = \varphi \in [0, 1)$ is tossed. If the outcome is heads, we select the queue or parcel 1. Otherwise, we select the urn or parcel 2.

- **Case 1:** If the queue is selected and if $M \geq 2$, a pointer selects the j^{th} cell with probability $\frac{1}{M-1+\alpha}$ if $1 \leq j \leq M-1$, and selects the last cell with a probability $\frac{\alpha}{M-1+\alpha}$, where $\alpha \geq 0$. If $M = 1$, then the pointer selects the only cell with probability 1.

- **Case 2:** If the urn is selected, a ball is drawn at random from it.

Based on the color of the ball selected by the pointer or drawn from the urn, a ball of the same color is inserted in the queue shifting its content and forcing out the ball in the last cell. Finally, the noise process $\{Z_i\}_{i=1}^{\infty}$ is generated as follows:

$$Z_i = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ experiment selects a red ball} \\ 0, & \text{otherwise.} \end{cases}$$

Note that for $M = 1$, the QBC is identical to the FMCC with $M = 1$ or the BFMNC.

The probability of an n-bit error pattern $\mathbf{z}_1^n$ can be written as follows:

- If $n \leq M$ then:

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= \frac{\prod_{j=0}^{n-d_1^n-1}\left[(1-\varphi)(1-p) + j\frac{\varphi}{M-1+\alpha}\right]}{\prod_{j=M-n}^{M-1}\left[1-(\alpha+j)\frac{\varphi}{M-1+\alpha}\right]} \\ &\quad \times \prod_{j=0}^{d_1^n-1}\left[(1-\varphi)p + j\frac{\varphi}{M-1+\alpha}\right] \qquad (2.7) \end{aligned}$$

  where $d_a^b = \sum_{i=a}^{b} z_i$ ($d_a^b = 0$ if $a > b$).

- If $n > M$ then:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = L^{(M)} \prod_{i=M+1}^{n} \left[(d_{i-M+1}^{i-1} + \alpha z_{i-M}) \frac{\varphi}{M-1+\alpha} + (1-\varphi)p \right]^{z_i} \left\{ \frac{\left[M-1-d_{i-M+1}^{i-1} + \alpha(1-z_{i-M})\right]\varphi}{M-1+\alpha} + (1-\varphi)(1-p) \right\}^{1-z_i} \tag{2.8}$$

where:

$$L^{(M)} = \frac{\prod_{j=0}^{M-d_1^M-1} \left[(1-\varphi)(1-p) + j\frac{\varphi}{M-1+\alpha}\right]}{\prod_{j=0}^{M-1}\left[1-(\alpha+j)\frac{\varphi}{M-1+\alpha}\right]} \times \prod_{j=0}^{d_1^M-1}\left[(1-\varphi)p + j\frac{\varphi}{M-1+\alpha}\right].$$

It can be shown that the channel bit error rate is p, and the correlation coefficient of the noise process is

$$\epsilon = \frac{\varphi}{(M-1+\alpha) - \varphi(M-2+\alpha)} \geq 0. \tag{2.9}$$

Equations (2.7) and (2.8) can be re-written in terms of p and ϵ as follows:

- If $n \leq M$ then:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = \frac{\prod_{j=0}^{n-d_1^n-1}\left[(1-p) + j\frac{\epsilon}{1-\epsilon}\right]\prod_{j=0}^{d_1^n-1}\left(p + j\frac{\epsilon}{1-\epsilon}\right)}{\prod_{j=0}^{n-1}\left(1+j\frac{\epsilon}{1-\epsilon}\right)}. \tag{2.10}$$

- If $n > M$ then:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \;=\; \frac{L^{(M)}}{\left[1 + (M - 1 + \alpha)\dfrac{\epsilon}{1-\epsilon}\right]^{n-M}} \prod_{i=M+1}^{n} \left[\frac{(d_{i-M+1}^{i-1} + \alpha z_{i-M})\epsilon}{1-\epsilon} + p\right]^{z_i} \left\{\frac{\left[M - 1 - d_{i-M+1}^{i-1} + \alpha(1 - z_{i-M})\right]\epsilon}{1-\epsilon} + 1 - p\right\}^{1-z_i} \tag{2.11}$$

where:

$$L^{(M)} \;=\; \frac{\prod_{j=0}^{M-d_1^M-1}\left[(1-p) + j\dfrac{\epsilon}{1-\epsilon}\right]\prod_{j=0}^{d_1^M-1}\left(p + j\dfrac{\epsilon}{1-\epsilon}\right)}{\prod_{j=0}^{M-1}\left(1 + j\dfrac{\epsilon}{1-\epsilon}\right)}.$$

Lemma 2.1. **[29, Lemma 2, Theorem 3]** *When $\alpha = 0$, the QBC with parameters ϵ, p and M reduces to the FMCC with parameters ϵ, p and $M - 1$.*

Lemma 2.2. *The all-zero error word $\mathbf{0}^n$ is the most likely among all error words of length n generated by the QBC.*

Proof. We consider the cases when $n \le M$ and $n > M$ separately.

- <u>Case 1: $n \le M$</u>

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= \frac{\prod_{j=0}^{n-d_1^n-1}\left[1 - p + j\dfrac{\epsilon}{1-\epsilon}\right]\prod_{j=0}^{d_1^n-1}\left(p + j\dfrac{\epsilon}{1-\epsilon}\right)}{\prod_{j=0}^{n-1}\left(1 + j\dfrac{\epsilon}{1-\epsilon}\right)} \\ &\le \frac{\prod_{j=0}^{n-d_1^n-1}\left[1 - p + j\dfrac{\epsilon}{1-\epsilon}\right]\prod_{j=0}^{d_1^n-1}\left(1 - p + j\dfrac{\epsilon}{1-\epsilon}\right)}{\prod_{j=0}^{n-1}\left(1 + j\dfrac{\epsilon}{1-\epsilon}\right)} \end{aligned}$$

$$\begin{aligned} & & (2.12) \\ &\leq& \frac{\prod_{j=0}^{n-1}\left[1-p+j\dfrac{\epsilon}{1-\epsilon}\right]}{\prod_{j=0}^{n-1}\left(1+j\dfrac{\epsilon}{1-\epsilon}\right)} \\ &=& P(\mathbf{Z}_1^n = \mathbf{0}^n), \end{aligned}$$

where the inequality in (2.12) follows from the assumption that $p < \frac{1}{2}$. Note that we get equality iff $\mathbf{z}_1^n = \mathbf{0}^n$.

- Case 2: $n > M$ First, note that $L^{(M)}$ in (2.11) is equal to the probability of the length M error word $\mathbf{z}_1^M$ generated by the QBC. We proved in Case 1 that $\mathbf{z}_1^M = \mathbf{0}^M$ (i.e., setting the first M bits to 0 in $\mathbf{z}_1^n$) maximizes $L^{(M)}$. Let

$$\begin{aligned} Q(\mathbf{z}_1^n) &=& \frac{1}{\left[1+(M-1+\alpha)\dfrac{\epsilon}{1-\epsilon}\right]^{n-M}} \prod_{i=M+1}^{n}\left[\frac{(d_{i-M+1}^{i-1}+\alpha z_{i-M})\epsilon}{1-\epsilon}+p\right]^{z_i} \\ && \left\{\frac{\left[M-1-d_{i-M+1}^{i-1}+\alpha(1-z_{i-M})\right]\epsilon}{1-\epsilon}+1-p\right\}^{1-z_i}. \end{aligned}$$

Then,

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \quad = \quad L^{(M)} Q(\mathbf{z}_1^n).$$

Hence,

$$Q(\mathbf{z}_1^n) \quad \leq \quad \frac{1}{\left[1+(M-1+\alpha)\dfrac{\epsilon}{1-\epsilon}\right]^{n-M}} \prod_{i=M+1}^{n}\left[\frac{(M-1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^{z_i}$$

$$
\begin{aligned}
&\left\{\frac{[M-1+\alpha]\,\epsilon}{1-\epsilon}+1-p\right\}^{1-z_i} \qquad (2.13)\\
= &\ \frac{\left[\dfrac{(M-1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^{n-M}}{\left[1+(M-1+\alpha)\dfrac{\epsilon}{1-\epsilon}\right]^{n-M}}\\
= &\ Q(\mathbf{0}^n),
\end{aligned}
$$

where (2.13) is a result of the following facts:

- $p < \frac{1}{2}$
- $d_i^j \leq (j-i+1)$ if $j >= i$
- $0 \leq z_i \leq 1$.

Equality is achieved iff $\mathbf{z}_1^n = \mathbf{0}^n$. Therefore, the all-zero error word $\mathbf{0}^n$ maximizes both $L^{(M)}$ and $Q(\mathbf{z}_1^n)$ and hence is the most likely error pattern generated by the QBC.

□

Remark 2.1. (Summary of the QBC features)

- *If $M = 1$, the QBC reduces to the FMCC with $M = 1$, or equivalently, the BFMNC with identical BER p and noise correlation coefficient ϵ.*
- *If $\varphi = 0$ (i.e., $\epsilon = 0$), the QBC reduces to a BSC with crossover probability p.*
- *If $\alpha = 1$, the QBC reduces to the FMCC with the same memory parameter M, the same BER p, and with the same correlation coefficient ϵ, or with $\delta = \frac{\epsilon}{1-\epsilon}$.*

- *If $\alpha = 0$, the QBC with parameters (memory M, BER p and correlation coefficient ϵ) reduces to the FMCC with parameters $(M-1, p, \epsilon)$.*

Remark 2.2. *The QBC is shown in [30] to model Rician fading channels more accurately than the Gilbert-Elliott channel.*

2.2.4 Gilbert-Elliott Channel

The Gilbert-Elliott channel (GEC) is one of the most widely used binary channel models in the literature (belonging to the class of finite-state Markov channels [10]) for describing burst error patterns in real communication channels. This channel is illustrated in Fig. 2.3 and is governed by an underlying two-state Markov chain where one state, denoted by G, represents the "good" state of the channel and the other state, denoted by B, represents its bad state. We set binary values to the two states: $s = 0$ for state G and $s = 1$ for state B. The state process is an irreducible stationary Markov source with transition probabilities $P(G|B) = g \in (0, 1)$ and

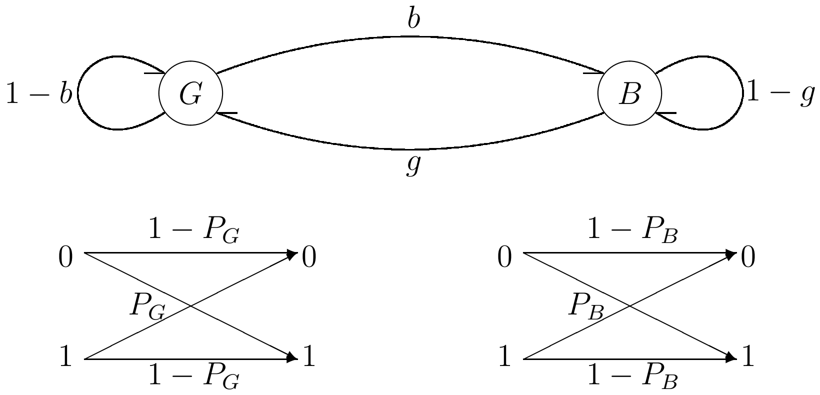


Figure 2.3: The Gilbert-Elliott channel model.

$P(B|G) = b \in (0, 1)$. The corresponding state transition matrix is

$$\mathbf{P} = \begin{bmatrix} 1-b & b \\ g & 1-g \end{bmatrix}.$$

The state process admits a stationary distribution (written as a column vector)

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_G \\ \pi_B \end{bmatrix} = \begin{bmatrix} \dfrac{g}{b+g} \\ \dfrac{b}{b+g} \end{bmatrix} \tag{2.14}$$

The channel is a BSC with crossover probability P_G in the "good" state and P_B in the "bad" state. Conventionally, we choose the states such that: $P_G < P_B$. We define the following two matrices:

$$\mathbf{P}(0) = \begin{bmatrix} (1-b)(1-P_G) & b(1-P_B) \\ g(1-P_G) & (1-g)(1-P_B) \end{bmatrix} \tag{2.15}$$

and

$$\mathbf{P}(1) = \begin{bmatrix} (1-b)P_G & bP_B \\ gP_G & (1-g)P_B \end{bmatrix} \tag{2.16}$$

For $t \in \{0, 1\}$, the ij^{th} entry of $\mathbf{P}(t)$ equals $P(Z_k = t, S_k = j | S_{k-1} = i)$, where $\{Z_k\}_{k=1}^{\infty}$ and $\{S_k\}_{k=1}^{\infty}$ are the noise and state processes associated with the GEC, respectively. The GEC can be described by $Y_k = X_k \oplus Z_k$, where Y_k, X_k and Z_k are the output, input and noise symbols, respectively, at time k. The noise process $\{Z_k\}_{k=1}^{\infty}$, which is independent of the input process, is a stationary ergodic hidden Markov source (of infinite memory). Note that the GEC's BER, which we denote by p, is given by:

$$p = \frac{g}{b+g}P_G + \frac{b}{b+g}P_B.$$

The probability of an $n-$bit error pattern $\mathbf{z}_1^n$ can be given by (in matrix form):

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = \boldsymbol{\pi}^T \left(\prod_{i=1}^{n} \mathbf{P}(z_i) \right) \underline{1}^n, \tag{2.17}$$

where T denotes transposition and $\underline{1}^n$ is the all-ones column vector of length n. When the channel state $\mathbf{s}_1^n$ is known, the probability of $\mathbf{z}^n$ given the channel states is:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n) = \prod_{i=1}^{n} \left[(1-P_B)^{1-z_i} p_B^{z_i} \right]_i^s \left[(1-P_G)^{1-z_i} P_G^{z_i} \right]^{1-s_i}. \tag{2.18}$$

Chapter 3

Channel Coding

Shannon's landmark paper on the "Mathematical theory of communication" [19] marks the emergence of the field of coding theory. In the early 1940s, sending information at a positive rate was thought to be impossible to achieve with a negligible probability of error. Shannon's channel coding theorem disproved the above claim by showing the existence of some codes, with sufficiently large block lengths and rates below a theoretical limit characteristic of the communication channel known as the channel capacity, that allow the probability of error at the receiver to be made arbitrarily small. However, the original proof was based on random coding and hence was not constructive. The next logical step was to try to construct and implement such codes which is the main task of the field of channel coding theory.

Informally, channel coding is the process of "smartly" adding redundancy to a message before sending it, in order to be able to protect that message against channel errors encountered during transmission. The role of the encoder is better illustrated with an example:

Consider an imaginary binary communication system where the transmitter is repeatedly tossing a coin and wants to transmit the outcomes to the receiver (e.g., 0 for *heads*, 1 for *tails*). The binary channel is noisy and it flips every transmitted bit with a probability $p = 0.1$ (the channel is a BSC with crossover probability p). If the message is transmitted as is, i.e., by sending only one bit to determining the outcome, then it will be received incorrectly with probability $p = 0.1$. Now, consider the simplest scheme where the transmitter sends the same bit three times. If the decoder uses a majority decoding rule, the probability of error at the receiver is:

$$\begin{aligned} P_{\text{err}} &= P(2 \text{ errors}) + P(3 \text{ errors}) \\ &= 3p^2(1-p) + p^3 \\ &= 0.0280 < 0.1. \end{aligned}$$

Hence, adding only 2 redundant bits reduces the probability of error by a factor of 3. We can easily prove that by sending the same bit 5 times the probability of error is reduced further to 0.0086. The problem with this approach is that information is transmitted at a rate of $1/n$, where n is total number of transmitted bits. Even though the probability of error goes to 0 as n goes to infinity, the information rate goes to 0 as well. Hence, this scheme is rather inefficient since we know from the channel coding theorem that there exists a scheme that transmits information at a rate arbitrarily close to the BSC's capacity $C = 1 - H_b(p) = 0.469$ with a negligible probability of error, where $H_b(\cdot)$ denotes the binary entropy function [6]. The field of coding theory considers the problem of finding methods of adding redundancy in a "smarter" way, in order to maximize the benefit in the error correction capability. In what follows, we provide a brief overview of the basic concepts in coding theory [4, 23].

3.1 Binary Block Codes over Binary Additive Noise Channels

By definition, "*block codes* map a block of information bits onto a channel codeword and there is no dependence on past information bits" [6].

Definition 3.1. *An (n, M) binary block code $\mathcal{C}$ is an injective mapping:*

$$\mathcal{E}: \qquad \mathbf{F}_2^k \mapsto \mathbf{F}_2^n,$$

where $k = \log_2(M)$, assuming M is a power of 2 and $n \geq k$. The parameters of the code are explained below:

- *n: is the length (or block length) of a codeword in $\mathcal{C}$, where a codeword is an element of $\mathcal{C} = \mathcal{E}(\mathbf{F}_2^k)$.*
- *M: is the code size.*
- *k: is the dimension of $\mathcal{C}$, or the message length.*
- *$\mathbf{F}_2 = GF(2)$: is the input and output alphabet.*

The rate of this code is:

$$R = \frac{log_2(M)}{n} = \frac{k}{n} \text{ message bits/code bits.}$$

In communication systems, a codeword $\mathbf{c} \in \mathcal{C}$ is sent over the communication channel and $\mathbf{y}_1^n = (y_1, ..., y_n) \in \mathbf{F}_2^n$ is received. Over an additive noise channel $\mathbf{y}_1^n = \mathbf{c} + \mathbf{e}_1^n$ where $\mathbf{e}_1^n = (e_1, ..., e_n) \in \mathbf{F}_2^n$ is the error pattern generated by the channel.

Definition 3.2. *The minimum Hamming distance of an (n, M) code $\mathcal{C}$ is:*

$$d_{min} \quad := \quad \min\{d_H(\mathbf{c}, \mathbf{c}') : (\mathbf{c}, \mathbf{c}') \in \mathcal{C}^2 \text{ and } \mathbf{c} \neq \mathbf{c}'\},$$

where $d_H(\mathbf{x}_1^n, \mathbf{y}_1^n)$ is the Hamming distance between the vectors $\mathbf{x}_1^n$ and $\mathbf{y}_1^n$, or the number of indices where these vectors (or n-tuples, or words of length n) differ. An (n, M) code $\mathcal{C}$ with a minimum Hamming distance d_{min} is described as an (n, M, d_{min}) code.

Definition 3.3. (Error set) *For a code $\mathcal{C}$, the error set of a received word $\mathbf{y}_1^n$ is defined as:*

$$\mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n) = \{\mathbf{e}_1^n \in \mathbf{F}_2^n : \mathbf{y}_1^n + \mathbf{e}_1^n \in \mathcal{C}\}.$$

This is the set of all possible error words that could have possibly occurred during transmission in order to receive $\mathbf{y}_1^n$.

Definition 3.4. *A decoder $\mathcal{D}_{dec}$ of a code $\mathcal{C}$ is a deterministic function:*

$$\mathcal{D}_{dec} : \mathbf{F}_2^n \mapsto \mathcal{C} \cup \{\emptyset\}. \tag{3.1}$$

The decoder maps every possible n-tuple in $\mathbf{F}_2^n$ to a codeword in $\mathcal{C}$ or declares a decoding failure. If $\exists$ at least one word $\mathbf{y} \in \mathbf{F}_2^n$ such that: $\mathcal{D}_{dec}(\mathbf{y}) = \emptyset$, then the *decoder is said to be incomplete. A complete decoder maps every word in $\mathbf{F}_2^n$ to a codeword in $\mathcal{C}$.*

For additive noise channels, the output of the decoder can be its estimate of the error word generated by the channel rather than its estimate of the codeword sent. Then subtracting this error pattern from the received word yields the decoded codeword. We denote such a decoder by

$$\mathcal{D} : \mathbf{y}_1^n \quad \mapsto \quad \mathbf{e}_1^n \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n).$$

For example, under minimum Hamming distance decoding, $\mathcal{D}$ outputs the error pattern of smallest Hamming weight in the corresponding error set. Such minimum distance (MD) decoder will be examined rigorously. For additive noise channels, the two decoders $\mathcal{D}_{dec}$ and $\mathcal{D}$ are related by:

$$\mathcal{D}_{dec} : \mathbf{y}_1^n \quad \mapsto \quad \mathbf{y}_1^n - \mathcal{D}(\mathbf{y}_1^n).$$

Thus, we will hereafter use the decoding function $\mathcal{D}$ instead of $\mathcal{D}_{dec}$.

Definition 3.5. *The packing radius r_{pac} of a code $\mathcal{C}$ is the maximum radius such that the Hamming spheres about each codeword do not intersect.*

It can be shown the packing radius is the guaranteed number of errors that can be corrected by the code under minimum Hamming distance decoding. For an (n, M, d_{min}) code $\mathcal{C}$, the packing radius is

$$r_{pac} \quad = \quad \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor .$$

Definition 3.6. *The covering radius r_{cov} of an (n, M, d_{min}) code $\mathcal{C}$ is the minimum radius of the Hamming spheres centered at each codeword such that every word in $\mathbf{F}_2^n$ is contained in at least one of the spheres. Formally,*

$$r_{cov} \;=\; \max_{\mathbf{y}_1^n \in \mathbf{F}_2^n} \min_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}_1^n, \mathbf{c}).$$

Definition 3.7. (Perfect and quasi-perfect codes) *A code $\mathcal{C}$ is called* perfect *iff its covering and packing radii are equal. In other words, $\mathcal{C}$ is a perfect code iff every $\mathbf{y}_1^n \in \mathbf{F}_2^n$ is within a Hamming distance of at most r_{pac} from exactly one codeword $\mathbf{c} \in \mathcal{C}$.*

Note that a *quasi-perfect* code satisfies $r_{cov} = r_{pac} + 1$. For a perfect code $\mathcal{C}$, $\forall \mathbf{y}_1^n \in \mathbf{F}_2^n, \exists$ a unique $\mathbf{e}_1^n \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)$ s.t $w_H(\mathbf{e}_1^n) \leq r_{pac(\mathcal{C})}$, where $w_H(\mathbf{e}_1^n)$ denotes the Hamming weight of the vector $\mathbf{e}_1^n$ or the number of non-zero entries in the vector.

Definition 3.8. (Decoding set) *For a code $\mathcal{C}$, the decoding set of a decoder $\mathcal{D}$ that outputs for every received word its error estimate, is defined as:*

$$\mathcal{D}_{\mathcal{C}} = \{\mathcal{D}(\mathbf{y}_1^n) : \mathbf{y}_1^n \in \mathbf{F}_2^n\},$$

where $\mathcal{D}(\mathbf{y}_1^n) \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)$. This is the set of all error patterns that the decoder outputs.

3.2 Binary Linear Codes over Binary Additive Noise Channels

Definition 3.9. *An $[n, k]$ binary linear code $\mathcal{C}$ is a k-dimensional subspace of $\mathbf{F}_2^n$, where $\mathbf{F}_2 = GF(2)$. Its rate is $R = \frac{k}{n}$. An encoder for $\mathcal{C}$ is a bijective map:*

$$\mathcal{E} : \mathbf{F}_2^k \mapsto \mathcal{C}.$$

An $[n, k, d_{min}]$ binary linear code is an $[n, k]$ binary linear code with minimum Hamming distance d_{min}.

Lemma 3.1. *For a binary linear code, the minimum Hamming distance is equal to the minimum Hamming weight of its non-zero codewords.*

Definition 3.10. *A generator matrix $\mathbf{G}$ of an $[n, k, d_{min}]$ code $\mathcal{C}$ is a $k \times n$ $\mathbf{F}_2$-valued matrix whose rows generate $\mathcal{C}$, i.e.*

$$\mathcal{C} \quad = \quad \{\mathbf{aG} : \mathbf{a} \in \mathbf{F}_2^k\}.$$

If the generator matrix is written as follows: $\mathbf{G} = (\mathbf{I}_k|\mathbf{P})$, where $\mathbf{I}_k$ is the k-dimensional identity matrix and $\mathbf{P}$ is a $k \times (n-k)$ matrix, then we say that $\mathbf{G}$ is in its standard form.

Remark 3.1. *Every linear code is equivalent to a code with a standard form $\mathbf{G}$, where two codes $\mathcal{C}_1$ and $\mathcal{C}_2$ are called equivalent if both codes have the same set of codewords but a different linear mapping, or if the they can be obtained from the other by permuting the bits in their codewords.*

Definition 3.11. *A parity-check matrix* $\mathbf{H}$ *of an* $[n, k, d_{min}]$ *code* $\mathcal{C}$ *is an* $(n-k)\times n$ *matrix satisfying* $\mathbf{G}\mathbf{H}^T = \mathbf{0}$*, where* $\mathbf{0}$ *is the all-zero* $k \times (n-k)$ *matrix and* $\mathbf{G}$ *is the generator matrix of* $\mathcal{C}$*. It is immediate to prove that:*

$$\mathbf{c} \in \mathcal{C} \quad \Longleftrightarrow \quad \mathbf{c}\mathbf{H}^T = \mathbf{0}^{n-k},$$

where $\mathbf{0}^{n-k}$ *is all-zero word of length* $n-k$*.*

An important observation is that the minimum Hamming distance d_{min} of $\mathcal{C}$ is:

$$d_{min} \;=\; \max\{d \in \mathbb{N} : \text{ any } d-1 \text{ columns of } \mathbf{H} \text{ are linearly independent.}\}.$$

Definition 3.12. *The syndrome of a received word* $\mathbf{y} \in \mathbf{F}_2^n$ *is the vector* $\mathbf{s} = \mathbf{y}\mathbf{H}^T$*. The set of all words in* $\mathbf{F}_2^n$ *that have the same syndrome is called a coset of* $\mathcal{C}$*.*

For a linear code $\mathcal{C}$, the coset associated with a received word $\mathbf{y}_1^n$ is the same as the error set $\mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)$. In fact, assume $\mathbf{c} \in \mathcal{C}$ is transmitted. The received word is $\mathbf{y}_1^n = \mathbf{c} + \mathbf{e}_1^n$ where $\mathbf{e}_1^n$ is a possible error pattern that might have occured, i.e. $\mathbf{e}_1^n \in \mathcal{S}_C(\mathbf{y}_1^n)$. The syndrome of $\mathbf{y}_1^n$ is:

$$\begin{aligned}\mathbf{s}(\mathbf{y}_1^n) &= \mathbf{y}_1^n\mathbf{H}^T \\ &= (\mathbf{c} + \mathbf{e}_1^n)\mathbf{H}^T \\ &= \mathbf{e}_1^n\mathbf{H}^T \\ &= \mathbf{s}(\mathbf{e}_1^n).\end{aligned}$$

Hence, $\mathbf{e}_1^n$ is in the same coset of $\mathbf{y}_1^n$.

Definition 3.13. *A coset leader is an element of the coset with minimum Hamming weight. It corresponds to the error pattern estimate of a minimum Hamming distance decoder (defined later in more details).*

3.3 Probabilistic Decoding over Binary Additive Noise Channels

3.3.1 Maximum A-Posteriori (MAP) Decoder

The MAP decoding is the optimal decoding rule, in terms of minimizing the probability of codeword error. Given a received word $\mathbf{y} \in \mathbf{F}_2^n$, the decoder $\mathcal{D}$ outputs the error pattern $\hat{\mathbf{e}} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})$ given by:

$$\begin{aligned}
\hat{\mathbf{e}} &= \arg\max_{\mathbf{e}\in\mathbf{F}_2^n} P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y}+\mathbf{e} \text{ was transmitted}|\mathbf{y} \text{ is received}\right) \\
&= \arg\max_{\mathbf{e}\in\mathbf{F}_2^n} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}+\mathbf{e} \text{ was transmitted}) P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y} \text{ is received}|\mathbf{y}+\mathbf{e} \text{ was transmitted}\right) \\
&= \arg\max_{\mathbf{e}\in\mathbf{F}_2^n} P(\mathbf{X}=\mathbf{y}+\mathbf{e}) P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{y}+\mathbf{e}\right) \\
&= \arg\max_{\mathbf{e}\in\mathcal{S}_{\mathcal{C}}(\mathbf{y})} P(\mathbf{X}=\mathbf{y}+\mathbf{e}) P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{y}+\mathbf{e}\right). \qquad (3.2)
\end{aligned}$$

3.3.2 Maximum Likelihood (ML) Decoder

Given a received word $\mathbf{y} \in \mathbf{F}_2^n$, the ML decoder outputs the error pattern $\hat{\mathbf{e}}$ such that:

$$\hat{\mathbf{e}} = \arg\max_{\mathbf{e}\in\mathcal{S}_{\mathcal{C}}(\mathbf{y})} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y}=\mathbf{y}|\mathbf{X}=\mathbf{y}+\mathbf{e}).$$

If the codewords are equiprobable, the MAP decoding rule reduces to the ML decoding rule since $P(X = \mathbf{y} + \mathbf{e})$ becomes a constant and can be dropped from (3.2). Typically in communication systems, we can assume that all codewords are equiprobable.

3.3.3 Minimum Hamming Distance (MD) Decoder

Given a received word $\mathbf{y} \in \mathbf{F}_2^n$, the MD decoder outputs the error pattern:

$$\hat{\mathbf{e}} = \arg \min_{\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} w_H(\mathbf{e}).$$

Remark 3.2. *For a BSC(p) with crossover probability $p < 1/2$, given a received word $\mathbf{y} \in \mathbf{F}_2^n$, the output of the ML decoder is:*

$$\begin{aligned}
\hat{\mathbf{e}} &= \arg \max_{\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{y} + \mathbf{e}) \\
&= \arg \max_{\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} P(\mathbf{Z} = \mathbf{e}) \\
&= \arg \max_{\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} (1-p)^n \left(\frac{p}{1-p} \right)^{w_H(\mathbf{e})} \\
&= \arg \max_{\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} \left(\frac{p}{1-p} \right)^{w_H(\mathbf{e})} \\
&= \arg \min_{\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} w_H(\mathbf{e}).
\end{aligned}$$

Hence, we clearly see that the ML decoding rule reduces to MD decoding when the channel is a BSC with crossover probability $p < 1/2$. This is the main reason why the Hamming distance is the metric used in deriving the basic results in coding theory.

Unlike the above three decoders, the strict minimum Hamming distance (SMD) decoder is an incomplete decoder. It has exactly the same decoding rule as the MD decoder except that it declares a decoding failure if more than one error word have the same minimum Hamming weight.

Lemma 3.2. [**10**] *A binary perfect or quasi-perfect code are optimal on the BSC with a crossover probability* $p < 1/2$*, in the sense that ML decoding (or MD decoding since we proved that the two decoding methods are equivalent in this case) achieves the minimum average probability of codeword error among all codes of the same length and dimension.*

3.4 Hamming Codes

Definition 3.14. [**23, p. 38**] *The* $[n, n-k]$ *binary Hamming code where* $n = (2^k - 1)$ *and* $k \geq 2$ *is a code whose parity-check matrix has columns that are pairwise linearly independent, i.e. the columns are a maximal set of pairwise linearly independent vectors.*

Hamming codes are perfect codes with minimum distance $d_{min} = 3$ and $r_{cov} = r_{pac} = 1$.

The $[7, 4, 3]$ Hamming code is an example of the family of Hamming codes. Its generator and parity-check matrices $\mathbf{G}$ and $\mathbf{H}$ are given by:

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix},$$

$$\mathbf{H} \; = \; \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

A $[2^k - 1, 2^k - 1 - k]$ Hamming code $\mathcal{C}$ can be extended to a $[2^k, 2^k - 1 - k]$ code $\mathcal{C}'$ by adding an extra all-one column to the generator matrix $\mathbf{G}$. $\mathcal{C}'$ is a quasi-perfect code.

3.5 The Binary Golay Code

Another code that we use in our simulations later on is the $[23, 12, 7]$ binary Golay code $\mathcal{G}_{23}$ and its extended $[24, 12, 8]$ code $\mathcal{G}_{24}$. The Golay code $\mathcal{G}_{23}$ is one of the most famous binary codes; it is the only non-trivial binary perfect code other than the binary Hamming codes [22, 24]. The extended Golay $\mathcal{G}_{24}$ is a quasi-perfect code; the weight of every codeword in $\mathcal{G}_{24}$ is a multiple of 4 and its minimum distance is $d_{min} = 8$. The extended Golay code is also self-dual, i.e., its generator matrix $\mathbf{G}_{24}$ is also its parity-check matrix matrix ($\mathbf{G}\mathbf{G}^T = \mathbf{0}$, where $\mathbf{0}$ is the 12×12 all-zero matrix).

3.6 Reed-Muller Codes

Reed-Muller codes are some of the oldest linear binary error correcting codes. They were discovered by Muller and provided with a decoding algorithm by Reed. There are several ways to describe these codes. For the sake of brevity, we only give a recursive description. Interested readers are encouraged to read relevant sections in [23] for alternative descriptions of these codes.

Given integers $m \geq 1$ and $0 \leq r \leq m$, $\mathcal{RM}(r,m)$ denotes the r^{th}-order Reed-Muller code of length $n = 2^m$. The dimension of this code is

$$k = \sum_{i=0}^{r} \binom{m}{i}$$

and its minimum distance is $d_{min} = 2^{m-r}$. $\mathcal{RM}(r,m)$ can be constructed recursively as follows:

- $\mathcal{RM}(0,m)$ is the repetition code of length $n = 2^m$. It consists of two codewords: the all-zero codeword $\mathbf{0}^n = (00...0)$ and the all-one codeword $\mathbf{1}^n = (11...1)$.

- $\mathcal{RM}(m,m)$ is the entire space of 2^m-tuples, $\mathbf{F}_2^n$ where $n = 2^m$.

- For $0 < r < m$,

$$\begin{aligned} \mathcal{RM}(r,m) &= \{(\mathbf{c}_1, \mathbf{c}_1 \oplus \mathbf{c}_2) : \\ &\qquad \mathbf{c}_1 \in \mathcal{RM}(r,m-1), \mathbf{c}_2 \in \mathcal{RM}(r-1,m-1)\}, \quad (3.3) \end{aligned}$$

  where $\oplus$ is the modulo-2 addition and where $(\mathbf{u}, \mathbf{v})$ denotes the concatenation of the vectors $\mathbf{u}$ and $\mathbf{v}$.

Example: $\mathcal{RM}(1,2)$

Note that $\mathcal{RM}(0,1) = \{00, 11\}$ and $\mathcal{RM}(1,1) = \{00, 01, 10, 11\}$. According to (3.3),

$$\mathcal{RM}(1,2) \;\; = \;\; \{(\mathbf{c}_1, \mathbf{c}_1 \oplus \mathbf{c}_2) :$$

$$
\begin{aligned}
& \mathbf{c}_1 \in \mathcal{RM}(1,1), \mathbf{c}_2 \in \mathcal{RM}(0,1)\} \\
= \;& \{(00, 00 \oplus 00), (00, 00 \oplus 11), \\
& (01, 01 \oplus 00), (01, 01 \oplus 11), \\
& (10, 10 \oplus 00), (10, 10 \oplus 11), \\
& (11, 11 \oplus 00), (11, 11 \oplus 11)\} \\
= \;& \{0000, 0011, 0101, 0110, 1010, 1001, 1111, 1100\}.
\end{aligned}
$$

3.7 Construction of Nonlinear Perfect Codes

It is proven in [22, 24] that any nontrivial perfect binary code has the parameters of a Hamming code or a Golay code. In this section, we briefly describe a method for constructing nonlinear binary perfect codes from linear Hamming codes.

Theorem 3.1. (Vasil'ev code) [25, 26, Theorem 4.2]*Let $\mathcal{V}$ be an $[n, n-k]$ binary Hamming code, where $n = (2^k - 1)$. Let $f : \mathcal{V} \mapsto \mathbf{F}_2 = \{0, 1\}$ be an arbitrary nonlinear mapping such that:*

- $f(\mathbf{0}^n) = 0$*, where* $\mathbf{0}^n$ *is the all-zero codeword*
- $f(\mathbf{u}) + f(\mathbf{v}) \neq f(\mathbf{u} + \mathbf{v})$ *for some* $\mathbf{u}, \mathbf{v} \in \mathcal{V}$*.*

We define the code $\mathcal{C}$

$$\mathcal{C} = \{(\mathbf{x}, \mathbf{x} \oplus \mathbf{v}, p_H(\mathbf{x}) \oplus f(\mathbf{v})) : \mathbf{x} \in \mathbf{F}_2^n, \mathbf{v} \in \mathcal{V}\} \tag{3.4}$$

where $\oplus$ is the modulo-2 addition, $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ *denotes the concatenation of the vectors*

$\mathbf{a}, \mathbf{b}$ *and* $\mathbf{c}$ *and*

$$p_H(\mathbf{x}) \quad = \quad w_H(\mathbf{x}) \mod 2,$$

where $w_H(\cdot)$ *denotes the Hamming weight.* $\mathcal{C}$ *is a* $(2n+1, 2^{2n-k}, 3)$ *perfect nonlinear code. Clearly,* $\mathcal{C}$ *has the same parameters as a* $[n', n'-k']$ *Hamming code* $\mathcal{V}'$*, where* $k' = k+1$ *and* $n' = (2^{k+1}-1) = 2n+1$*.*

Theorem 3.2. *Let* $\mathcal{C}$ *be a* $(2n+1, 2^{2n-k}, 3)$ *nonlinear perfect code constructed using the method in Theorem 3.1. Then,*

(a) $\mathcal{C}$ *has the all-one codeword* $\mathbf{1}^{2n+1}$*.*

(b) For any codeword $\mathbf{c} \in \mathcal{C}$*,* $\mathbf{1}^{2n+1} \oplus \mathbf{c}$ *is also a codeword.*

Proof. (a) Since $\mathbf{0}^n \in \mathcal{V}$, consider $\mathbf{x} = \mathbf{1}^n \in \mathbf{F}_2^n$. From (3.4), $\mathcal{C}$ has a codeword $\mathbf{c}$ given by

$$\begin{aligned} \mathbf{c} &= (\mathbf{1}^n, \mathbf{1}^n \oplus \mathbf{0}^n, p_H(\mathbf{1}^n) \oplus f(\mathbf{0}^n)) \\ &= (\mathbf{1}^n, \mathbf{1}^n, p_H(\mathbf{1}^n)) && (3.5) \\ &= \big(\mathbf{1}^{2n}, 1\big) && (3.6) \\ &= \mathbf{1}^{2n+1}, \end{aligned}$$

where the equality in (3.5) comes from the condition $f(\mathbf{0}^n) = 0$ and we have equality in (3.6) because $n = 2^k - 1$ is an odd number and hence $p_H(\mathbf{1}^n) = 1$. Hence, $\mathbf{c} = \mathbf{1}^{2n+1} \in \mathcal{C}$.

(b) Let $\mathbf{c}$ be a codeword in $\mathcal{C}$. Hence, from Theorem 3.1, there exists $\mathbf{x} \in \mathbf{F}_2^n$

and $\mathbf{v} \in \mathbf{V}$ such that:

$$\mathbf{c} = \left(\mathbf{x}, \mathbf{x} \oplus \mathbf{v}, p_H(\mathbf{x}) \oplus f(\mathbf{v})\right).$$

Let $\mathbf{x}' = \mathbf{1}^n \oplus \mathbf{x}$. Then, $\mathcal{C}$ has a codeword $\mathbf{c}'$ given by

$$\begin{aligned}
\mathbf{c}' &= (\mathbf{x}', \mathbf{x}' \oplus \mathbf{v}, p_H(\mathbf{x}') \oplus f(\mathbf{v})) \\
&= \left(\mathbf{1}^n \oplus \mathbf{x}, \mathbf{1}^n \oplus (\mathbf{x} \oplus \mathbf{v}), w_H(\mathbf{1}^n \oplus \mathbf{x}) + f(\mathbf{v}) \mod 2\right) \\
&= \left(\mathbf{1}^n \oplus \mathbf{x}, \mathbf{1}^n \oplus (\mathbf{x} \oplus \mathbf{v}), 1 \oplus p_H(\mathbf{x}) \oplus f(\mathbf{v})\right) \\
&= \mathbf{1}^{2n+1} \oplus (\mathbf{x}, \mathbf{x} \oplus \mathbf{v}, p_H(\mathbf{x}) \oplus f(\mathbf{v})) \\
&= \mathbf{1}^{2n+1} \oplus \mathbf{c}.
\end{aligned}$$

□

3.8 Generalized Concepts of Coding Theory

As we mentioned before, the basic concepts of coding theory are derived for the memoryless BSC(p) with $p < 1/2$. For this channel, the likelihood metric is equivalent to the Hamming distance. This equivalence does not hold when we change the channel model. In this section, we extend the basic definitions in coding theory to a wider class of channels with arbitrary additive noise processes with memory. The content of this section is discussed in more details in [13].

Consider a general binary additive noise communication channel with a probability transition matrix $P_{\mathbf{Y}|\mathbf{X}}(.|.)$. Let D_n be the following generalized distance:

$$\begin{aligned} D_n : \mathbf{F}_2^n \times \mathbf{F}_2^n &\mapsto \mathbb{R} \\ D_n(\mathbf{x}, \mathbf{y}) &= -\log_k \frac{P_{\mathbf{Y}|\mathbf{X}}(\mathbf{Y} = \mathbf{y}|\mathbf{X} = \mathbf{x})}{P(\mathbf{Z} = \mathbf{0}^n)} \\ &= -\log_k \frac{P(\mathbf{Z} = \mathbf{y} + \mathbf{x})}{P(\mathbf{Z} = \mathbf{0}^n)} \end{aligned}$$

where $k > 1$ is a constant. We denote by $\mathcal{K}(D_n)$ the domain of this distance, i.e.

$$\mathcal{K}(D_n) = \{t \in \mathbb{R} : \exists \mathbf{x}, \mathbf{y} \in \mathbf{F}_2^n, D_n(\mathbf{x}, \mathbf{y}) = t\}.$$

It is natural to associate with the distance D_n a weight function W_n defined as follows:

$$\begin{aligned} W_n : \mathbf{F}_2^n &\mapsto \mathbb{R} \\ W_n(\mathbf{e}) &= -\log_k \frac{P(\mathbf{Z} = \mathbf{e})}{P(\mathbf{Z} = \mathbf{0}^n)}. \end{aligned}$$

Definition 3.15. *The generalized minimum distance of an (n, M) code $\mathcal{C}$ is:*

$$\rho_{min} := \min\{D_n(\mathbf{c}, \mathbf{c}') : \mathbf{c}, \mathbf{c}' \in \mathcal{C} \textit{ and } \mathbf{c} \neq \mathbf{c}'\}.$$

Definition 3.16. *The generalized packing radius ρ_{pac} of an (n, M) code $\mathbf{C}$ with generalized minimum distance ρ_{min} is:*

$$\begin{aligned} \rho_{pac} := \ & \max\{t \in \mathcal{K}(D_n) : \forall \mathbf{y} \in \mathbf{F}_2^n, \exists \textit{ at most one error word } \mathbf{z} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}) \\ & \textit{such that } W_n(\mathbf{z}) \leq t\}. \end{aligned}$$

In other words, ρ_{pac} is the maximum generalized weight of an error word that the code $\mathcal{C}$ can correct under minimum generalized distance decoding, which is equivalent to ML decoding [13].

Definition 3.17. *The generalized covering radius ρ_{cov} of an (n, M) code $\mathcal{C}$ is:*

$$\rho_{cov} \quad = \quad \max_{\mathbf{y} \in \mathbf{F}_2^n} \min_{\mathbf{c} \in \mathcal{C}} D_n(\mathbf{c}, \mathbf{y}).$$

In other words, any received word $\mathbf{y}$ is within at most ρ_{cov} from at least one codeword in $\mathcal{C}$.

Definition 3.18 (Generalized perfect and quasi-perfect codes). *A code $\mathcal{C}$ is called a generalized perfect code iff $\rho_{cov} = \rho_{pac}$. For such a code, every received word $\mathbf{y}$ is within at most ρ_{cov} from **exactly one** codeword in $\mathcal{C}$.*
A code $\mathcal{C}$ is called a generalized quasi-perfect code iff $\rho_{pac} < \rho_{cov}$ and $\nexists t^* \in \mathcal{K}(D_n)$ such that $\rho_{pac} < t^* < \rho_{cov}$.

Remark 3.3. *The generalized definition of perfect (quasi-perfect) codes reduces to the conventional definition when the distance D_n is given by the Hamming distance.*

Theorem 3.3. **[13, Theorem 1]** *Generalized perfect and quasi-perfect codes are optimal (i.e., have minimal codeword error probability) under ML decoding among all codes with the same lengths and dimensions.*

We close this chapter by proving the following two results.

Lemma 3.3. Optimality of classical perfect codes over channels with memory *Let $\mathcal{C}$ be an (n, M, d_{min}) perfect code (in the classical sense) to be used* *over the general binary additive noise channel. Consider the set:*

$$\mathcal{D}_{\mathcal{C}} \;=\; \left\{ \mathbf{e} \in \mathbf{F}_2^n : w_H(\mathbf{e}) \leq r_{cov} = \left\lfloor \frac{d_{min}-1}{2} \right\rfloor \right\}.$$

Note that $\mathcal{D}_{\mathcal{C}}$ is the MD decoding set of $\mathcal{C}$. Consider the following condition Condition $(*)$: For any $\mathbf{e} \in \mathcal{D}_{\mathcal{C}}$ and any $\mathbf{e}' \in \mathbf{F}_2^n$, $w_H(\mathbf{e}) < w_H(\mathbf{e}') \implies P(\mathbf{Z} = \mathbf{e}) > P(\mathbf{Z} = \mathbf{e}')$.

If condition $(*)$ holds, then $\mathcal{C}$ is a generalized perfect code and hence is optimal among all codes of the same length and dimension under minimum generalized distance decoding (which is equivalent to ML decoding).

Proof. Let $\mathcal{C}$ be an (n, M, d_{min}) perfect code satisfying condition $(*)$. Its generalized covering radius is:

$$\begin{aligned}
\rho_{cov} &= \max_{\mathbf{y}\in\mathbf{F}_2^n} \min_{\mathbf{c}\in\mathcal{C}} D_n(\mathbf{c},\mathbf{y}) \\
&= \max_{\mathbf{y}\in\mathbf{F}_2^n} \min_{\mathbf{c}\in\mathcal{C}} W_n(\mathbf{c}\oplus\mathbf{y}) \\
&= \max_{\mathbf{y}\in\mathbf{F}_2^n} \min_{\mathbf{z}\in\mathcal{S}_{\mathcal{C}}(\mathbf{y})} W_n(\mathbf{z}) \\
&= \max_{\mathbf{y}\in\mathbf{F}_2^n} W_n\left(\mathbf{z}^*(\mathbf{y})\right), \qquad (3.7)
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{z}^*(\mathbf{y}) &= \arg \min_{\mathbf{z}\in\mathcal{S}_{\mathcal{C}}(\mathbf{y})} W_n(\mathbf{z}) \\
&= \arg \min_{\mathbf{z}\in\mathcal{S}_{\mathcal{C}}(\mathbf{y})} -\log_k \frac{P(\mathbf{Z}=\mathbf{z})}{P(\mathbf{Z}=\mathbf{0}^n)} \\
&= \arg \max_{\mathbf{z}\in\mathcal{S}_{\mathcal{C}}(\mathbf{y})} P(\mathbf{Z}=z).
\end{aligned}$$

Now since $\mathcal{C}$ is a perfect code in the classical sense, then $\forall \mathbf{y} \in \mathbf{F}_2^n, \exists$ a unique error pattern $\hat{\mathbf{z}}(\mathbf{y}) \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})$ (of minimal Hamming weight) such that $w_H\left((\hat{\mathbf{z}}(\mathbf{y})\right) \leq r_{cov}$. From condition $(*)$, $\forall \mathbf{z} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})$ such that $\mathbf{z} \neq \hat{\mathbf{z}}(\mathbf{y}), P(\mathbf{Z} = \hat{\mathbf{z}}(\mathbf{y})) > P(\mathbf{Z} = \mathbf{z})$. Hence, $\mathbf{z}^*(\mathbf{y}) = \hat{\mathbf{z}}(\mathbf{y})$. Since $\mathcal{C}$ is a perfect code, then:

$$\{\hat{\mathbf{z}}(\mathbf{y}) : \mathbf{y} \in \mathbf{F}_2^n\} \quad = \quad \{\mathbf{z} \in \mathbf{F}_2^n : w_H(\mathbf{z}) \leq r_{cov}\} = \mathcal{D}_{\mathcal{C}},$$

which is the set of all coset leaders for $\mathcal{C}$. Therefore,

$$\begin{aligned} \rho_{cov} &= \max_{\mathbf{y} \in \mathbf{F}_2^n} W_n\left(\hat{\mathbf{z}}(\mathbf{y})\right) \\ &= \max_{\mathbf{z} \in \mathcal{D}_{\mathcal{C}}} W_n\left(\mathbf{z}\right) \\ &= \max_{\mathbf{z} \in \mathcal{D}_{\mathcal{C}} : w_H(\mathbf{z}) = r_{cov}} W_n\left(\mathbf{z}\right). \end{aligned}$$

The last equality is a result of condition $(*)$.

We now prove that the generalized packing radius ρ_{pac} of $\mathcal{C}$ is the same as its generalized covering radius ρ_{cov}. By definition, $\rho_{pac} \leq \rho_{cov}$.

Assume $\rho_{pac} < \rho_{cov}$, then there exists at least one word $\mathbf{y} \in \mathbf{F}_2^n$ with two error patterns $\mathbf{z}_1$ and $\mathbf{z}_2 \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})$ such that $W_n(\mathbf{z_1}) \leq \rho_{cov}$ and $W_n(\mathbf{z_2}) \leq \rho_{cov}$. Now for any $\mathbf{z}' \in \mathbf{F}_2^n$,

$$\begin{aligned} W_n(\mathbf{z}') \leq \rho_{cov} \quad &\Longleftrightarrow \quad W_n(\mathbf{z}') \leq \max_{\mathbf{z} \in \mathcal{D}_{\mathcal{C}} : w_H(\mathbf{z}) = r_{cov}} W_n(\mathbf{z}) \\ &\Longleftrightarrow \quad P(\mathbf{Z} = \mathbf{z}') \geq \max_{\mathbf{z} \in \mathcal{D}_{\mathcal{C}} : w_H(\mathbf{z} = r_{cov})} P(\mathbf{Z} = z) \\ &\Longleftrightarrow \quad P(\mathbf{Z} = \mathbf{z}') \geq P(\mathbf{Z} = \mathbf{z}^*), \end{aligned}$$

where

$$\mathbf{z}^* = \arg \max_{\mathbf{z}\in\mathcal{D}_\mathcal{C}:w_H(\mathbf{z})=r_{cov}} P(\mathbf{Z}=\mathbf{z}).$$

Since $\mathbf{z}^* \in \mathcal{D}_\mathcal{C}$, from condition $(*)$:

$$P(\mathbf{Z}=\mathbf{z}') \geq P(\mathbf{Z}=\mathbf{z}^*) \quad \Longleftrightarrow \quad w_H(\mathbf{z}') \leq w_H(\mathbf{z}^*) = r_{cov}.$$

Therefore, both $\mathbf{z}_1$ and $\mathbf{z}_2$ have a Hamming weight of at most r_{cov} and they both belong to the error set of $\mathbf{y}$, $\mathcal{S}_\mathcal{C}(\mathbf{y})$. This is a contradiction since $\mathcal{C}$ is a perfect code and hence any error set can contain at most one error word with a Hamming weight less than or equal to the covering radius of the code. Therefore, $\rho_{pac} = \rho_{cov}$ and hence $\mathcal{C}$ is a generalized perfect code. □

Lemma 3.4. (Optimality of classical quasi-perfect codes over channels with memory) *Let $\mathcal{C}$ be an (n, M, d_{min}) quasi-perfect code (in the classical sense) to be used over the general binary additive noise channel. Consider the set:*

$$\Gamma_\mathcal{C} = \left\{\mathbf{e}\in\mathbf{F}_2^n : w_H(\mathbf{e}) \leq r_{cov} = \left\lfloor \frac{d_{min}-1}{2} \right\rfloor + 1\right\}.$$

Consider the following condition, Condition $(**)$: For any $\mathbf{e} \in \Gamma_\mathcal{C}$ and any $\mathbf{e}' \in \mathbf{F}_2^n$, $w_H(\mathbf{e}) < w_H(\mathbf{e}') \iff P(\mathbf{Z}=\mathbf{e}) > P(\mathbf{Z}=\mathbf{e}')$.

If condition $(**)$ holds, then $\mathcal{C}$ is a generalized quasi-perfect code and hence is optimal among all codes of the same length and dimension under minimum generalized distance decoding (which is equivalent to ML decoding).

Proof. Let $\mathcal{C}$ be an (n, M, d_{min}) quasi-perfect code that satisfies condition $(**)$.

Note that condition $(**)$ means that:

$$\begin{aligned} \forall \mathbf{z}_1, \mathbf{z}_2 \in \Gamma_{\mathcal{C}}, w_H(z_1) = w_H(z_2) \quad &\Longleftrightarrow \quad P(\mathbf{Z} = \mathbf{z}_1) = P(\mathbf{Z} = \mathbf{z}_2) \\ \text{and } w_H(z_1) < w_H(z_2) \quad &\Longleftrightarrow \quad P(\mathbf{Z} = \mathbf{z}_1) > P(\mathbf{Z} = \mathbf{z}_2). \end{aligned}$$

Hence, the generalized weight W_n of the error words in $\Gamma_{\mathcal{C}}$ is only a function of their Hamming weight. We will denote by $\omega_n(s)$ for $s \in \{0, 1, ..., r_{cov}\}$, the generalized weight of error words in $\Gamma_{\mathcal{C}}$ having a Hamming weight s. From condition $(**)$, $\omega_n(s)$ is strictly increasing in s.
From (3.7),

$$\rho_{cov} \;=\; \max_{\mathbf{y} \in \mathbf{F}_2^n} W_n\left(\mathbf{z}^*(\mathbf{y})\right),$$

where

$$\begin{aligned} \mathbf{z}^*(\mathbf{y}) &= \arg \max_{\mathbf{z} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} P(\mathbf{Z} = z) \\ &= \arg \min_{\mathbf{z} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})} w_H(\mathbf{z}), \end{aligned}$$

where the last equality follows from condition $(**)$. Define

$$d_{min}^H(\mathbf{y}) = \min_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}, \mathbf{c}),$$

where $d_H(\cdot, \cdot)$ denotes the Hamming distance. Then,

$$\begin{aligned} \rho_{cov} &= \max_{\mathbf{y} \in \mathbf{F}_2^n} \omega_n\left(d_{min}^H(\mathbf{y})\right) \\ &= \omega_n\left(\max_{\mathbf{y} \in \mathbf{F}_2^n} d_{min}^H(\mathbf{y}\right) \end{aligned} \tag{3.8}$$

$$= \quad \omega_n(r_{cov}). \tag{3.9}$$

The equality in (3.8) follows from the fact that $\omega_n(\cdot)$ is a strictly increasing function, whereas (3.9) follows immediately from Definition 3.6.

By definition, we know that $\rho_{pac} \leq \rho_{cov} = \omega_n(r_{cov})$. Hence, from condition $(**)$, $\rho_{pac} = \omega_n(s)$ for some $s \leq r_{cov}$. We want to prove that $\rho_{pac} = \omega_n(r_{pac})$; i.e., that the latter s is actually $s = r_{pac}$.

- Assume $\rho_{pac} = \omega_n(s)$ for some $s < r_{pac}$, which means that $\rho_{pac} < \omega_n(r_{pac})$. Thus, by the definiton of ρ_{pac}, there exists at least one word $\mathbf{y} \in \mathbf{F}_2^n$ with two error patterns $\mathbf{z}_1$ and $\mathbf{z}_2 \in \mathcal{S}_C(\mathbf{y})$ such that $W_n(\mathbf{z}_1) \leq \omega_n(r_{pac})$ and $W_n(\mathbf{z}_2) \leq \omega_n(r_{pac})$. From condition $(**)$, we deduce that $w_H(\mathbf{z}_1) \leq r_{pac}$ and $w_H(\mathbf{z}_2) \leq r_{pac}$. This is a contradiction since for any $\mathbf{y} \in \mathbf{F}_2^n$, there exists at most one error word $\mathbf{e} \in \mathcal{S}_{\mathcal{C}}(\mathbf{y})$ such that $w_H(\mathbf{e}) \leq r_{pac}$. Therefore, $\rho_{pac} = \omega_n(s)$ for some $r_{pac} \leq s \leq r_{cov} = r_{pac} + 1$.

- Since $\omega_n(\cdot)$ is a strictly increasing function over the set $\mathcal{S} = \{0, 1, ..., r_{cov}\}$, then:

  $$r_{pac} \leq s \leq r_{cov} = r_{pac} + 1 \quad \Longleftrightarrow \quad \omega_n(r_{pac}) \leq \rho_{pac} = \omega_n(s) \leq \rho_{cov} = \omega_n(r_{cov}).$$

  Since $\nexists i \in \mathcal{S}$ such that $r_{pac} < i < r_{cov}$ (since $r_{cov} = r_{pac}+1$), then $\nexists t \in \mathcal{K}(D_n)$ such that $\omega_n(r_{pac}) < t < \rho_{cov}$. Therefore, we have either $\rho_{pac} = \rho_{cov}$ which means that $\mathcal{C}$ is a generalized perfect code or $\rho_{pac} = \omega_n(r_{pac})$ which means that $\mathcal{C}$ is a generalized quasi-perfect code.

- Since $r_{cov} = r_{pac} + 1 > r_{pac}$, $\exists$ at least one word $\mathbf{y} \in \mathbf{F}_2^n$ with two error patterns $\mathbf{z}_1$ and $\mathbf{z}_2 \in \mathcal{S}_C(\mathbf{y})$ such that $w_H(\mathbf{z}_1) \leq r_{cov}$ and $w_H(\mathbf{z}_2) \leq r_{cov}$. From condition $(**)$, $W_n(\mathbf{z}_1) \leq \omega_n(r_{cov}) = \rho_{cov}$ and $W_n(\mathbf{z}_2) \leq \omega_n(r_{cov}) = \rho_{cov}$. From the definition of the generalized packing radius, $\rho_{pac} < \omega_n(r_{cov})$. Therefore, $\rho_{pac} = \omega_n(r_{pac})$ and hence $\mathcal{C}$ is a generalized quasi-perfect code.

□

Chapter 4

Optimality of MD Decoding over Infinite Memory Channels

In this chapter, we study two infinite-memory channel models: the IMCC and the GEC. We determine necessary and sufficient conditions on binary codes under which minimum Hamming distance decoding is equivalent to ML decoding over the IMCC. We start by re-deriving few properties of ML decoding over the Polya channel that are useful in obtaining the conditions on binary codes. We also show that under these conditions, classical perfect and quasi-perfect codes are also generalized perfect and quasi-perfect codes, and hence are optimal for the IMCC among all codes of the same lengths and dimensions. For the GEC, we study separately the case when the state vector is unknown and known at the decoder. In the former case, we present sufficient conditions on binary codes under which strict MD decoding and ML decoding are equivalent. In the second case (when the state vector is available at the decoder), we present partial results pertaining to the equivalence between the Hamming weight and the likelihood of error patterns.

4.1 Properties of ML Decoding over the IMCC

The IMCC was introduced in [1]. Let $\mathbf{z}_1^n \in \mathbf{F}^n$ be an error word generated by this channel. The probability of this word is given by (2.4):

$$P(\mathbf{Z} = \mathbf{z}_1^n) = \frac{\Gamma(\frac{1}{\delta})\Gamma(\frac{p}{\delta}+d)\Gamma(\frac{1-p}{\delta}+n-d)}{\Gamma(\frac{p}{\delta})\Gamma(\frac{1-p}{\delta})\Gamma(\frac{1}{\delta}+n)}$$

where d is the Hamming weight of $\mathbf{z}_1^n$, p is the channel BER and $\delta > 0$ is a channel parameter related to the noise correlation coefficient ϵ as follows:

$$\epsilon = \frac{\delta}{1+\delta}.$$

We can clearly see that the probability distribution of error words is only a function of their Hamming weights. Hence, error patterns with the same Hamming weight occur with the same probability. Denote by $\mathbf{z}_1^n(t)$ an error word with Hamming weight t. For $0 \leq m \leq n-1$ and $1 \leq i \leq n-m$, we define $\rho(m,i)$ as follows:

$$\begin{aligned}
\rho(m,i) &:= \frac{\mathrm{P}\left(\mathbf{Z}_1^n = \mathbf{z}_1^n(m+i)\right)}{\mathrm{P}\left(\mathbf{Z}_1^n = \mathbf{z}_1^n(m)\right)} \\
&= \frac{\Gamma\left(\frac{p}{\delta}+m+i\right)\Gamma\left(\frac{1-p}{\delta}+n-m-i\right)}{\Gamma\left(\frac{p}{\delta}+m\right)\Gamma\left(\frac{1-p}{\delta}+n-m\right)} \\
&= \prod_{j=1}^{i}\left(\frac{\frac{p}{\delta}+m+i-j}{\frac{1-p}{\delta}+n-m-j}\right),
\end{aligned}$$

where the last equality follows from the identity:

$$\Gamma(x+1) = x\Gamma(x).$$

We can easily see that:

$$\begin{aligned} \rho(m,i) \le 1 \quad &\Longleftrightarrow \quad \frac{p}{\delta} + m + i \le \frac{1-p}{\delta} + n - m \\ &\Longleftrightarrow \quad m + i - \mu \le \mu - m \\ &\Longleftrightarrow \quad |m+i-\mu| \le |m-\mu|, \end{aligned} \tag{4.1}$$

where

$$\mu = \frac{1-2p}{2\delta} + \frac{n}{2}. \tag{4.2}$$

Lemma 4.1. *Let l, i, k be three integers such that: $1 \le k < i \le n - l$ and $0 \le l \le n-2$, then $P\left(\mathbf{Z} = \mathbf{z}_1^n(l+k)\right) \le P\left(\mathbf{Z} = \mathbf{z}_1^n(l)\right)$ or $P\left(\mathbf{Z} = \mathbf{z}_1^n(l+k)\right) \le P\left(\mathbf{Z} = \mathbf{z}_1^n(l+i)\right)$.*

Proof. The above claim is true when:

$$\begin{aligned} P\left(\mathbf{Z} = \mathbf{z}_1^n(l+k)\right) \le P\left(\mathbf{Z} = \mathbf{z}_1^n(l)\right) \text{ or } P\left(\mathbf{Z} = \mathbf{z}_1^n(l+k)\right) \le P\left(\mathbf{Z} = \mathbf{z}_1^n(l+i)\right) \\ \Longleftrightarrow \rho(l,k) \le 1 \text{ or } \rho(l+k, i-k) \ge 1 \\ \Longleftrightarrow |l+k-\mu| \le |l-\mu| \text{ or } |l+k-\mu| \le |l+i-\mu| \end{aligned}$$

where the equivalence in the last step follows from (4.1). Finally noting that the expression in (4.3) always holds concludes the proof. □

Theorem 4.1. *ML decoding over the IMCC is equivalent to either minimum distance decoding or maximum distance decoding.*

Proof. Let $\mathbf{y}_1^n$ be the received word. We order the error words in the error set $\mathcal{S}_{\mathcal{C}}(\mathbf{y})$ in increasing order of their Hamming weights, i.e., we define the ordered sequence $\{\mathbf{z}_1^n[i]\}_{i=1}^{|\mathcal{C}|}$ as follows:

$$\{\mathbf{z}_1^n[i]\}_{i=1}^{|\mathcal{C}|} \;=\; \{\mathbf{z}_1^n[i] \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)\colon\, w_H\left(\mathbf{z}_1^n[i]\right) \leq w_H\left(\mathbf{z}_1^n[j]\right) \quad \forall i,j : 1 \leq i \leq j \leq |\mathcal{C}|\}.$$

Define:

$$\begin{aligned} d_{min}(\mathbf{y}_1^n) &= \min_{\mathbf{z}_1^n \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)} w_H(\mathbf{z}_1^n) \\ &= w_H\left(\mathbf{z}_1^n[1]\right) \qquad (4.3) \\ d_{max}(\mathbf{y}_1^n) &= \max_{\mathbf{z}_1^n \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)} w_H(\mathbf{z}_1^n) \\ &= w_H\left(\mathbf{z}_1^n[|\mathcal{C}|]\right). \qquad (4.4) \end{aligned}$$

From Lemma 4.1, we conclude that

$$P(\mathbf{Z} = \mathbf{z}_1^n(i)) \leq P(\mathbf{Z} = \mathbf{z}_1^n(1)) \text{ or } P(\mathbf{Z} = \mathbf{z}_1^n(i)) \leq P(\mathbf{Z} = \mathbf{z}_1^n(|\mathcal{C}|)),$$

for $i \in \{1, ..., |\mathcal{C}|\}$.

Therefore, the most likely error pattern is either the one with the minimum weight or the one with the maximum weight which concludes the proof.

The decoder chooses $\mathbf{z}_1^n[1]$ when:

$$P(\mathbf{Z} = \mathbf{z}_1^n[1]) \geq P(\mathbf{Z} = \mathbf{z}_1^n[|\mathcal{C}|]) \quad \Longleftrightarrow \quad |d_{min}(\mathbf{y}_1^n) - \mu| \geq |d_{max}(\mathbf{y}_1^n) - \mu|.$$

□

The result in Theorem 4.1 can also be found in the original paper introducing the channel model [1]. However, we have provided an alternative proof for the same result.

4.2 On the Equivalence of MD and ML Decoding over the IMCC

Theorem 4.2. *For any (n, M, d_{min}) code $\mathcal{C}$ used over the IMCC, if the (classical) covering radius of this code*

$$\begin{aligned} r_{cov} &\leq \frac{1-2p}{\delta} \\ &= \frac{(1-2p)(1-\epsilon)}{\epsilon}, \end{aligned}$$

then the outputs of the MD and ML decoders are identical.

Proof. Let $\mathbf{y}_1^n$ be the received word. We know that $d_{min}(\mathbf{y}_1^n) \leq r_{cov}$ from the definition of the covering radius (with equality achieved for at least one word). On the other hand, $d_{max}(\mathbf{y}_1^n) \leq n$. (Note that there are codes for which both inequalities can be satisfied with equality for the same word $\mathbf{y}_1^n$.) Hence, $|d_{max}(\mathbf{y}_1^n) - \mu| \leq \frac{n}{2} - \frac{1-2p}{2\delta}$ and $|d_{min}(\mathbf{y}_1^n) - \mu| \geq \frac{n}{2} - \frac{1-2p}{2\delta}$, where μ is defined in (4.2) and we have used the assumption that $r_{cov} \leq \frac{1-2p}{\delta}$. Therefore, for any received word $\mathbf{y}$, $|d_{min}(\mathbf{y}_1^n) - \mu| \geq |d_{max}(\mathbf{y}_1^n) - \mu|$ which means that the MD decoding rule is always used. □

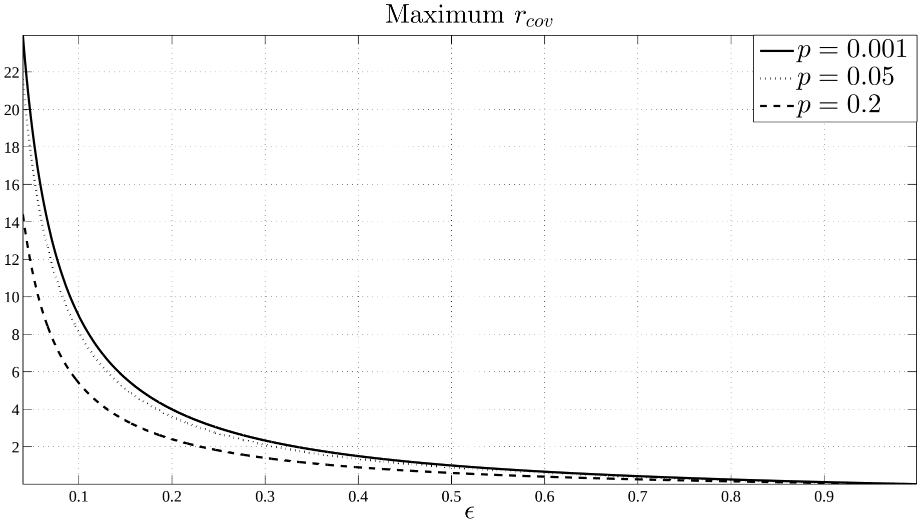


Figure 4.1: Plot of the condition in Theorem 4.2: maximum allowable r_{cov} over the IMCC with respect to the correlation coefficient ϵ and for different values of the BER p.

We illustrate the condition of Theorem 4.2 in Fig. 4.2 by plotting the maximum allowable value for a code's covering radius over the IMCC.

We can tighten the condition in Theorem 4.2 to obtain a necessary and sufficient condition on the code $\mathcal{C}$.

Lemma 4.2. *Let* $\mathbf{y}_1^n$ *be the received word, then*

$$d_{max}(\mathbf{y}_1^n) \quad = \quad n - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n),$$

where d_{min} *and* d_{max} *are defined in (4.3) and (4.4), respectively, and where* $\mathbf{1}^n = (1, ..., 1)$ *is the all-one word of length* n*.*

Proof. First, it is easy to see that $d_H(\mathbf{y}_1^n, \mathbf{c}) = n - d_H(\mathbf{1}^n \oplus \mathbf{y}_1^n, \mathbf{c})$. Then,

$$d_{max}(\mathbf{y}_1^n) \quad = \quad \max_{\mathbf{c} \in \mathcal{C}} d_H(\mathbf{y}_1^n, \mathbf{c})$$

$$
\begin{aligned}
&= \max_{\mathbf{c}\in\mathcal{C}} \{n - d_H(\mathbf{1}^n \oplus \mathbf{y}_1^n, \mathbf{c})\} \\
&= n - \min_{\mathbf{c}\in\mathcal{C}} d_H(\mathbf{1}^n \oplus \mathbf{y}_1^n, \mathbf{c}) \\
&= n - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n).
\end{aligned}
$$

□

Definition 4.1. *Let $\mathbf{y}_1^n$ be the received word. We define:*

$$
d_{sum}(\mathbf{y}_1^n) \;:=\; |d_{min}(\mathbf{y}_1^n) - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n)|,
$$

and let

$$
d_{sum}(\mathcal{C}) \;:=\; \max_{\mathbf{y}_1^n \in \mathbf{F}_2^n} d_{sum}(\mathbf{y}_1^n).
$$

Theorem 4.3. *For any (n, M, d_{min}) code $\mathcal{C}$ used over the IMCC, the outputs of the MD and ML decoders are identical iff*

$$
\begin{aligned}
d_{sum}(\mathcal{C}) &\leq \frac{1-2p}{\delta} \\
&= \frac{(1-2p)(1-\epsilon)}{\epsilon}.
\end{aligned}
$$

Proof. We start by proving the first direction ($\Longrightarrow$):

Assume $d_{sum}(\mathcal{C}) \leq \frac{1-2p}{\delta}$, and let $\mathbf{y}_1^n$ be the received word. Then:

$$
d_{min}(\mathbf{y}_1^n) + d_{max}(\mathbf{y}_1^n) \;=\; d_{min}(\mathbf{y}_1^n) + n - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n)
$$

$$
\begin{aligned}
&\leq n + d_{sum}(\mathbf{y}_1^n) \\
&\leq n + d_{sum}(\mathcal{C}) \\
&\leq n + \frac{1-2p}{\delta} \\
&= 2\mu.
\end{aligned}
$$

Hence,

$$d_{min}(\mathbf{y}_1^n) + d_{max}(\mathbf{y}_1^n) \leq 2\mu \quad \Longrightarrow \quad |d_{min}(\mathbf{y}_1^n) - \mu| \geq |d_{max}(\mathbf{y}_1^n) - \mu|,$$

which means (in light of Theorem 4.1) that for every $\mathbf{y}_1^n \in \mathbf{F}_2^n$, the ML decoder picks the minimum Hamming weight error word and hence reduces to the MD decoder.

We now prove the other direction ($\Longleftarrow$):
Assume $d_{sum}(\mathcal{C}) > \frac{1-2p}{\delta}$. From the definition of $d_{sum}(\mathcal{C})$, we know that there exist at least one word $\bar{\mathbf{y}}_1^n$ such that:

$$d_{min}(\bar{\mathbf{y}}_1^n) - d_{min}(\mathbf{1^n} \oplus \bar{\mathbf{y}}_1^n) \quad > \quad \frac{1-2p}{\delta}.$$

For this received word, we have

$$
\begin{aligned}
d_{min}(\bar{\mathbf{y}}_1^n) + d_{max}(\bar{\mathbf{y}}_1^n) &= d_{min}(\bar{\mathbf{y}}_1^n) + n - d_{min}(\mathbf{1}^n \oplus \bar{\mathbf{y}}_1^n) \\
&> n + \frac{1-2p}{\delta} \\
&= 2\mu.
\end{aligned}
$$

Hence,

$$d_{min}(\bar{\mathbf{y}}_1^n) + d_{max}(\bar{\mathbf{y}}_1^n) > 2\mu \quad \Longrightarrow \quad |d_{min}(\bar{\mathbf{y}}_1^n) - \mu| < |d_{max}(\bar{\mathbf{y}}_1^n) - \mu|.$$

Which means that for this received word $\bar{\mathbf{y}}_1^n$, the ML decoder picks the maximum Hamming weight error word and hence the ML decoder is not equivalent to the MD decoder.

□

Corollary 4.1. *Let $\mathcal{C}$ be an (n, M, d_{min}) perfect (quasi-perfect) code, in the classical sense, used over the IMCC. If*

$$\begin{aligned} r_{cov} &\leq \frac{1-2p}{\delta} &(4.5)\\ &= \frac{(1-2p)(1-\epsilon)}{\epsilon} &(4.6)\end{aligned}$$

then $\mathcal{C}$ is a generalized perfect (generalized quasi-perfect) code for this channel and hence is optimal under ML decoding among all codes of the same length and dimension sent over the same channel.

Proof. Immediate from Lemma 3.3 (3.4) and Theorem 4.2. □

Remark 4.1. *It is worth noting that Theorem 4.2 implies Theorem 4.3. Indeed, assume that the condition $r_{cov} \leq \frac{1-2p}{\delta}$ holds. This means that for any received word $\mathbf{y}_1^n$:*

$$0 \leq d_{min}(\mathbf{y}_1^n) \leq r_{cov} \leq \frac{1-2p}{\delta}.$$

Therefore,

$$\begin{aligned}
-\frac{1-2p}{\delta} \leq d_{min}(\mathbf{y}_1^n) - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n) &\leq \frac{1-2p}{\delta} \\
\iff |d_{min}(\mathbf{y}_1^n) - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n)| &\leq \frac{1-2p}{\delta} \\
\iff d_{sum}(\mathcal{C}) &\leq \frac{1-2p}{\delta},
\end{aligned}$$

which is the condition in Theorem 4.3.

Remark 4.2. *The original paper introducing the contagion channel [1] gives a sufficient condition under which, for any (n, M, d) binary code, the ML decoder becomes equivalent to the MD decoder. The condition is:*

$$n - 1 \quad < \quad \frac{1-2p}{\delta}. \tag{4.7}$$

Comparing it to the condition that we give in Theorem 4.2, we can clearly see that the latter is much tighter. In fact, the only time the condition in Theorem 4.2 is as restrictive as the one given in (4.7) is when the covering radius of the code $\mathcal{C}$ is $n - 1$, which is almost never the case for most codes of interest.

There is another condition in [5] for binary *linear* codes under which ML and MD decoding are equivalent.

Lemma 4.3. **[5]** *For a linear code containing the all-one codeword, if $p < 0.5$ then ML decoding over the IMCC reduces to MD decoding.*

We explain the above result by using our condition in Theorem 4.3.
Let $\mathcal{C}$ be an $[n, k]$ binary linear code that contains the all-one codeword $\mathbf{1}^n$ and let $\mathbf{y}_1^n$ be the received word.

Since $\mathbf{1}^n \in \mathcal{C}$ and since $\mathcal{C}$ is a linear code, then $(\mathbf{1}^n \oplus \mathbf{y}_1^n)$ and $\mathbf{y}_1^n$ belong to the same coset and hence:

$$d_{min}(\mathbf{y}_1^n) = d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n) = w_H(\bar{\mathbf{z}}_1^n),$$

where $\bar{\mathbf{z}}_1^n$ is the corresponding coset leader. Hence, for any received word $\mathbf{y}_1^n$:

$$\begin{aligned} d_{sum}(\mathbf{y}) &= |d_{min}(\mathbf{y}_1^n) - d_{min}(\mathbf{1}^n \oplus \mathbf{y}_1^n)| \\ &= 0. \end{aligned}$$

Therefore, $d_{sum}(\mathcal{C}) = 0 < \frac{1-2p}{\delta}$ iff $p < 0.5$ which is not a restrictive assumption.

4.3 Numerical Results

We illustrate the condition of Theorem 4.3 by simulating the performance of the $[7, 4, 3]$ or $(7, 2^4, 3)$ Hamming code, the $[8, 4, 4]$ Reed-Muller code and a $(15, 2^{11}, 3)$ perfect nonlinear code under ML and MD decoding over the IMCC. The $[7, 4, 3]$ Hamming code, the $[8, 4, 4]$ Reed-Muller code are both linear codes containing the all-one codeword. Hence, it follows from Lemma 4.3 that ML decoding and MD decoding are identical for these two codes over the IMCC. This is indeed shown in Tables 4.1 and 4.2 where we tabulate the probability of codeword error (PCE) for both codes under ML and MD decoding over the IMCC with parameters $p = 0.1$ and for different values of ϵ. We notice as expected that we get exactly the same performance under MD and ML decoding since they are identical for $p < 0.5$.

Finally, we simulate the performance of the $(15, 2^{11}, 3)$ Vasil'ev nonlinear per-

ϵ	0.2	0.3	0.4	0.5	0.6	0.7
max $d_{sum}(\mathcal{C})$	3.2	1.8667	1.2	0.8	0.5333	0.3429
PCE under MD	0.17747	0.18514	0.165086	0.167472	0.149	0.139856
PCE under ML	0.17747	0.18514	0.165086	0.167472	0.149	0.139856

Table 4.1: Verifying Theorem 4.3 and Lemma 4.3 for the $[7, 4, 3]$ Hamming code over the IMCC with parameters $p = 0.1$ and ϵ.

ϵ	0.1	0.2	0.25	0.3	0.4	0.5
max $d_{sum}(\mathcal{C})$	7.2	3.2	2.4	1.8667	1.2	0.8
PCE under MD	0.0717504	0.0895768	0.0914395	0.100567	0.0930972	0.100681
PCE under ML	0.0717504	0.0895768	0.0914395	0.100567	0.0930972	0.100681

Table 4.2: Verifying Theorem 4.3 and Lemma 4.3 for the $[8, 4, 4]$ Reed-Muller code $\mathcal{RM}(1, 3)$ over the IMCC with parameters $p = 0.1$ and ϵ.

fect code $\mathcal{C}$ constructed using the method described in Theorem 3.1 using

$$f(\mathbf{x}_1^7) = x_1x_2x_7 \oplus x_2x_4x_6 \oplus x_1x_3x_5.$$

According to Theorem 3.2, if $\mathbf{c} \in \mathcal{C}$ then $\mathbf{1}^{15} \oplus \mathbf{c} \in \mathcal{C}$ as well. It is easy to generalize the result in Lemma 4.3 as follows:

Lemma 4.4. *Consider any (n, M, d) code $\mathcal{C}$ that includes every codeword along with its complement, i.e., if $\mathbf{c} \in \mathcal{C}$ then $(\mathbf{1}^n \oplus \mathbf{c}) \in \mathcal{C}$. If $p < 0.5$ then ML decoding over the IMCC reduces to MD decoding.*

We verify Lemma 4.4 in Table 4.3 where we can see that the PCE under ML is identical to the PCE under MD.

ϵ	0.2	0.3	0.4	0.5	0.6	0.7
max $d_{sum}(\mathcal{C})$	3.2	1.8667	1.2	0.8	0.5333	0.3429
PCE under MD	0.2656	0.2308	0,18166	0.2126	0.2002	0.1604
PCE under ML	0.2656	0.2308	0,18166	0.2126	0.2002	0.1604

Table 4.3: Verifying Theorem 4.3 and Lemma 4.4 for the $(15, 2^{11}, 3)$ Vasil'ev nonlinear code over the IMCC with parameters $p = 0.1$ and ϵ.

4.4 On the Equivalence of MD and ML decoding over the GEC

The GEC, introduced in [9, 12], is a widely used model of burst-noise binary channels. It belongs to the class of finite-state Markov channels. Let $\mathbf{z}_1^n$ be an error word generated by this channel. The probability of this word is given (as a matrix product) by (2.17):

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \quad = \quad \boldsymbol{\pi}^T \left(\prod_{i=1}^{n} \mathbf{P}(z_i) \right) \underline{1}^n,$$

where $\boldsymbol{\pi}$ and $\mathbf{P}(\cdot)$ are given by (2.14), (2.15) and (2.16) and $\underline{1}^n$ is the all-one column vector of length n. Let $\mathbf{s}_1^n \in \mathbf{F}_2^n$ be a state vector where for convenience we set binary values to the two states: $s = 0$ for state G and $s = 1$ for state B. The joint distribution of the error word and the state vector is given by:

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n) \quad &= \quad \pi_0^{1-s_0}\pi_1^{s_0} \prod_{i=1}^{n} \left[(1-P_B)^{1-z_i} P_B^{z_i}\right]^{s_i} \left[(1-P_G)^{1-z_i} P_G^{z_i}\right]^{1-s_i} \\ &\quad \prod_{i=2}^{n} (1-b)^{(1-s_{i-1})(1-s_i)} b^{(1-s_{i-1})s_i} g^{s_{i-1}(1-s_i)} (1-g)^{s_{i-1}s_i} \\ &= \quad L(\mathbf{s}_1^n) \prod_{i=1}^{n} \left[(1-P_B)^{1-z_i} P_B^{z_i}\right]^{s_i} \left[(1-P_G)^{1-z_i} P_G^{z_i}\right]^{1-s_i}, \end{aligned}$$

where

$$L(\mathbf{s}_1^n) \;\;=\;\; \pi_0^{1-s_0}\pi_1^{s_0}\prod_{i=2}^{n}(1-b)^{(1-s_{i-1})(1-s_i)}b^{(1-s_{i-1})s_i}g^{s_{i-1}(1-s_i)}(1-g)^{s_{i-1}s_i}$$

depends only on the state vector. We assume throughout that $P_G < P_B < \frac{1}{2}$.

4.4.1 Case 1: State Vector is Unknown at the Decoder

Theorem 4.4. *Define:*

$$\hat{m}(n, P_G, P_B) \;\;:=\;\; \frac{\log\left(\dfrac{1-P_G}{P_B}\right) - n\log\left(\dfrac{1-P_G}{1-P_B}\right)}{\log\left(\dfrac{P_B(1-P_B)}{P_G(1-P_G)}\right)}.$$

For any two error words $\mathbf{z}_1^n$ and $\bar{\mathbf{z}}_1^n$ generated by the GEC satisfying

- $w_H(\mathbf{z}_1^n) = m$, *where* $0 \le m \le n-1$
- $w_H(\bar{\mathbf{z}}_1^n) = m+i$, *where* $1 \le i \le n-m$

where $w_H(\cdot)$ denotes the Hamming weight, we have that

$$m < \hat{m}(n, P_G, P_B) \quad\Longrightarrow\quad P\left(\mathbf{Z}_1^n = \mathbf{z}_1^n\right) > P\left(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n\right).$$

Proof. Let $\mathbf{s}_1^n \in \mathbf{F}_2^n$ be a state vector. Hence,

$$\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)} \;\;=\;\; \prod_{j=1}^{n}\left[\frac{(1-P_B)^{1-\bar{z}_j}P_B^{\bar{z}_j}}{(1-P_B)^{1-z_j}P_B^{z_j}}\right]^{s_j}\left[\frac{(1-P_G)^{1-\bar{z}_j}P_G^{\bar{z}_j}}{(1-P_G)^{1-z_j}P_G^{z_j}}\right]^{1-s_j}$$

$$\begin{aligned} &= \frac{P_B^{n_1(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)} P_G^{m+i-n_1(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)} (1-P_B)^{n_2(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)}}{P_B^{n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} P_G^{m-n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} (1-P_B)^{n_2(\mathbf{z}_1^n,\mathbf{s}_1^n)}} \\ &\quad \frac{(1-P_G)^{n-m-i-n_2(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)}}{(1-P_G)^{n-m-n_2(\mathbf{z}_1^n,\mathbf{s}_1^n)}} \\ &= \frac{P_B^{n_1(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)} P_G^{i-n_1(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)} (1-P_B)^{n_2(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n)}}{P_B^{n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} P_G^{-n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} (1-P_B)^{n_2(\mathbf{z}_1^n,\mathbf{s}_1^n)}} \\ &\quad \frac{(1-P_G)^{-n_2(\mathbf{z}_1^n,\mathbf{s}_1^n)}}{(1-P_G)^{i-n_2(\mathbf{z}_1^n,\mathbf{s}_1^n)}}, \end{aligned}$$

where

$$\begin{aligned} n_1(\mathbf{z}_1^n,\mathbf{s}_1^n) &= |\{j \in \{1,..,n\} : z_j = 1 \text{ and } s_j = 1\}| \\ n_2(\mathbf{z}_1^n,\mathbf{s}_1^n) &= |\{j \in \{1,..,n\} : z_j = 0 \text{ and } s_j = 1\}|. \end{aligned}$$

Clearly,

$$\begin{aligned} 0 \leq \quad n_1(\mathbf{z}_1^n,\mathbf{s}_1^n) &\leq w_H(\mathbf{z}_1^n) = m \\ 0 \leq \quad n_2(\mathbf{z}_1^n,\mathbf{s}_1^n) &\leq n - w_H(\mathbf{z}_1^n) = n-m. \end{aligned}$$

Since $P_G < P_B < \frac{1}{2}$,

$$\begin{aligned} \frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)} &\leq \frac{P_B^{m+i}(1-P_G)^{n-m-i}}{P_G^m(1-P_B)^{n-m}} && (4.8) \\ &= \left[\frac{(1-P_G)}{(1-P_B)}\right]^n \left[\frac{P_B(1-P_B)}{P_G(1-P_G)}\right]^m \left[\frac{P_B}{1-P_G}\right]^i \\ &\leq \left[\frac{(1-P_G)}{(1-P_B)}\right]^n \left[\frac{P_B(1-P_B)}{P_G(1-P_G)}\right]^m \left[\frac{P_B}{1-P_G}\right] && (4.9) \\ &< 1, && (4.10) \end{aligned}$$

where (4.8) is obtained by setting $n_1(\mathbf{z}_1^n,\mathbf{s}_1^n) = 0, n_2(\mathbf{z}_1^n,\mathbf{s}_1^n) = n-m, n_1(\bar{\mathbf{z}}_1^n,\mathbf{s}_1^n) =$

$m + i$ and $n_2(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) = 0$. Also, (4.9) is obtained by upper-bounding $\left[\frac{P_B}{1-P_G}\right]^i$ by $\left[\frac{P_B}{1-P_G}\right]$ since $P_B < 1-P_G$. Finally, the last inequality is a result of the assumption that $m < \hat{m}(n, P_G, P_B)$. Hence, we established that under the condition $m < \hat{m}(n, P_G, P_B)$ for any state vector $\mathbf{s}_1^n$, $P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n) < P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)$. Therefore,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)} &= \frac{\sum_{\mathbf{s}_1^n \in \mathbf{F}_2^n} P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)}{\sum_{\mathbf{s}_1^n \in \mathbf{F}_2^n} P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)} \\
&< \frac{\sum_{\mathbf{s}_1^n \in \mathbf{F}_2^n} P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)}{\sum_{\mathbf{s}_1^n \in \mathbf{F}_2^n} P(\mathbf{Z}_1^n = \mathbf{z}_1^n, \mathbf{S}_1^n = \mathbf{s}_1^n)} \\
&= 1,
\end{aligned}$$

which concludes the proof. □

Theorem 4.5. *Let $\mathcal{C}$ be any (n, M, d) code used over the GEC. Denote by r_{cov} the classical covering radius of this code. If $r_{cov} < \hat{m}(n, P_G, P_B)$, then the output of the SMD decoder (when it does not declare a decoding failure) is identical to the output of the ML decoder for this code.*

Proof. Let $\mathbf{y}_1^n$ be the received word. Let

$$m := \min_{\mathbf{z}_1^n \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)} w_H(\mathbf{z}_1^n),$$

where $w_H(\cdot)$ denotes the Hamming weight. Clearly, $m \le r_{cov} < \hat{m}(n, P_G, P_B)$. If there exists a unique error word $\hat{\mathbf{z}}_1^n \in \mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)$ such that $w_H(\hat{\mathbf{z}}_1^n) = m$, then the SMD decoder does not declare a failure and outputs $\hat{\mathbf{z}}_1^n$. Since $m < \hat{m}(n, P_G, P_B)$, it follows from Theorem 4.4 that $\hat{\mathbf{z}}_1^n$ is the most likely error word in $\mathcal{S}_{\mathcal{C}}(\mathbf{y}_1^n)$ and hence will also be the output of the ML decoder. Therefore, SMD and ML decoding are equivalent. □

Corollary 4.2. *Let $\mathcal{C}$ be an (n, M, d_{min}) perfect code (in the classical sense) used over a GEC with parameters P_B and P_G. If*

$$r_{cov} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor \; < \; \hat{m}(n, P_G, P_B),$$

then $\mathcal{C}$ is a generalized perfect code for this channel and hence is optimal among all codes of the same lengh and dimension sent over the same channel.

Proof. It follows from Lemma 3.3 and Theorem 4.4. □

Remark 4.3. *Note that we cannot make a statement similar to the one in Corollary 4.2 for quasi-perfect codes, since two error words of the same weight are not guaranteed to have the same probability.*

In Fig. 4.2, we illustrate the condition in Theorem 4.5 by plotting $\hat{m}(n, P_G, P_B)$ with respect to P_B for $P_G = 0.001$ and different values of n. We notice from Fig. 4.2 that the sufficient condition in Theorem 4.5 becomes increasingly restrictive as the value P_B further exceeds that of P_G. Indeed, for $P_G = 0.001$ when $P_B > 0.03$ all codes (with $n \geq 5$) stop satisfying the condition. We can also see that when $P_B \to P_G^+$ (i.e., when the GEC tends to the BSC), unsurprisingly all codes satisfy the condition. It is worth noting that the condition is independent of the underlying Markov chain parameters (b and g) which is a factor in it being stringent.

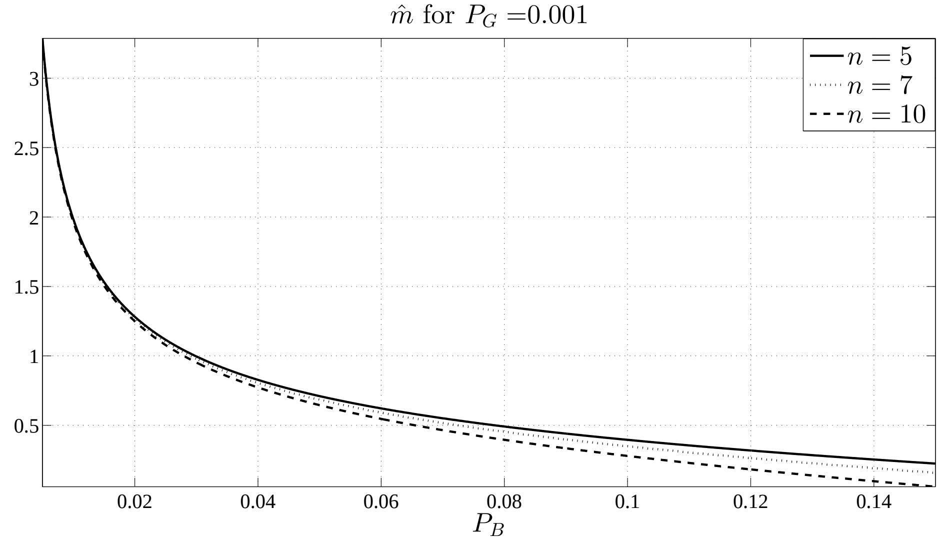


Figure 4.2: Plot of $\hat{m}(n, P_G, P_B)$ for the GEC with respect to P_B for different values of n and for $P_G = 0.001$.

4.4.2 Case 2: State Vector is Known at the Decoder

In this case, the probability of an error pattern $\mathbf{z}_1^n$ given a state vector $\mathbf{s}_1^n$ known at the decoder is given by:

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n) &= \prod_{i=1}^{n} \left[(1-P_B)^{1-z_i} P_B^{z_i}\right]^{s_i} \left[(1-P_G)^{1-z_i} P_G^{z_i}\right]^{1-s_i} \\ &= P_B^{n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} P_G^{m-n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} (1-P_B)^{t-n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} \\ &\quad (1-P_G)^{n-m-t+n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} \\ &= \left[\underbrace{\frac{P_B(1-P_G)}{P_G(1-P_B)}}_{\geq 1}\right]^{n_1(\mathbf{z}_1^n,\mathbf{s}_1^n)} \left[\frac{P_G}{1-P_G}\right]^m \left[\frac{1-P_B}{1-P_G}\right]^t (1-P_G)^n, \end{aligned}$$

where

$$n_1(\mathbf{z}_1^n, \mathbf{s}_1^n) = |\{j \in \{1,..,n\} : z_j = 1 \text{ and } s_j = 1\}|$$

and t and m are the Hamming weights of $\mathbf{s}_1^n$ and $\mathbf{z}_1^n$, respectively. Clearly, given a state vector $\mathbf{s}_1^n$ and among all words of Hamming weight m the error pattern with the smallest $n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)$ is the least likely and the pattern with the largest $n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)$ is the most likely. It is easy to show that

$$\max\{t + m - n, 0\} \le n_1(\mathbf{z}_1^n, \mathbf{s}_1^n) \le \min\{m, t\}.$$

Let $\mathbf{z}_1^n$ and $\bar{\mathbf{z}}_1^n$ be two error words such that:

- $w_H(\mathbf{z}_1^n) = m$ where $0 \le m \le n - 1$
- $w_H(\bar{\mathbf{z}}_1^n) = m + i$ where $1 \le i \le n - m$.

We can write the ratio of error patterns $\bar{\mathbf{z}}_1^n$ and $\mathbf{z}_1^n$ given $\mathbf{s}_1^n$ as follows

$$\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)} = \left[\frac{P_B(1 - P_G)}{P_G(1 - P_B)}\right]^{n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) - n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)} \left[\frac{P_G}{1 - P_G}\right]^i$$

We treat several cases separately and we present, in each case, necessary and sufficient conditions under which the ratio

$$\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}$$

is less than 1. These cases depend on the Hamming weights of $\mathbf{z}_1^n$, $\bar{\mathbf{z}}_1^n$ and $\mathbf{s}_1^n$ and hence can only give some partial insight on the relation between SMD and ML decoding.

Case 1: $0 \leq t < \min\{n-m, m+i\}$:

In this case, $\max\{0, t+m-n\} = 0$ which means that $n_1(\mathbf{z}_1^n, \mathbf{s}_1^n) \geq 0$ and $\min\{t, m+i\} = t$ which means that $n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) \leq t$. Hence,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)} &= \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) - n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)} \left[\frac{P_G}{1-P_G}\right]^i \\
&\leq \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^t \left[\frac{P_G}{1-P_G}\right]^i \\
&\leq \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^t \left[\frac{P_G}{1-P_G}\right] \\
&< 1 \text{ iff } t < \frac{\log\left(\dfrac{1-P_G}{P_G}\right)}{\log\left[\dfrac{P_B(1-P_G)}{P_G(1-P_B)}\right]}.
\end{aligned}$$

Case 2: $\min\{n-m, m+i\} \leq t \leq n$:

We consider three subcases:

Subcase 1: $n-m \leq t \leq m+i$: In this case, $\max\{0, t+m-n\} = t+m-n$ and $\min\{t, m+i\} = t$ which means that $n_1(\mathbf{z}_1^n, \mathbf{s}_1^n) \geq t+m-n$ and $n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) \leq t$. Hence,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)} &= \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) - n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)} \left[\frac{P_G}{1-P_G}\right]^i \\
&\leq \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n-m} \left[\frac{P_G}{1-P_G}\right]^i \\
&\leq \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n-m} \left[\frac{P_G}{1-P_G}\right] \\
&< 1 \text{ iff } m > \frac{n \log\left[\dfrac{P_B(1-P_G)}{P_G(1-P_B)}\right] - \log\left(\dfrac{1-P_G}{P_G}\right)}{\log\left[\dfrac{P_B(1-P_G)}{P_G(1-P_B)}\right]}.
\end{aligned}$$

<u>**Subcase 2:** $m+i \le t \le n-m:$</u> In this case, $\max\{0, t+m-n\} = 0$ and $\min\{t, m+i\} = m+i$ which means that $n_1(\mathbf{z}_1^n, \mathbf{s}_1^n) \ge 0$ and $n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) \le m+i$. Hence,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)} &= \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) - n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)} \left[\frac{P_G}{1-P_G}\right]^i \\
&\le \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{m+i} \left[\frac{P_G}{1-P_G}\right]^i \\
&= \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{m} \left[\frac{P_B}{1-P_B}\right]^i \\
&\le \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{m} \left[\frac{P_B}{1-P_B}\right] \\
&< 1 \text{ iff } m < \frac{\log\left(\dfrac{1-P_B}{P_B}\right)}{\log\left[\dfrac{P_B(1-P_G)}{P_G(1-P_B)}\right]}.
\end{aligned}$$

<u>**Subcase 3:** $t \ge m+i$ **and** $t \ge n-m:$</u> In this case, $\max\{0, t+m-n\} = t+m-n$ and $\min\{t, m+i\} = m+i$ which means that $n_1(\mathbf{z}_1^n, \mathbf{s}_1^n) \ge t+m-n$ and $n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) \le m+i$. Hence,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n | \mathbf{S}_1^n = \mathbf{s}_1^n)} &= \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n_1(\bar{\mathbf{z}}_1^n, \mathbf{s}_1^n) - n_1(\mathbf{z}_1^n, \mathbf{s}_1^n)} \left[\frac{P_G}{1-P_G}\right]^i \\
&\le \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n+i-t} \left[\frac{P_G}{1-P_G}\right]^i \\
&= \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n-t} \left[\frac{P_B}{1-P_B}\right]^i \\
&\le \left[\frac{P_B(1-P_G)}{P_G(1-P_B)}\right]^{n-t} \left[\frac{P_B}{1-P_B}\right] \\
&< 1 \text{ iff } t > \frac{n\log\left[\dfrac{P_B(1-P_G)}{P_G(1-P_B)}\right] - \log\left(\dfrac{1-P_B}{P_B}\right)}{\log\left[\dfrac{P_B(1-P_G)}{P_G(1-P_B)}\right]}.
\end{aligned}$$

Chapter 5

Optimality of MD decoding over the BFMNC

In this chapter, we determine sufficient conditions on binary codes under which strict minimum Hamming distance decoding is equivalent to ML decoding over the BMNC. As we mentioned earlier, we will only treat the case where the correlation coefficient of this channel, ϵ, is non-. This channel is a special case of the QBC and the FMCC obtained by setting $M = 1$. It is also a special case of the Gilbert-Elliott channel (realized when the error probability is set to zero in the "good" state and to one in the "bad state"). We start by deriving the noise block distribution over this channel and then we present necessary and sufficient conditions on the error words under which the output of the strict minimum distance (SMD) decoder is identical to that of an ML decoder. From this result, we can determine tight sufficient conditions on binary codes under which the two decoders are equivalent.

5.1 Block Transition Probability for the BFMNC

We have shown in Section 2.2.1 that the transition matrix associated with this channel can be written in terms of the BER p and the correlation coefficient ϵ as follows:

$$\mathbf{P} \;=\; \begin{bmatrix} \epsilon + (1-\epsilon)(1-p) & (1-\epsilon)p \\ (1-\epsilon)(1-p) & \epsilon + (1-\epsilon)p \end{bmatrix}. \tag{5.1}$$

Let $\mathbf{z}_1^n$ be an error pattern generated by this channel, its probability is given by:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \;=\; P(Z_1 = z_1)\prod_{i=2}^{n} P(Z_i = z_i | Z_{i-1} = z_{i-1}). \tag{5.2}$$

We define $t_{ij}(\mathbf{z}_1^n)$ as follows:

$$t_{ij}(\mathbf{z}_1^n) \;:=\; \sum_{k=2}^{n} \delta[z_{k-1}, i]\delta[z_k, j], \tag{5.3}$$

where $i, j \in \mathbf{F}_2$ and where

$$\delta[a,b] \;=\; \begin{cases} 1, & \text{if a=b} \\ 0, & \text{otherwise.} \end{cases}$$

In other words, $t_{ij}(\mathbf{z}_1^n)$ counts the number of times that the pattern ij occurs in $\mathbf{z}_1^n$. When there is no confusion, we will write t_{ij} to denote $t_{ij}(\mathbf{z}_1^n)$. We can now re-write the probability distribution given in (5.2) as follows:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \;=\; P(Z_1 = z_1)\prod_{i=2}^{n} P(Z_i = z_i | Z_{i-1} = z_{i-1})$$

$$
\begin{aligned}
&= p^{z_1}(1-p)^{1-z_1}\left[\epsilon+(1-\epsilon)(1-p)\right]^{t_{00}}\left[(1-\epsilon)p\right]^{t_{01}} \\
&\quad \left[(1-\epsilon)(1-p)\right]^{t_{10}}\left[\epsilon+(1-\epsilon)p\right]^{t_{11}}.
\end{aligned}
$$

From the definition of t_{ij}, we can obtain the following identities:

$$t_{10} = n - w_H(\mathbf{z}_1^n) - t_{00} - (1-z_1) \tag{5.4}$$

$$t_{01} = w_H(\mathbf{z}_1^n) - t_{11} - z_1, \tag{5.5}$$

where $w_H(\cdot)$ denotes the Hamming weight. Hence,

$$
\begin{aligned}
P(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= p^{z_1}(1-p)^{1-z_1}\left[\epsilon+(1-\epsilon)(1-p)\right]^{t_{00}}\left[(1-\epsilon)p\right]^{w_H(\mathbf{z}_1^n)-t_{11}-z_1} \\
&\quad \left[(1-\epsilon)(1-p)\right]^{n-w_H(\mathbf{z}_1^n)-t_{00}-(1-z_1)}\left[\epsilon+(1-\epsilon)p\right]^{t_{11}} \\
&= (1-\epsilon)^{n-1}(1-p)^n\left(\frac{p}{1-p}\right)^{w_H(\mathbf{z}_1^n)}\left[\frac{\epsilon+(1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{t_{00}} \\
&\quad \left[\frac{\epsilon+(1-\epsilon)p}{(1-\epsilon)p}\right]^{t_{11}}.
\end{aligned}
$$

5.2 On the Equivalence of SMD and ML decoding over the BFMNC

Remark 5.1. *Let $\mathbf{z}_1^n \in \mathbf{F}_2^n$ be an error pattern with Hamming weight $0 < w_H(\mathbf{z}_1^n) < n$, then*

- *$t_{00} \leq n - w_H(\mathbf{z}_1^n) - 1$ with equality iff all zeros in $\mathbf{z}_1^n$ are consecutive.*
- *$t_{11} \leq w_H - 1$ with equality iff all ones in $\mathbf{z}_1^n$ are consecutive.*

Hence, t_{00} and t_{11} can be maximized simultaneously iff all zeros and all ones in $\mathbf{z}_1^n$ are consecutive.

Remark 5.2. *Let* $\mathbf{z}_1^n \in \mathbf{F}_2^n$ *be an error pattern with Hamming weight* $0 < w_H(\mathbf{z}_1^n) < \frac{n}{2}$*, then* $t_{00} \geq n - 2w_H(\mathbf{z}_1^n) - 1$ *with equality iff* $z_1 = 0$ *and* $t_{10} = w_H(\mathbf{z}_1^n)$*.*

Proof. Immediate from (5.4). □

Lemma 5.1. *The error pattern* $\mathbf{z}_1^n$ *with a Hamming weight* $0 < m < n$ *where all zeros and ones are consecutive (e.g.,* $\mathbf{z}_1^n = 00...011...1$*) is the most likely among all other error patterns generated by the BFMNC of the same length and weight.*

Proof. Let $\mathbf{z}_1^n$ be the error pattern with a Hamming weight $0 < m < n$ having all zeros and ones consecutive, and let $\tilde{\mathbf{z}}_1^n$ be an error pattern of the same Hamming weight m. We know that

$$t_{00}(\tilde{\mathbf{z}}_1^n) \;\; \leq \;\; n - m - 1 = t_{00}(\mathbf{z}_1^n)$$

and that

$$t_{11}(\tilde{\mathbf{z}}_1^n) \;\; \leq \;\; m - 1 = t_{11}(\mathbf{z}_1^n).$$

Then,

$$\begin{aligned} \frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\tilde{\mathbf{Z}}_1^n = \mathbf{z}_1^n)} &= \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{t_{00}(\mathbf{z}_1^n) - t_{00}(\tilde{\mathbf{z}}_1^n)} \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{t_{11}(\mathbf{z}_1^n) - t_{11}(\tilde{\mathbf{z}}_1^n)} \\ &= \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{n-m-1-t_{00}(\tilde{\mathbf{z}}_1^n)} \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{m-1-t_{11}(\tilde{\mathbf{z}}_1^n)} \\ &\geq 1, \end{aligned}$$

which concludes the proof. The probability of the most likely error pattern of

Hamming weight m with $0 < m < n$ is given by:

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= (1-\epsilon)^{n-1}(1-p)^n \left(\frac{p}{1-p}\right)^m \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{n-m-1} \\ &\quad \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{m-1}. \end{aligned} \tag{5.6}$$

□

Lemma 5.2. *Let $\mathbf{z}_1^n$ be an error pattern with a Hamming weight $0 < m < \frac{n}{2}$ and with $z_1 = 0$ and let all ones in $\mathbf{z}_1^n$ be followed each by a zero (e.g., $\mathbf{z}_1^n = 01010...0$). Then $\mathbf{z}_1^n$ is the least likely error pattern among all error patterns with the same length and Hamming weight.*

Proof. Let $\tilde{\mathbf{z}}_1^n$ be an error word with a Hamming weight $0 < m < \frac{n}{2}$. From Remark 5.2, we know that

$$t_{00}(\tilde{\mathbf{z}}_1^n) \geq n - 2m - 1 = t_{00}(\mathbf{z}_1^n).$$

Therefore,

$$\begin{aligned} \frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\tilde{\mathbf{Z}}_1^n = \mathbf{z}_1^n)} &= \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{t_{00}(\mathbf{z}_1^n) - t_{00}(\tilde{\mathbf{z}}_1^n)} \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{t_{11}(\mathbf{z}_1^n) - t_{11}(\tilde{\mathbf{z}}_1^n)} \\ &= \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{n-2m-1-t_{00}(\tilde{\mathbf{z}}_1^n)} \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{-t_{11}(\tilde{\mathbf{z}}_1^n)} \\ &\leq \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{-t_{11}(\tilde{\mathbf{z}}_1^n)} \\ &\leq 1, \end{aligned}$$

which concludes the proof. The probability of this error pattern is given by:

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \;=\; (1-\epsilon)^{n-1}(1-p)^n\left(\frac{p}{1-p}\right)^m\left[\frac{\epsilon+(1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{n-2m-1} \tag{5.7}$$

□

Theorem 5.1. *Define:*

$$m_1(\epsilon,p) := \frac{\ln\left(\dfrac{\epsilon+(1-\epsilon)(1-p)}{(1-\epsilon)p}\right)}{\ln\left(\dfrac{\epsilon+(1-\epsilon)p}{(1-\epsilon)p}\right)+\ln\left(\dfrac{\epsilon+(1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right)}$$

and

$$m_2(n,\epsilon,p) := \frac{(n-1)\ln\left(\dfrac{\epsilon+(1-\epsilon)(1-p)}{\epsilon+(1-\epsilon)p}\right)+\ln\left(\dfrac{1-p}{p}\right)}{2\ln\left(\dfrac{\epsilon+(1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right)+\ln\left(\dfrac{1-p}{p}\right)}.$$

For ***any*** *two BFMNC error words $\mathbf{z}_1^n$ and $\bar{\mathbf{z}}_1^n$ satisfying*

i. $w_H(\mathbf{z}_1^n) = m$, *where* $0 \le m < \frac{n}{2}$

ii. $w_H(\bar{\mathbf{z}}_1^n) = m + i$, *where* $1 \le i \le n - m$

we have that

$$m < m^*(n,\epsilon,p) := \min\{m_1(\epsilon,p), m_2(n,\epsilon,p)\} \quad\Longleftrightarrow\quad \frac{P(\mathbf{Z}_1^n = \mathbf{z}^n)}{P(\bar{\mathbf{Z}}_1^n = \mathbf{z}^n)} > 1.$$

Remark 5.3. *This theorem is an improvement on [2, Lemma 3] since the provided conditions are necessary and sufficient, while the conditions of [2, Lemma 3] are only sufficient.*

Proof. We first prove the first direction ($\Longrightarrow$):

Consider the following three cases:

Case 1: $m = 0$

In this case, $\mathbf{z}_1^n$ is the all-zero error pattern. From Lemma 2.2, $P(\mathbf{Z}_1^n = 0^n) > P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)$.

Case 2: $0 < m < \frac{n}{2}$ and $0 < i < n - m$

The error pattern $\bar{\mathbf{z}}_1^n$ is not the all one error word. Hence,

$$\begin{align}
\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} &\geq \frac{\min_{\mathbf{z}_1^n \in \mathbf{F}_2^n : w_H(\mathbf{z}_1^n) = m} P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{\max_{\bar{\mathbf{z}}_1^n \in \mathbf{F}_2^n : w_H(\bar{\mathbf{z}}_1^n) = m+i} P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \nonumber \\
&= \frac{P(\mathbf{Z}_1^n = 01010...00)}{P(\mathbf{Z}_1^n = 00000...11)} \tag{5.8} \\
&= \left(\frac{1-p}{p}\right)^i \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{(n-2m-1)-(n-m-i-1)} \nonumber \\
&\quad \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{-(m+i-1)} \tag{5.9} \\
&= \left(\frac{1-p}{p}\right)^i \left[\frac{(1-\epsilon)(1-p)}{\epsilon + (1-\epsilon)(1-p)}\right]^{m-i} \left[\frac{(1-\epsilon)p}{\epsilon + (1-\epsilon)p}\right]^{m+i-1} \nonumber \\
&= \left[\frac{\epsilon + (1-\epsilon)(1-p)}{\epsilon + (1-\epsilon)p}\right]^i \left[\frac{(1-\epsilon)(1-p)}{\epsilon + (1-\epsilon)(1-p)}\right]^m \nonumber \\
&\quad \left[\frac{(1-\epsilon)p}{\epsilon + (1-\epsilon)p}\right]^{m-1} \nonumber \\
&\geq \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)p}\right] \nonumber \\
&\quad \left\{\frac{(1-\epsilon)^2(1-p)p}{[\epsilon + (1-\epsilon)(1-p)][\epsilon + (1-\epsilon)p]}\right\}^m \tag{5.10} \\
&> 1. \nonumber
\end{align}$$

Here, (5.8) follows from Lemmas 5.1 and 5.2 and (5.9) is obtained by replacing the numerator by (5.7) and the denominator by (5.6). We have an inequality in (5.10) since we set $i = 1$ (indeed, the term raised to the power i is greater than 1 and hence is increasing in i where $i \geq 1$). The last strict inequality is a result of the condition $m < m_1(\epsilon, p)$.

Case 3: $0 < m < \frac{n}{2}$ and $i = n - m$

In this case, $\bar{\mathbf{z}}_1^n$ is the all-one codeword, and hence $t_{00}(\bar{\mathbf{z}}_1^n) = 0$ and $t_{11}(\bar{\mathbf{z}}_1^n) = n - 1$. Hence,

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} &= \left(\frac{1-p}{p}\right)^{n-m} \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{t_{00}(\mathbf{z}_1^n)} \\
&\quad \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{t_{11}(\mathbf{z}_1^n)-n+1} \\
&\geq \left(\frac{1-p}{p}\right)^{n-m} \left[\frac{\epsilon + (1-\epsilon)(1-p)}{(1-\epsilon)(1-p)}\right]^{n-2m-1} \\
&\quad \left[\frac{\epsilon + (1-\epsilon)p}{(1-\epsilon)p}\right]^{-n+1} \\
&> 1,
\end{aligned}$$

where the last strict inequality is a result of the condition $m < m_2(n, \epsilon, p)$.

We now prove the other direction ($\Longleftarrow$):

- Assume $m \geq m_1$: In the proof of Case 2, all the inequalities except the last one can be met with equality by choosing the error patterns as follows: $\mathbf{z}_1^n = 01010...00$ and $\bar{\mathbf{z}}_1^n = 00000...11$, and by letting $w_H(\bar{\mathbf{z}}_1^n) = m + 1$. As a result of the assumption that $m \geq m_1(\epsilon, p)$, we get

$$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \leq 1.$$

Therefore, we proved that there exist at least two words $\mathbf{z}^n$ and $\bar{\mathbf{z}}^n$ satisfying:

i. $w(\mathbf{z}^n) = m$, where $0 < m \leq \frac{n}{2}$

ii. $w(\bar{\mathbf{z}}^n) = m + i$, where $1 \leq i \leq n - m$

such that:

$$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}^n)} \leq 1.$$

- Assume $m \geq m_2(n, \epsilon, p)$: The proof follows a similar reasoning as the case where $m \geq m_1(\epsilon, p)$, only this time we choose $\bar{\mathbf{z}}_1^n = \mathbf{1}^n$ (while $\mathbf{z}_1^n$ is unchanged).

□

Theorem 5.2. *Let $\mathcal{C}$ be any (n, K, d) code used over a BFMNC with parameters ϵ and p. Denote by r_{cov} the classical covering radius of this code. If $r_{cov} < \min\left\{m^*(n, \epsilon, p), \frac{n}{2}\right\}$, then the output of the SMD decoder (when it does not declare a decoding failure) is identical to the output of the ML decoder for this code.*

Proof. Let $\mathbf{y}$ be the received word. Let $m := \min_{\mathbf{c} \in \mathcal{C}} \mathrm{d}_H(\mathbf{y}, \mathbf{c})$,where $d_H(\cdot, \cdot)$ denotes the Hamming distance. Clearly, $m \leq r_{cov} < \frac{n}{2}$ (from the definition of the covering radius). If there exists a unique codeword $\hat{\mathbf{c}}$ such that $\mathrm{d}_H(\mathbf{y}, \hat{\mathbf{c}}) = m$, then SMD decoding gives a valid codeword. Since $m < m^*(n, \epsilon, p)$, it follows from Theorem 5.1 that all other error words of larger Hamming weights have a smaller probability than the error word corresponding to the SMD decision. Hence the ML decoder will give the same output. □

Remark 5.4. *Theorem 5.2 improves on [2, Lemma 4] since it applies to any code, whereas [2, Lemma 4] only applies to linear perfect codes.*

Corollary 5.1. *Let $\mathcal{C}$ be an (n, M, d_{min}) perfect code (in the classical sense) used over a BFMNC with parameters ϵ and p. If*

$$r_{cov} = \left\lfloor \frac{d_{min} - 1}{2} \right\rfloor \ < \ \min\left\{m^*(n, \epsilon, p), \frac{n}{2}\right\},$$

then $\mathcal{C}$ is a generalized perfect code for this channel and hence is optimal (under ML decoding) among all codes of the same length and dimension sent over the same channel.

Proof. Immediate from Lemma 3.3 and Theorems 5.1, 5.2. □

Remark 5.5. *Note that we cannot make a statement similar to the one in Corollary 5.1 for quasi-perfect codes, since two error words of the same weight are not guaranteed to have the same probability.*

5.3 Numerical Results

We illustrate the condition of Theorem 5.2 by simulating the performance of different codes under both SMD and ML decoding over the BFMNC with parameters ϵ and p. Since Theorem 5.2 does not treat the case when SMD declares a failure (i.e., when ties occur), we disregard this case in our simulations as well by only using the ML decoder when the SMD decoder does not declare a failure.

The first code we simulate is the $[7, 4, 3]$ perfect Hamming code. It has a covering radius $r_{cov} = 1$. We show the results in Table 5.1.

According to Theorem 5.2, if $m^*(n, \epsilon, p) > 1$ the output of the SMD decoder (when it does not declare a decoding failure) is identical to the output of the ML decoder for this code. Indeed, as Table 5.1 shows, the probabilities of codeword

ϵ	0.2	0.3	0.4	0.5	0.6	0.7
$m^*(n,\epsilon,p)$	1.6305	1.2591	1.0691	0.9362	0.8467	0.7775
PCE under SMD	0.17338	0.18111	0.18366	0.18602	0.18113	0.171275
PCE under ML	0.17338	0.18111	0.18366	0.182835	0.170805	0.153265

Table 5.1: Verifying Theorem 5.2 for the $[7,4,3]$ Hamming code over the BFMNC with parameters $p = 0.1$ and ϵ.

error for SMD and ML decoding (when SMD does not declare a decoding failure) over the BFMNC are identical. We start noticing some discrepancy when $m^*(n,\epsilon,p) < 1$. Similarly, we simulate the performance of the $(15, 2^{11}, 3)$ Vasil'ev nonlinear perfect code constructed using the method described in Theorem 3.1 using

$$f(\mathbf{x}_1^7) = x_1x_2x_7 \oplus x_2x_4x_6 \oplus x_1x_3x_5.$$

This code has also a covering radius $r_{cov} = 1$. We show the results in Table 5.2. The results are similar to those observed in Table 5.1 for the $[7,4,3]$ Hamming code; i.e., we notice that the probabilities of codeword error for SMD and ML decoding (when SMD does not declare a decoding failure) over the BFMNC are identical when $m^*(n,\epsilon,p) > 1$ and they start to differ when $m^*(n,\epsilon,p) < 1$. Similarly, we simulate the performance of the $[8,4,4]$ Reed-Muller code ($r_{cov} = 2$) and the $[24,12,8]$ extended Golay code ($r_{cov} = 3$). The results are shown in Tables 5.3 and 5.4, respectively.

ϵ	0.2	0.3	0.4	0.5	0.6	0.7
$m^*(n,\epsilon,p)$	1.6305	1.2591	1.0691	0.9362	0.8467	0.7775
PCE under SMD	0.4206	0.415	0.3872	0.3718	0.3486	0.3083
PCE under ML	0.4206	0.415	0.3872	0.3669	0.3341	0.2878

Table 5.2: Verifying Theorem 5.2 for the $(15, 2^{11}, 3)$ Vasil'ev nonlinear perfect code over the BFMNC with parameters $p = 0.1$ and ϵ.

ϵ	0.08	0.1	0.12	0.6	0.7	0.8
$m^*(n,\epsilon,p)$	3.1889	2.6790	2.3353	0.8467	0.7775	0.6167
PCE under SMD	0.0523871	0.0546158	0.0578775	0.113368	0.11782	0.116573
PCE under ML	0.0523871	0.0546158	0.0578775	0.108908	0.103598	0.0992633

Table 5.3: Verifying Theorem 5.2 for the $[8,4,4]$ Reed Muller code $\mathcal{RM}(1,3)$ over the BFMNC with parameters $p=0.1$ and ϵ.

ϵ	0.03	0.04	0.05	0.2	0.3	0.4
$m^*(n,\epsilon,p)$	7.3567	5.5679	4.6990	1.6305	1.2591	1.0691
PCE under SMD	0.0853387	0.0889826	0.0861075	0.101059	0.11657	0.130687
PCE under ML	0.0853387	0.0889826	0.0861075	0.100353	0.110665	0.102962

Table 5.4: Verifying Theorem 5.2 for the $[24,12,8]$ extended Golay code over the BFMNC with parameters $p=0.1$ and ϵ.

5.4 A Closer Look at Theorems 5.1 and 5.2

We illustrate the condition of Theorem 5.1 in Figs 5.1-5.3 by plotting $m^*(m,\epsilon,p)$ versus the BFMNC correlation coefficient ϵ for different values of the BER p and the block length n.

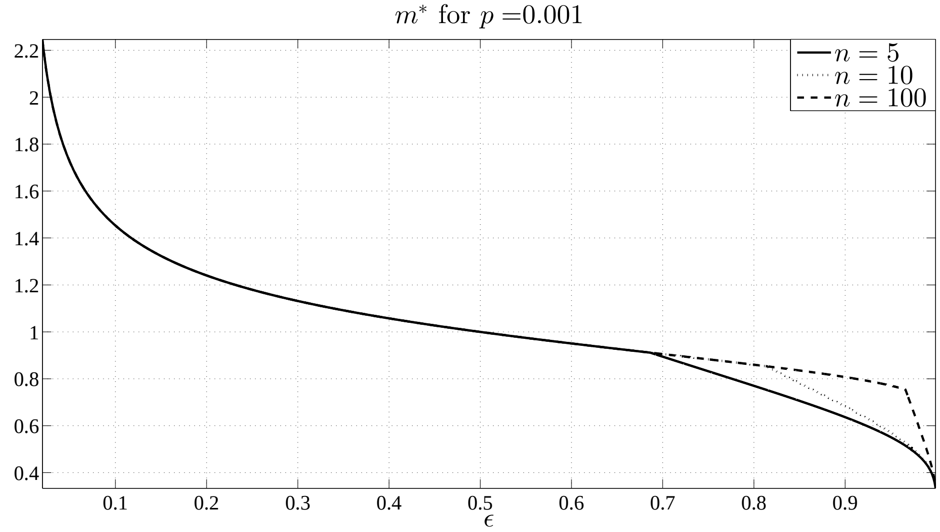


Figure 5.1: Plot of $m^*(n,\epsilon,p)$ with respect to ϵ for different values of n and for $p=0.001$.

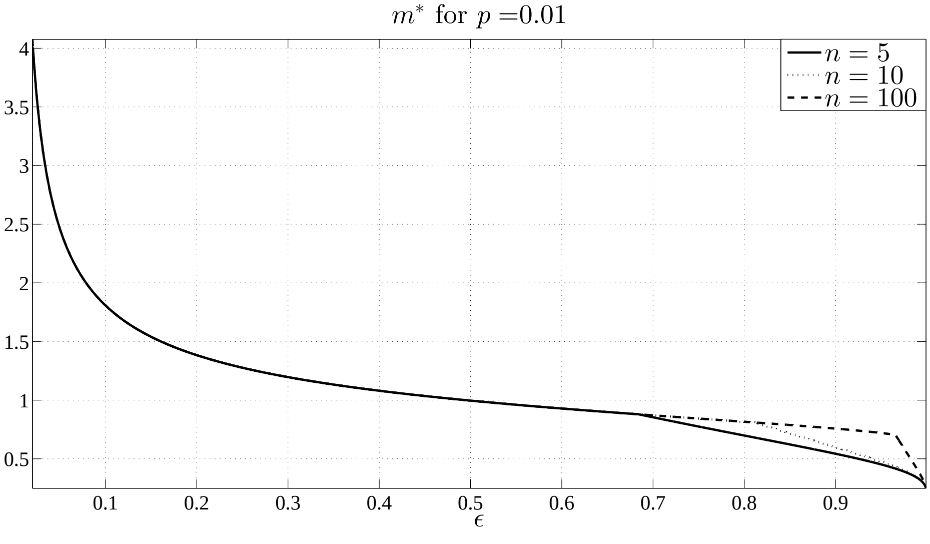


Figure 5.2: Plot of $m^*(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.01$.

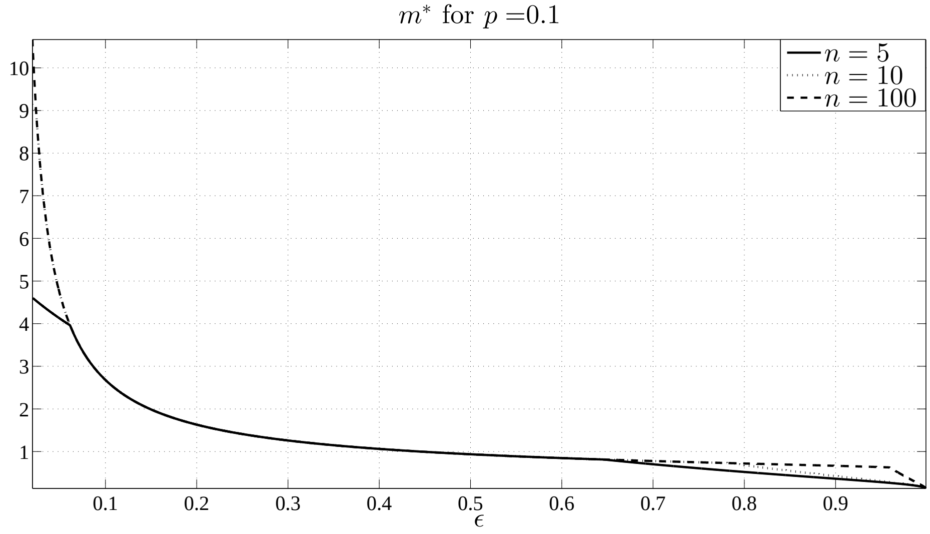


Figure 5.3: Plot of $m^*(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.1$.

We notice from the above figures that the condition of Theorem 5.1 is quite restrictive for channels with $\epsilon > 0.1$. In fact, for these channels, only codes with a covering radius $r_{cov} = 1$ satisfy the condition (e.g., the family of Hamming codes). For smaller ϵ, more codes satisfy the condition, and when $\epsilon = 0$ (i.e., when the BFMNC reduces to the BSC), unsurprisingly all block codes satisfy it.

Corollary 5.2. *Every binary (classical) perfect code of minimum Hamming distance 3 is optimal among all other codes of the same lengths and dimensions over any BFMNC with* $\epsilon < \frac{1-2p}{2(1-p)}$.

Proof. The condition

$$\epsilon < \frac{1-2p}{2(1-p)} \quad \Longleftrightarrow \quad m_1(\epsilon, p) > 1$$

and it also implies that $\forall n \geq 3$, $m_2(n, \epsilon, p) > 1$. Hence,

$$\epsilon < \frac{1-2p}{2(1-p)} \quad \Longrightarrow \quad m^*(n, \epsilon, p) > 1 \qquad (\forall n \geq 3).$$

Hence, according to Corollary 5.1, any perfect code (linear or non-linear) of minimum Hamming distance 3 ($r_{cov} = 1$) is optimal. □

Corollary 5.2 is also proven in [14, Theorem 1]. Next, we generalize the above result to any (classical) perfect code.

Lemma 5.3. *$m^*(n, \epsilon, p)$ is strictly decreasing in ϵ.*

Proof. The proof is basic but long, and hence will be skipped for conciseness. □

We now define $\epsilon_i(p, n)$ $\forall p \in (0, \frac{1}{2})$ and $n \in \mathbb{N}^*$ as follows:

$$\epsilon_i(p, n) \quad := \quad \sup\left\{\epsilon \in (0, 1) : m^*(n, \epsilon, p)) > \left\lfloor \frac{i-1}{2} \right\rfloor\right\} \qquad i \in \mathbb{N}^* : i \leq n.$$

We know from Lemma 5.4 that $\epsilon_i(p,n)$ exists for any such i, n and p.

Theorem 5.3. *For arbitrary,* $p \in (0,1)$ *,* $\forall i, n \in \mathbb{N}^*$ *such that* $i < n$*, and* $\forall \epsilon < \epsilon_i(p,n)$*, all (classical) perfect length* n *codes* $\mathcal{C}$ *with minimum distance* $d_{min} \leq i$ *are optimal on the BFMNC with parameters* ϵ *and* p*.*

Proof. Perfect codes $\mathcal{C}$ satisfy:

$$\begin{aligned} r_{cov} &= \left\lfloor \frac{d_{min}-1}{2} \right\rfloor \\ &\leq \left\lfloor \frac{i-1}{2} \right\rfloor \\ &< m^*(n,\epsilon,p)), \end{aligned}$$

where the last inequality follows from the definition of $\epsilon_i(p,n)$. From Corollary 5.1, $\mathcal{C}$ is optimal on the BFMNC with parameters ϵ and p; this concludes the proof. □

Chapter 6

Optimality of MD Decoding over the QBC

In this chapter, we present results similar to those presented in Chapter 5 for n-block codes transmitted over the QBC. We have shown in Chapter 2 that the block transition probability over this channel admits two different expressions depending on whether $M \geq n$ or $M < n$. Consequently, those two cases will be treated separately. For the case when $M < n$, we restrict our study to the QBC with $M = 2$. We assume throughout this chapter that $\alpha \leq 1$. This assumption, while restrictive, is practical since it only does not allow the oldest ball in the queue to contribute more to the current draw than the more recent balls in the queue.

We will first briefly present the results for the case when $M > n$. The treatment for the QBC with $M = 2$ will be organized similarly to the treatment of the BFMNC. We will start by deriving the noise block distribution over this channel in order to eventually obtain sufficient conditions on binary codes under which the strict MD decoder and the ML decoder produce the same output.

6.1 On the equivalence of MD and ML Decoding of n-block codes over the QBC with M $\geq$ n

Let $\mathbf{z}_1^n$ be an error word generated by this channel. Its probability is given by (2.7):

$$P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \quad = \quad \frac{\prod_{j=0}^{n-d_1^n-1}\left[(1-p)+j\dfrac{\epsilon}{1-\epsilon}\right]\prod_{j=0}^{d_1^n-1}\left(p+j\dfrac{\epsilon}{1-\epsilon}\right)}{\prod_{j=0}^{n-1}\left(1+j\dfrac{\epsilon}{1-\epsilon}\right)}$$

where d is the Hamming weight of $\mathbf{z}_1^n$. We clearly notice that the probability of error patterns depends only on their Hamming weight. In other words, two error patterns have the same probability if their Hamming weight is the same. We denote by $\mathbf{z}_1^n(t)$ an error word with Hamming weight t. For $0 \leq m \leq n-m$ and $1 \leq i \leq n-m$, we define $p(m,i)$ as follows:

$$\begin{aligned}
\rho(m,i) \quad &:= \quad \frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n(m+i))}{P(\mathbf{Z}_1^n = \mathbf{z}_1^n(m))} \\
&= \quad \frac{\prod_{j=m}^{m+i-1}\left[p+j\dfrac{\epsilon}{1-\epsilon}\right]}{\prod_{j=n-m-i}^{n-m-1}\left[(1-p)+j\dfrac{\epsilon}{1-\epsilon}\right]} \\
&= \quad \prod_{j=0}^{i-1}\frac{p+(m+j)\dfrac{\epsilon}{1-\epsilon}}{(1-p)+(n-m-i+j)\dfrac{\epsilon}{1-\epsilon}}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\rho(m,i) \leq 1 \quad &\Longleftrightarrow \quad (1-p) + (n-m-i)\frac{\epsilon}{1-\epsilon} \geq p + m\frac{\epsilon}{1-\epsilon} \\
&\Longleftrightarrow \quad m + i - \mu \leq \mu - m \\
&\Longleftrightarrow \quad |m + i - \mu| \leq |m - \mu|,
\end{aligned}$$

where

$$\mu \;=\; \frac{(1-2p)(1-\epsilon)}{2\epsilon} + \frac{n}{2}.$$

It is clear that this case is very similar to what we saw in Chapter 4 for the IMCC. To avoid redundancy, we will only list the theorems that apply to this without proof.

Theorem 6.1. *ML decoding over the QBC with $M \geq n$ is equivalent to either minimum Hamming distance decoding or maximum Hamming distance decoding. The output of the ML decoder is given by:*

$$\hat{\mathbf{e}}_{ML} \;=\; arg \max_{\mathbf{e} \in \{\mathbf{e}_{min}, \mathbf{e}_{max}\}} |w_H(\mathbf{e}) - \mu|,$$

where $\mathbf{e}_{min}$ and $\mathbf{e}_{max}$ are the outputs of the minimum and the maximum Hamming distance decoders, respectively.

Theorem 6.2. *For any (n, M, d_{min}) code $\mathcal{C}$ used over the QBC with $M \geq n$, if the (classical) covering radius of this code satisfies*

$$r_{cov} \leq \frac{(1-2p)(1-\epsilon)}{\epsilon},$$

then the outputs of the MD and ML decoders are the same.

Remark 6.1. *A plot of this condition is already given in Chapter 4 (see Fig. 4.2).*

Theorem 6.3. *For any (n, M, d_{min}) code $\mathcal{C}$ used over the QBC with $M \geq n$, the outputs of the MD and ML decoders are the same iff*

$$d_{sum}(\mathcal{C}) \ \leq \ \frac{(1-2p)(1-\epsilon)}{\epsilon},$$

where $d_{sum}(\mathcal{C})$ is defined in Definition 4.1.

Corollary 6.1. *For a linear code containing the all-one codeword, if $p < 0.5$ and $\epsilon > 0$ then ML decoding over the QBC with $M > n$ reduces to minimum Hamming distance decoding.*

6.2 Block Transition Probability for the QBC with M = 2

Let $\mathbf{z}_1^n$ be an error word generated by the QBC with $M < n$, its probability is given by (2.11) as follows:

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n) \ &= \ \frac{L^{(M)}}{\left[1 + (M-1+\alpha)\frac{\epsilon}{1-\epsilon}\right]^{n-M}} \prod_{i=M+1}^{n} \left[\frac{(d_{i-M+1}^{i-1} + \alpha z_{i-M})\epsilon}{1-\epsilon} + p\right]^{z_i} \\ &\quad \left\{\frac{\left[M-1-d_{i-M+1}^{i-1} + \alpha(1-z_{i-M})\right]\epsilon}{1-\epsilon} + 1 - p\right\}^{1-z_i} \end{aligned}$$

where:

$$L^{(M)} = \frac{\prod_{j=0}^{M-d_1^M-1}\left[(1-p)+j\frac{\epsilon}{1-\epsilon}\right]\prod_{j=0}^{d_1^M-1}\left(p+j\frac{\epsilon}{1-\epsilon}\right)}{\prod_{j=0}^{M-1}\left(1+j\frac{\epsilon}{1-\epsilon}\right)}.$$

For M=2,

$$\begin{aligned} P(\mathbf{Z}_1^n = \mathbf{z}_1^n) &= \frac{\left[(1-p)p^{z_2}\left(1-p+\frac{\epsilon}{1-\epsilon}\right)^{1-z_2}\right]^{1-z_1}\left[p(1-p)^{1-z_2}\left(p+\frac{\epsilon}{1-\epsilon}\right)^{z_2}\right]^{z_1}}{1+\frac{\epsilon}{1-\epsilon}} \\ &\times \frac{\prod_{i=3}^{n}\left[\frac{(z_{i-1}+\alpha z_{i-2})\epsilon}{1-\epsilon}+p\right]^{z_i}\left\{\frac{[(1-z_{i-1})+\alpha(1-z_{i-2})]\,\epsilon}{1-\epsilon}+1-p\right\}^{1-z_i}}{\left[1+(1+\alpha)\frac{\epsilon}{1-\epsilon}\right]^{n-2}}. \end{aligned} \tag{6.1}$$

We define $t_{ijk}(\mathbf{z}_1^n)$ as follows:

$$t_{ijk}(\mathbf{z}_1^n) := \sum_{l=3}^{n} \delta[z_{l-2}, i]\delta[z_{l-1}, j]\delta[z_l, k],$$

where $i, j, k \in \mathbf{F}_2$ and where

$$\delta[a,b] = \begin{cases} 1, & \text{if } a = b \\ 0, & \text{otherwise.} \end{cases}$$

In other words, $t_{ijk}(\mathbf{z}_1^n)$ counts the number of times the pattern ijk occurs in $\mathbf{z}_1^n$. When there is no confusion, we will write t_{ijk} to denote $t_{ijk}(\mathbf{z}_1^n)$. We now re-write

(6.1) as follows:

$$\begin{aligned}
P(\mathbf{z}_1^n) &= \frac{(1-p)\left(\frac{\epsilon}{1-\epsilon}+1-p\right)}{\left[1+(1+\alpha)\frac{\epsilon}{1-\epsilon}\right]^{n-2}\left[1+\frac{\epsilon}{1-\epsilon}\right]}\left[\frac{\left(\frac{\epsilon}{1-\epsilon}+1-p\right)\left(\frac{\epsilon}{1-\epsilon}+p\right)}{p(1-p)}\right]^{z_1z_2} \\
&\quad \left[\frac{p}{\frac{\epsilon}{1-\epsilon}+1-p}\right]^{z_1+z_2}\left[(1+\alpha)\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{000}}p^{t_{001}}\left[\alpha\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{010}} \\
&\quad \left[\frac{\epsilon}{1-\epsilon}+p\right]^{t_{011}}\left[\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{100}}\left[\alpha\frac{\epsilon}{1-\epsilon}+p\right]^{t_{101}}(1-p)^{t_{110}} \\
&\quad \left[(1+\alpha)\frac{\epsilon}{1-\epsilon}+p)\right]^{t_{111}}. \qquad (6.2)
\end{aligned}$$

From the definition of t_{ijk}, we obtain the following identities:

$$t_{001}+t_{101} = t_{01}-(1-z_1)z_2 \qquad (6.3)$$

$$t_{011}+t_{111} = t_{11}-z_1z_2 \qquad (6.4)$$

$$t_{100}+t_{000} = t_{00}-(1-z_1)(1-z_2) \qquad (6.5)$$

$$t_{110}+t_{010} = t_{10}-z_1(1-z_2), \qquad (6.6)$$

where t_{ij} are defined in (5.3). Hence,

$$\begin{aligned}
P(\mathbf{Z}_1^n=\mathbf{z}_1^n) &= \frac{(1-p)\left(\frac{\epsilon}{1-\epsilon}+1-p\right)}{\left[1+(1+\alpha)\frac{\epsilon}{1-\epsilon}\right]^{n-2}\left[1+\frac{\epsilon}{1-\epsilon}\right]}\left[\frac{\left(\frac{\epsilon}{1-\epsilon}+1-p\right)\left(\frac{\epsilon}{1-\epsilon}+p\right)}{p(1-p)}\right]^{z_1z_2} \\
&\quad \left[\frac{p}{\frac{\epsilon}{1-\epsilon}+1-p}\right]^{z_1+z_2}\left[(1+\alpha)\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{000}}p^{t_{01}-t_{101}-z_2+z_1z_2}
\end{aligned}$$

$$
\begin{aligned}
& \left[\alpha\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{010}} \left[\frac{\epsilon}{1-\epsilon}+p\right]^{t_{11}-t_{111}-z_1z_2} \left[\alpha\frac{\epsilon}{1-\epsilon}+p\right]^{t_{101}} \\
& \left[\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{00}-t_{000}-1-z_1z_2+z_1+z_2} \left[(1+\alpha)\frac{\epsilon}{1-\epsilon}+p)\right]^{t_{111}} \\
& (1-p)^{t_{10}-z_1+z_1z_2-t_{010}} \\
= & \frac{(1-p)}{\left[1+(1+\alpha)\frac{\epsilon}{1-\epsilon}\right]^{n-2}\left[1+\frac{\epsilon}{1-\epsilon}\right]} \left[\frac{p}{1-p}\right]^{z_1} p^{t_{01}}(1-p)^{t_{10}} \\
& \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+1-p}{\frac{\epsilon}{1-\epsilon}+1-p}\right]^{t_{000}} \left[\frac{\alpha\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{t_{010}} \left[\frac{\alpha\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{t_{101}} \\
& \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+p)}{\frac{\epsilon}{1-\epsilon}+p}\right]^{t_{111}} \left[\frac{\epsilon}{1-\epsilon}+1-p\right]^{t_{00}} \left[\frac{\epsilon}{1-\epsilon}+p\right]^{t_{11}}. \qquad (6.7)
\end{aligned}
$$

Equation (6.7) can be further simplified by using the identities in (5.4) and (5.5). After some algebraic manipulations, we obtain this final expression of the probability:

$$
\begin{aligned}
P(\mathbf{Z}_1^n = \mathbf{z}_1^n) = & \frac{(1-p)^n}{\left[1+(1+\alpha)\frac{\epsilon}{1-\epsilon}\right]^{n-2}\left[1+\frac{\epsilon}{1-\epsilon}\right]} \left[\frac{p}{1-p}\right]^{m} \left[\frac{\alpha\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{t_{010}} \\
& \left[\frac{\alpha\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{t_{101}} \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+p)}{\frac{\epsilon}{1-\epsilon}+p}\right]^{t_{111}} \left[\frac{\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{t_{00}} \\
& \left[\frac{\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{t_{11}} \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+1-p}{\frac{\epsilon}{1-\epsilon}+1-p}\right]^{t_{000}}, \qquad (6.8)
\end{aligned}
$$

where m is the Hamming weight of $\mathbf{z}_1^n$.

6.3 On the Equivalence of SMD and ML Decoding over the QBC with M = 2

Remark 6.2. *Let $\mathbf{z}_1^n \in \mathbf{F}_2^n$ be an error word with Hamming weight $m = w_H(\mathbf{z}_1^n)$, then:*

$$t_{000} \leq n - m - 2 \quad \textit{if} \quad m < n - 1 \tag{6.9}$$

$$t_{111} \leq m - 2 \quad \textit{if} \quad m > 1 \tag{6.10}$$

$$t_{101} + t_{11} \leq m - 1 \quad \textit{if} \quad m > 0 \tag{6.11}$$

$$t_{010} + t_{00} \leq n - m - 1 \quad \textit{if} \quad m < n \tag{6.12}$$

Proof. The inequalities in (6.9) and (6.10) are trivial.

From the definition of t_{ijk} and t_{ij} :

$$\begin{aligned}
t_{101} + t_{11} &= \sum_{i=3}^{n} z_{i-2}(1 - z_{i-1})z_i + \sum_{i=2}^{n} z_{i-1}z_i \\
&= z_1 z_2 + \sum_{i=3}^{n} z_{i-2}(1 - z_{i-1})z_i + z_{i-1}z_i \\
&= z_1 z_2 + \sum_{i=3; z_i=1}^{n} z_{i-2}(1 - z_{i-1}) + z_{i-1} \\
&= z_1 z_2 + \sum_{i=3; z_i=1}^{n} s_i,
\end{aligned}$$

where $s_i = z_{i-2}(1 - z_{i-1}) + z_{i-1} \leq 1$. Hence, the sum can increase by at most one every time there is a one in $\mathbf{z}_1^n$.

- **Case 1:** $z_1 = 0, z_2 = 0$

  Let i_0 be the index of the first non-zero entry in $\mathbf{z}_3^n$. Then $s_{i_0} = 0$ since

$z_{i_0-1} = z_{i_0-2} = 0$. For the other $m-1$ ones in $\mathbf{z}_1^n$, s_i can be at most one. Hence, $t_{101} + t_{11} \leq m-1$.

- **Case 2:** $z_1 = 0, z_2 = 1$

  In that case, $z_1 z_2 = 0$ and there are $m-1$ summands in the sum. Hence, $t_{101} + t_{11} \leq m-1$.

- **Case 3:** $z_1 = 1, z_2 = 0$

  Similar to case 2.

- **Case 4:** $z_1 = 1, z_2 = 1$

  In that case, $z_1 z_2 = 1$, but there are $m-2$ summands in the sum. Hence, $t_{101} + t_{11} \leq 1 + m - 2 = m-1$.

Indeed, $t_{101} + t_{11} \leq m-1$. By symmetry, the inequality in (6.12) holds. □

Remark 6.3. *Let $\mathbf{z}_1^n \in \mathbf{F}_2^n$ be an error word with Hamming weight $m = w_H(\mathbf{z}_1^n) < \frac{n}{3}$, then*

$$t_{000} \geq n - 3m - 2. \tag{6.13}$$

Proof. The proof is immediate from (6.5) and Remark 5.2. The pattern

$$(001001....000),$$

for example, achieves the above inequality with equality. □

Lemma 6.1. *The error pattern $\mathbf{z}_1^n$ of Hamming weight $0 < m < n$ where all zeros and ones are consecutive (e.g., $\mathbf{z}_1^n = 00...011...1$) is the most likely among all other error patterns of the same length and weight generated by the QBC with $M = 2$.*

Proof. Let $\mathbf{z}_1^n$ be an error word with Hamming weight $0 < m < n$.

$$\max_{t_{101},t_{11}} \left[\frac{\frac{\alpha\epsilon}{1-\epsilon}+p}{p}\right]^{t_{101}} \left[\frac{\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{t_{11}} \leq \max_{(i,j):i+j=m-1} \left[\frac{\frac{\alpha\epsilon}{1-\epsilon}+p}{p}\right]^{i} \left[\frac{\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{j} \tag{6.14}$$

$$= \left[\frac{\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{m-1}, \tag{6.15}$$

where the inequality in (6.14) comes from (6.11), and (6.15) follows from the assumption that $\alpha \leq 1$. We have equality when all ones are consecutive.

Similarly, using (6.12) we can prove that:

$$\max_{t_{010},t_{00}} \left[\frac{\frac{\alpha\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{t_{010}} \left[\frac{\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{t_{00}} = \left[\frac{\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{n-m-1},$$

where the maximum is achieved when all zeros are consecutive.

When all zeros and all ones are consecutive, t_{000} and t_{111} are also maximized and hence the expression of the probability of $\mathbf{z}_1^n$ given by (6.8) is maximized. □

Lemma 6.2. *Consider the error words of length n having a Hamming weight $0 < m < \frac{n}{3}$. The pattern $\mathbf{b}_1^n = (001001....00)$ is the least likely among all patterns of the same length and weight generated by the QBC with $M = 2$.*

Proof. For the pattern $\mathbf{b}_1^n$:

- $t_{000} = n - 3m - 2$ which is the minimum possible value of t_{000} among all other error patterns of length n and Hamming weight $0 < m < \frac{n}{3}$.

- $t_{00} = n - 2m - 1$ which is also the minimum value of t_{00} among all other error patterns of length n and Hamming weight $0 < m < \frac{n}{2}$ (see Remark 5.1).

- $t_{101} = t_{111} = t_{11} = 0$ are all minimized.
- $t_{010} = m$ is maximized.

The only way to minimize further the probability given by (6.8) is by reducing t_{010} since all the other terms are already minimized. We can reduce t_{010} by having successive ones in the error pattern. This means that every time t_{010} is reduced by one, at least t_{11} is increased by one. Assuming all the other terms keep their minimal values, the probability of new error pattern changes by a factor of at least:

$$\begin{aligned}
\left[\frac{\frac{\epsilon}{1-\epsilon}+p}{p}\right]\left[\frac{1-p}{\alpha\frac{\epsilon}{1-\epsilon}+1-p}\right] &= \left[\frac{\frac{\epsilon+p(1-\epsilon)}{1-\epsilon}}{p}\right]\left[\frac{1-p}{\frac{\alpha\epsilon+(1-p)(1-\epsilon)}{1-\epsilon}}\right] \\
&= \frac{\epsilon(1-p)+p(1-p)(1-\epsilon)}{\alpha p\epsilon+p(1-p)(1-\epsilon)} \\
&> 1 \qquad \text{iff } \alpha < \frac{1-p}{p}.
\end{aligned}$$

Since we already assume that $\alpha \leq 1 < \frac{1-p}{p}$, then the new error pattern is more likely than $\mathbf{b}_1^n$, which concludes the proof. □

Theorem 6.4. *Define:*

$$m_1(\alpha,\epsilon,p) \;:=\; \frac{\log\left\{\dfrac{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)}\right\}}{\log\left\{\dfrac{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^2}{p(\dfrac{\alpha\epsilon}{1-\epsilon}+1-p)(\dfrac{\epsilon}{1-\epsilon}+1-p)}\right\}}$$

$$m_2(n,\alpha,\epsilon,p) \;:=\; \begin{cases} \dfrac{n\log\left[\dfrac{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p}{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p}\right]+A}{C}, & \textit{if } \alpha > \dfrac{(1-2p)(1-\epsilon)}{\epsilon}-1 \\[2ex] \dfrac{n\log\left[\dfrac{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p}{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p}\right]+B}{C}, & \textit{otherwise,} \end{cases}$$

where

$$A \;:=\; \log\left[\frac{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]^2(1-p)\left(\dfrac{\epsilon}{1-\epsilon}+1-p\right)}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^2}\right],$$

$$B \;:=\; \log\left[\frac{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]^3\left(\dfrac{\epsilon}{1-\epsilon}+1-p\right)}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^2}\right]$$

$$C \;:=\; \log\left[\frac{\left(\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right)^3}{p\left(\dfrac{\alpha\epsilon}{1-\epsilon}+1-p\right)\left(\dfrac{\epsilon}{1-\epsilon}+1-p\right)}\right].$$

*For **any** two words $\mathbf{z}_1^n$ and $\bar{\mathbf{z}}_1^n$ generated by the QBC with $M=2$ satisfying* *i.* $w_H(\mathbf{z}_1^n) = m$, *where* $0 \geq m < \frac{n}{3}$,

ii. $w_H(\bar{\mathbf{z}}_1^n) = m + i$, *where* $1 \leq i \leq n - m$

we have that

$$m < \tilde{m}(n, \alpha, \epsilon, p) := \min\{m_1(\alpha, \epsilon, p), m_2(n, \alpha, \epsilon, p)\} \iff \frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} > 1.$$

Proof. Define $\mathbf{a}_1^n(t)$ and $\mathbf{b}_1^n(t)$ be the following n-bit patterns of Hamming weight t:

$$\begin{aligned}
\mathbf{a}_1^n(t) &= (000...111) \qquad \text{For } 0 < t < n \\
\mathbf{b}_1^n(t) &= (001001...000) \qquad \text{For } 0 < t < \frac{n}{3}
\end{aligned}$$

We first prove the first direction ($\Longrightarrow$): Consider the following two cases:

Case 1: $m = 0$

In this case, $\mathbf{z}_1^n$ is the all-zero error pattern. From Lemma 2.2, $P(\mathbf{Z}_1^n = 0^n) > P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)$.

- Case 2: $0 < m < \frac{n}{3}$ and $1 \leq i < n - m - 1$

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} &\geq \frac{\min_{\mathbf{z}_1^n \in \mathbf{F}_2^n : w_H(\mathbf{z}_1^n) = m} P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{\max_{\mathbf{z}_1^n \in \mathbf{F}_2^n : w_H(\mathbf{z}_1^n) = m+i} P(\mathbf{Z}_1^n = \mathbf{z}_1^n)} \\
&= \frac{P(\mathbf{Z}_1^n = \mathbf{b}_1^n(m))}{P(\mathbf{Z}_1^n = \mathbf{a}_1^n(m+i))} \\
&= \left[\frac{1-p}{p}\right]^i \left[\frac{(1+\alpha)\dfrac{\epsilon}{1-\epsilon} + 1 - p}{\dfrac{\epsilon}{1-\epsilon} + 1 - p}\right]^{(n-3m-2)-(n-m-i-2)}
\end{aligned}$$

$$
\begin{aligned}
& \left[\frac{\alpha\dfrac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{m}\left[\frac{(1+\alpha)\dfrac{\epsilon}{1-\epsilon}+p}{\dfrac{\epsilon}{1-\epsilon}+p}\right]^{-(m+i-2)} \\
& \left[\frac{\dfrac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{(n-2m-1)-(n-m-i-1)} \qquad \left[\frac{\dfrac{\epsilon}{1-\epsilon}+p}{p}\right]^{-(m+i-1)} \\
= \; & \left[\frac{\dfrac{\alpha\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{m}\left[\frac{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p}{\dfrac{\epsilon}{1-\epsilon}+p}\right]^{-(m+i-2)} \qquad \left[\frac{\dfrac{\epsilon}{1-\epsilon}+p}{p}\right]^{-(m+i-1)} \\
& \left[\frac{1-p}{p}\right]^{i}\left[\frac{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p}{\dfrac{\epsilon}{1-\epsilon}+1-p}\right]^{i-2m}\left[\frac{\dfrac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{i-m}
\end{aligned}
$$

where the first equality follows from Lemmas 6.1 and 6.2. Thus,

$$
\begin{aligned}
\frac{P(\mathbf{Z}_1^n=\mathbf{z}_1^n)}{P(\mathbf{Z}_1^n=\bar{\mathbf{z}}_1^n)} \;\geq\; & \left\{\frac{p(\dfrac{\alpha\epsilon}{1-\epsilon}+1-p)(\dfrac{\epsilon}{1-\epsilon}+1-p)}{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^{2}}\right\}^{m} \\
& \left[\frac{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p}{\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p}\right]^{i}\frac{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]^{2}}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)} \\
\geq\; & \left\{\underbrace{\frac{p(\dfrac{\alpha\epsilon}{1-\epsilon}+1-p)(\dfrac{\epsilon}{1-\epsilon}+1-p)}{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^{2}}}_{\leq 1}\right\}^{m} \\
& \frac{\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]\left[\dfrac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)}
\end{aligned}
\tag{6.16}
$$

$$> \quad 1. \tag{6.17}$$

We have inequality in (6.16) since we set $i = 1$ (indeed, the term raised to the power i is greater than 1 and hence is increasing in i where $i \geq 1$). The last strict inequality is a result of the condition $m < m_1(\alpha, \epsilon, p)$.

- Case 3: $0 < m < \frac{n}{3}$ and $n - m - 1 \leq i \leq n - m$

In this case, $t_{000}(\bar{\mathbf{z}}_1^n) = t_{00}(\bar{\mathbf{z}}_1^n) = 0$. We have

$$\begin{aligned}
\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} &\geq \frac{P(\mathbf{Z}_1^n = \mathbf{b}_1^n(m))}{P(\mathbf{Z}_1^n = \mathbf{a}_1^n(m+i))} \\
&= \left[\frac{1-p}{p}\right]^i \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+1-p}{\frac{\epsilon}{1-\epsilon}+1-p}\right]^{(n-3m-2)} \\
&\quad \left[\frac{\alpha\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^m \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+p}{\frac{\epsilon}{1-\epsilon}+p}\right]^{-(m+i-2)} \\
&\quad \left[\frac{\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^{(n-2m-1)} \left[\frac{\frac{\epsilon}{1-\epsilon}+p}{p}\right]^{-(m+i-1)} \\
&= \left[\frac{(1+\alpha)\frac{\epsilon}{1-\epsilon}+1-p}{1-p}\right]^n \left[\frac{1-p}{(1+\alpha)\frac{\epsilon}{1-\epsilon}+p}\right]^i \\
&\quad \left[\frac{p\left(\frac{\epsilon}{1-\epsilon}+1-p\right)\left(\frac{\alpha\epsilon}{1-\epsilon}+1-p\right)(1-p)}{\left(\frac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right)^3\left[\frac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]}\right]^m \\
&\quad \frac{\left[\frac{(1+\alpha)\epsilon}{1-\epsilon}+p\right]^2(1-p)\left(\frac{\epsilon}{1-\epsilon}+1-p\right)}{p\left(\frac{\epsilon}{1-\epsilon}+p\right)\left[\frac{(1+\alpha)\epsilon}{1-\epsilon}+1-p\right]^2} \\
&> 1.
\end{aligned}$$

The last inequality is a result of the condition $m < m_2(n, \alpha, \epsilon, p)$.

We now prove the other direction ($\Longleftarrow$):

- Assume $m \geq m_1(\alpha, \epsilon, p)$: In the proof of Case 2, all the inequalities except the last one can be met with equality by choosing the error patterns as follows: $\mathbf{z}_1^n = \mathbf{b}_1^n(m)$ and $\bar{\mathbf{z}}_1^n = \mathbf{a}_1^n(m+1)$. Under the assumption that $m \geq m_1(\alpha, \epsilon, p)$, we get:

  $$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \leq 1.$$

  Therefore, we proved that there exist at least two words $\mathbf{z}_1^n$ and $\bar{\mathbf{z}}_1^n$ satisfying:

  i. $w_H(\mathbf{z}_1^n) = m$, where $0 < m \leq \frac{n}{2}$

  ii. $w_H(\bar{\mathbf{z}}_1^n) = m + i$, where $1 \leq i \leq n - m$

  such that:

  $$\frac{P(\mathbf{Z}_1^n = \mathbf{z}_1^n)}{P(\mathbf{Z}_1^n = \bar{\mathbf{z}}_1^n)} \leq 1.$$

- Assume $m \geq m_2(n, \alpha, \epsilon, p)$: The proof follows a similar reasoning as above, only this time we choose $\bar{\mathbf{z}}_1^n$ to be the all-one error word (while $\mathbf{z}_1^n$ is unchanged).

□

Theorem 6.5. *Let $\mathcal{C}$ be any (n, K, d) code used over the QBC with $M = 2$ and parameters α, ϵ and p. Denote by r_{cov} the classical covering radius of this code. If $r_{cov} < \min\left\{\tilde{m}(n, \alpha, \epsilon, p), \frac{n}{3}\right\}$, then the output of the SMD decoder (when it does* *not declare a decoding failure) is identical to the output of the ML decoder for this code.*

Proof. Let $\mathbf{y}$ be the received word. Let $m := \min_{\mathbf{c}\in\mathcal{C}} \mathrm{d}_H(\mathbf{y},\mathbf{c})$,where d_H denotes the Hamming distance. Clearly, $m \leq r_{cov} < \frac{n}{3}$ (from the definition of the covering radius). If there exists a unique codeword $\hat{\mathbf{c}}$ such that $\mathrm{d}_H(\mathbf{y},\hat{\mathbf{c}}) = m$, then the SMD decoding gives a valid codeword. Since $m < \tilde{m}(n,\alpha,\epsilon,p)$, it follows from Theorem 6.4 that all other error words of larger Hamming weights have a smaller probability than the error word corresponding to the SMD decision. Hence the ML decoder will give the same output. □

Corollary 6.2. *Let $\mathcal{C}$ be an (n, M, d_{min}) perfect code (in the classical sense) used over the QBC with parameters α, ϵ and p. If*

$$r_{cov} = \left\lfloor \frac{d_{min}-1}{2} \right\rfloor \quad < \quad \min\left\{\tilde{m}(n,\alpha,\epsilon,p), \frac{n}{3}\right\},$$

then $\mathcal{C}$ is a generalized perfect code for this channel and hence is optimal (under ML decoding) among all codes of the same length and dimension sent over the same channel.

Proof. Immediate from Lemma 3.3 and Theorems 6.4 and 6.5. □

Remark 6.4. *Note that we can not make a similar statement to the one in Corollary 6.2 for quasi-perfect codes, since two error words of the same weight are not guaranteed to have the same probability.*

Remark 6.5. *When $\alpha = 0$, the QBC with $M = 2$ reduces to the BFMNC (or QBC with $M = 1$) for the same BER p and correlation coefficient ϵ (see Remark 2.1)* *and, unsurprisingly, we can prove that Theorem 6.4 implies Theorem 5.1 and consequently Theorem 6.5 implies Theorem 5.2.*

Corollary 6.3. *The results in Theorems 6.4 and 6.5 and in Corollary 6.2 can be specialized to the FMCC with $M = 2$ by setting $\alpha = 1$. The expressions of $m_1(\alpha, \epsilon, p), m_2(n, \alpha, \epsilon, p)$ and $\tilde{m}(n, \alpha, \epsilon, p)$ reduce in this case to the following form:*

$$
\begin{aligned}
m_1(\epsilon, p) &= \frac{\log\left\{\dfrac{\left[\dfrac{2\epsilon}{1-\epsilon}+1-p\right]\left[\dfrac{2\epsilon}{1-\epsilon}+p\right]}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)}\right\}}{\log\left\{\dfrac{\left[\dfrac{2\epsilon}{1-\epsilon}+p\right]\left[\dfrac{2\epsilon}{1-\epsilon}+1-p\right]^2}{p\left(\dfrac{\epsilon}{1-\epsilon}+1-p\right)^2}\right\}} \\
m_2(n, \epsilon, p) &= \begin{cases} \dfrac{n\log\left[\dfrac{\dfrac{2\epsilon}{1-\epsilon}+1-p}{\dfrac{2\epsilon}{1-\epsilon}+p}\right]+A}{C}, & \textit{if } \dfrac{(1-2p)(1-\epsilon)}{\epsilon} < 2 \\ \dfrac{n\log\left[\dfrac{\dfrac{2\epsilon}{1-\epsilon}+1-p}{\dfrac{2\epsilon}{1-\epsilon}+p}\right]+B}{C}, & \textit{otherwise,} \end{cases}
\end{aligned}
$$

where

$$
A = \log\left[\frac{\left[\dfrac{2\epsilon}{1-\epsilon}+p\right]^2(1-p)\left(\dfrac{\epsilon}{1-\epsilon}+1-p\right)}{p\left(\dfrac{\epsilon}{1-\epsilon}+p\right)\left[\dfrac{2\epsilon}{1-\epsilon}+1-p\right]^2}\right],
$$

$$\begin{aligned} B &= \log\left[\frac{\left[\frac{2\epsilon}{1-\epsilon}+p\right]^3\left(\frac{\epsilon}{1-\epsilon}+1-p\right)}{p\left(\frac{\epsilon}{1-\epsilon}+p\right)\left[\frac{2\epsilon}{1-\epsilon}+1-p\right]^2}\right] \\ C &= \log\left[\frac{\left(\frac{2\epsilon}{1-\epsilon}+1-p\right)^3}{p\left(\frac{\epsilon}{1-\epsilon}+1-p\right)^2}\right]. \end{aligned}$$

and finally

$$\tilde{m}(n,\epsilon,p) \quad := \quad \min\{m_1(\epsilon,p), m_2(n,\epsilon,p)\}.$$

In Figs. 6.1-6.3, we plot $\tilde{m}(n,\epsilon,p)$ from Corollary 6.3 versus the FMCC with $M=2$ channel correlation coefficient ϵ for different values of the BER p and the block length n.

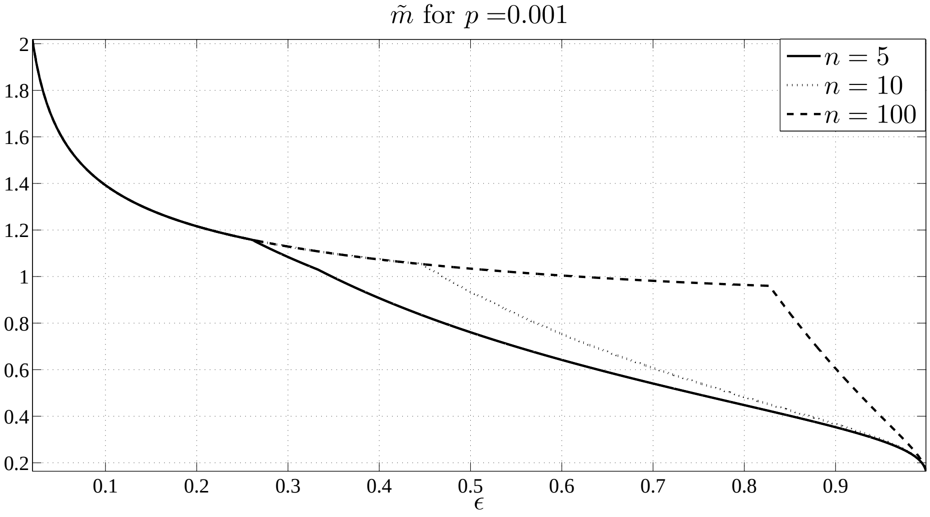


Figure 6.1: Plot of $\tilde{m}(n,\epsilon,p)$ with respect to ϵ for different values of n and for $p=0.001$, FMCC with $M=2$.

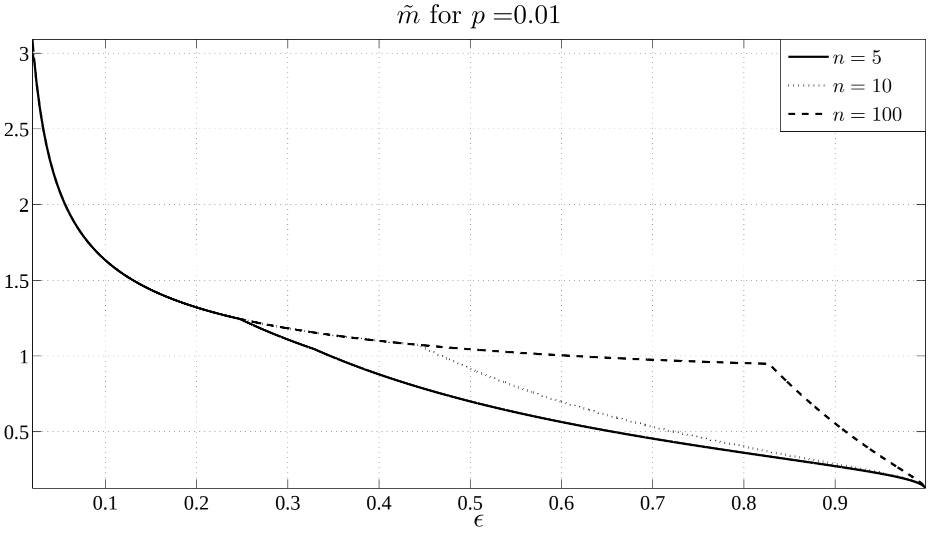


Figure 6.2: Plot of $\tilde{m}(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.01$, FMCC with $M = 2$.

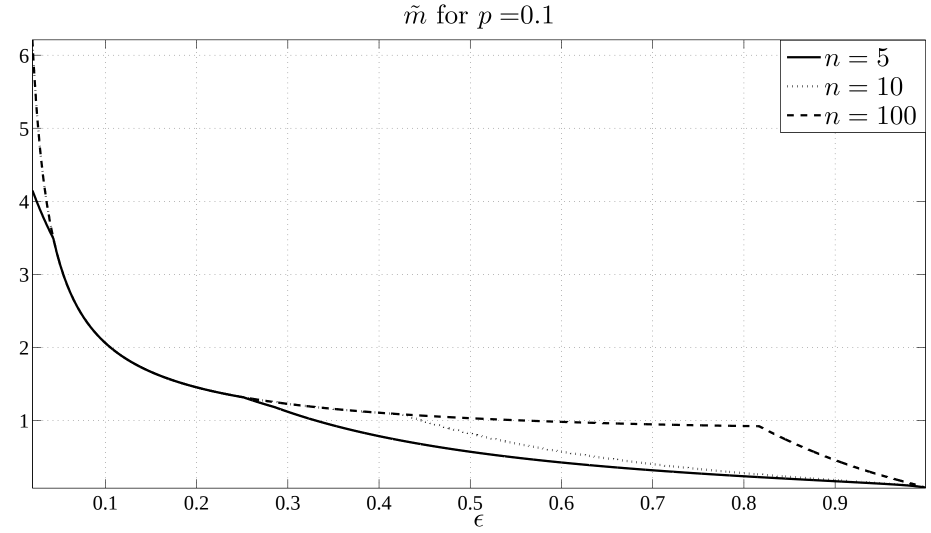


Figure 6.3: Plot of $\tilde{m}(n, \epsilon, p)$ with respect to ϵ for different values of n and for $p = 0.1$, FMCC with $M = 2$.

We notice from the above figures that the condition of Corollary 6.2 is restrictive for channels with $\epsilon > 0.1$. In fact, for these channels, only codes with a covering radius $r_{cov} = 1$ satisfy the condition (e.g., the family of Hamming codes). The result is unsurprisingly similar to the results we obtained for the BFMNC. For smaller ϵ, more codes satisfy the condition, and when $\epsilon = 0$ (i.e., when the QBC reduces to the BSC), unsurprisingly all block codes satisfy it.

In Figs. 6.4-6.6, we plot $\tilde{m}(n, \alpha, \epsilon, p)$ from Theorem 6.4 versus the QBC with $M = 2$ channel correlation coefficient ϵ for different values of the BER p and different values of α and for a block length $n = 10$.

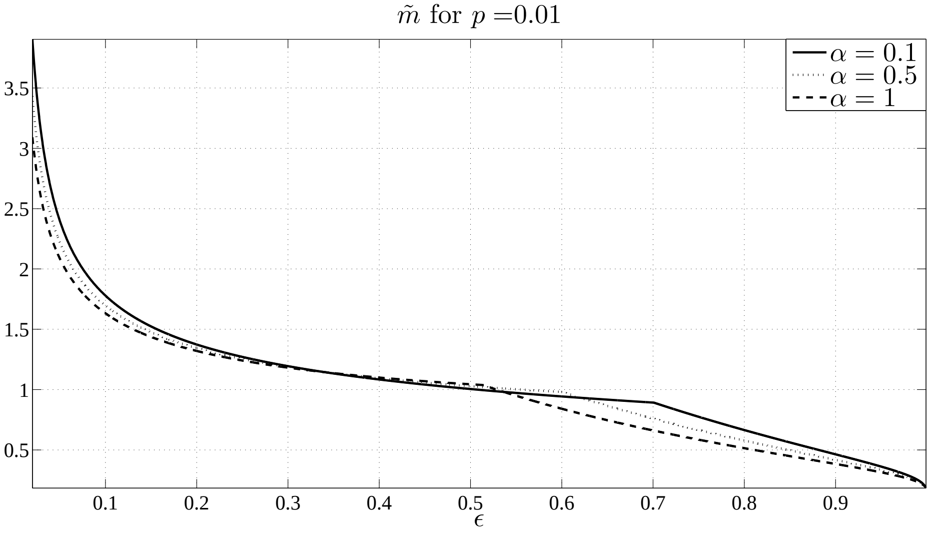


Figure 6.4: Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of α, for $p = 0.001$ and $n = 10$, QBC with $M = 2$.

We notice from the above figures that reducing α, while fixing the values of n, ϵ and p, results most of the time in increasing $\tilde{m}(n, \alpha, \epsilon, p)$. However, the gain is not significant.

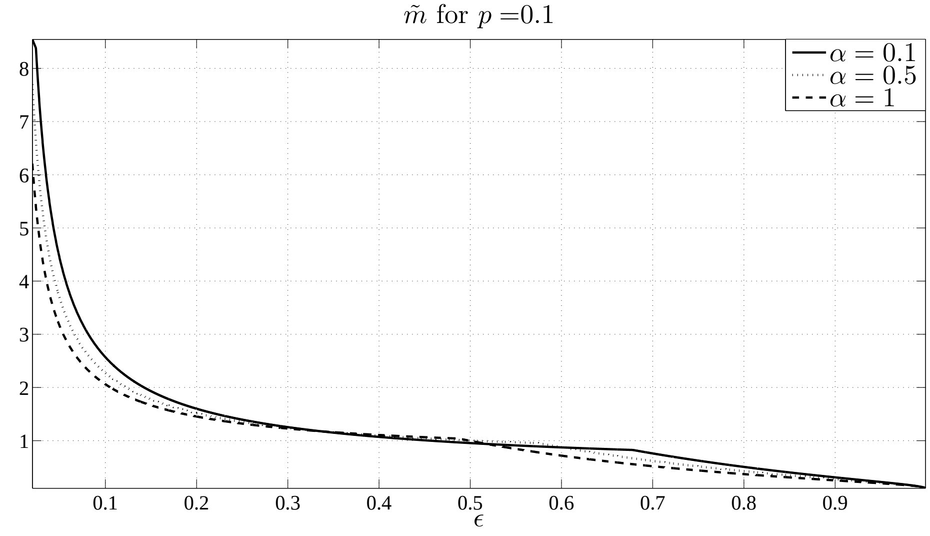


Figure 6.5: Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of α, for $p = 0.01$ and $n = 10$, QBC with $M = 2$.

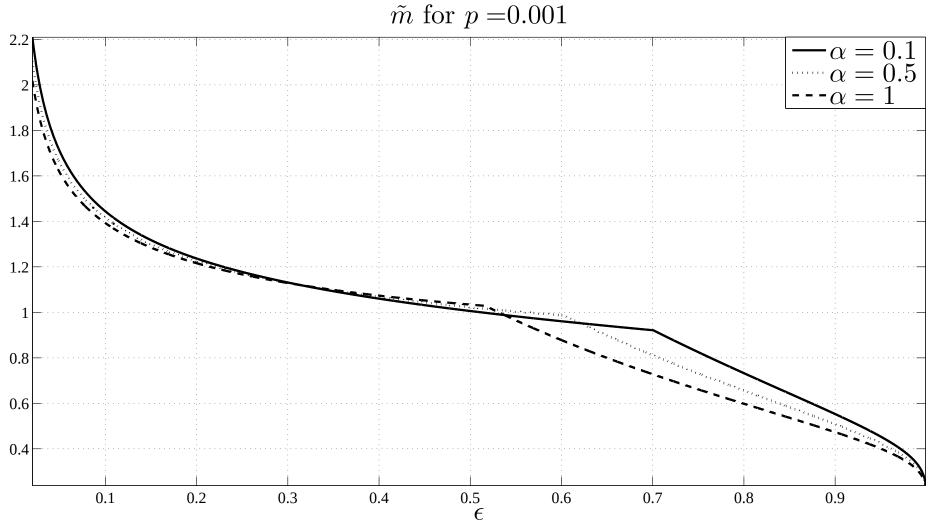


Figure 6.6: Plot of $\tilde{m}(n, \alpha, \epsilon, p)$ with respect to ϵ for different values of α, for $p = 0.1$ and $n = 10$, QBC with $M = 2$.

6.4 Numerical Results

We illustrate the condition of Theorem 6.5 by simulating the performance of different codes under both SMD and ML decoding over the FMCC with parameters

$M = 2$, ϵ and p. Since Theorem 6.5 does not treat the case when SMD declares a failure, we disregard this case in our simulations as well by only using the ML decoder when the SMD decoder does not declare a failure.

The first code we simulate is the $[7, 4, 3]$ perfect Hamming code. It has a covering radius $r_{cov} = 1$. We show the results in Table 6.1.

ϵ	0.1	0.2	0.3	0.6	0.7	0.8
$\tilde{m}(n, \epsilon, p)$	2.0606	1.4544	1.2279	0.4853	0.3529	0.2542
PCE under SMD	0.13799	0.12767	0.11344	0.07217	0.0604	0.03941
PCE under ML	0.13799	0.12767	0.11344	0.06831	0.05267	0.03511

Table 6.1: Verifying Theorem 6.5 for the $[7, 4, 3]$ Hamming code over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ.

According to Theorem 6.5, if $\tilde{m}(n, \epsilon, p) > 1$ the output of the SMD decoder (when it does not declare a decoding failure) is identical to the output of the ML decoder for this code. Indeed, as Table 6.1 shows, the probabilities of codeword error for SMD and ML decoding (when SMD does not declare a decoding failure) over the FMCC with $M = 2$ are identical. We start noticing some discrepancy when $\tilde{m}(n, \epsilon, p) < 1$. Similarly, we simulate the performance of the $(15, 2^{11}, 3)$ Vasil'ev nonlinear perfect code constructed using the method described in Theorem 3.1 using

$$f(\mathbf{x}_1^7) = x_1x_2x_7 \oplus x_2x_4x_6 \oplus x_1x_3x_5.$$

. This code has also a covering radius $r_{cov} = 1$. We show the results in Table 6.2. The results are similar to those observed in Table 6.1 for the $[7, 4, 3]$ Hamming code, i.e., we notice that the probabilities of codeword error for SMD and ML decoding (when SMD does not declare a decoding failure) over the FMCC with $M = 2$ are identical when $\tilde{m}(n, \epsilon, p) > 1$ and they start to differ when $\tilde{m}(n, \epsilon, p) <$

ϵ	0.1	0.2	0.3	0.6	0.7	0.8
$\tilde{m}(n,\epsilon,p)$	2.0606	1.4544	1.2279	0.7230	0.4873	0.3215
PCE under SMD	0.3875	0.3343	0.2841	0.1573	0.1184	0.0908
PCE under ML	0.3875	0.3343	0.2841	0.15	0.1124	0.0789

Table 6.2: Verifying Theorem 6.5 for the $(15, 2^{11}, 3)$ Vasil'ev nonlinear perfect code over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ.

1. Similarly, we simulate the performance of the $[8, 4, 4]$ Reed-Muller code ($r_{cov} = 2$) and the $[24, 12, 8]$ extended Golay code ($r_{cov} = 3$). The results are shown in Tables 6.3 and 6.4, respectively.

ϵ	0.05	0.075	0.1	0.6	0.7	0.8
$\tilde{m}(n,\epsilon,p)$	2.6667	2.4344	2.0606	0.5150	0.3697	0.2626
PCE under SMD	0.0413811	0.0419236	0.0437669	0.0274721	0.0344286	0.0214617
PCE under ML	0.0413811	0.0419236	0.0437669	0.0274721	0.0218041	0.0142804

Table 6.3: Verifying Theorem 6.5 for the $[8, 4, 4]$ Reed Muller code $\mathcal{RM}(1, 3)$ over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ.

ϵ	0.02	0.025	0.03	0.2	0.3	0.4
$\tilde{m}(n,\epsilon,p)$	6.2104	5.2032	4.5269	1.4544	1.2279	1.1068
PCE under SMD	0.0791618	0.0570431	0.0742459	0.061086	0.0588235	0.0536481
PCE under ML	0.0791618	0.0570431	0.0742459	0.0554299	0.0374332	0.0482833

Table 6.4: Verifying Theorem 6.5 for the $[24, 12, 8]$ extended Golay code over the FMCC with parameters $M = 2$, $p = 0.1$ and ϵ.

Chapter 7

Conclusion and Future Work

7.1 Concluding Remarks

In this work, we presented sufficient conditions on general binary codes under which SMD and ML are equivalent over the QBC with $M = 1$ (or the BFMNC), the QBC with $M = 2$, and the GEC (when the state vector is not available at the decoder). We also determined sufficient conditions under which classical perfect codes are optimal under ML decoding over these channels. For the IMCC and the QBC with $M \geq n$, we provided both necessary and sufficient conditions on binary codes for which ML and MD are equivalent. We also determined sufficient conditions under which classical perfect and quasi-perfect codes are optimal under ML decoding over these channels. For the GEC when the state vector is available at the decoder, we gave partial results pertaining to equivalence between the Hamming metric and the likelihood of error patterns generated by this channel.

7.2 Application to Syndrome Source Coding

The results that we derived in this work can be applied to syndrome source coding with or without side information [3, 28]. Similar work has been done to Markov sources in [27]. The syndrome source coding scheme without side information is shown in Fig. 7.1.

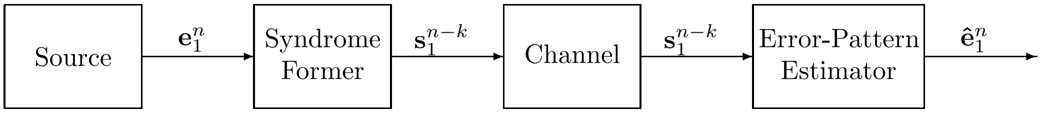


Figure 7.1: Syndrome source coding method of using error correcting codes for data compression

The scheme uses an $[n, k, d_{min}]$ *linear* code $\mathcal{C}$. The source encoder computes the syndrome $\mathbf{s}_1^{n-k}$ of the source output $\mathbf{e}_1^n$ and sends it over a noiseless communication channel. The source decoder outputs its estimate $\hat{\mathbf{e}}_\mathbf{1}^\mathbf{n}$ of the source output from its syndrome. If the n-bit source output is identically distributed to the n-bit error pattern generated by one of the channels that we considered in this work and if the *linear* code $\mathcal{C}$ satisfies the conditions we presented for that channel, then we obtain the same equivalence relation between the ML and the MD (or SMD) decoders.

Alternatively, the syndrome source coding scheme with side information at the decoder is shown in Fig. 7.2.

The source gives two outputs Y_i and X_i, where the latter is only available at the decoder. Let $U_i = X_i \oplus Y_i$. In this scheme, the source encoder computes the syndrome $\mathbf{z}_1^{n-k}$ of the n-bit source output $\mathbf{y}_1^n$ and transmits it over a noiseless

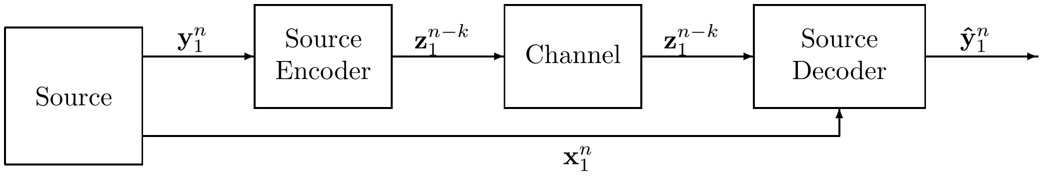


Figure 7.2: Source encoder with side information at the receiver

communication channel. The source decoder computes the syndrome of the second n-bit source output $\mathbf{x}_1^n$ and adds it bitwise to $\mathbf{z}_1^{n-k}$ (modulo-2). It can be easily proven that the result is the syndrome of $\mathbf{u}_1^n = \mathbf{x}_1^n \oplus \mathbf{y}_1^n$. Hence, the decoder computes its estimate $\hat{\mathbf{u}}_1^n$ of $\mathbf{u}_1^n$ from its syndrome and outputs $\hat{\mathbf{y}}_1^n = \hat{\mathbf{u}}_1^n \oplus \mathbf{x}_1^n$. Similarly, if the $\mathbf{U}_1^n$ is identically distributed to the n-bit error pattern of one of the channels that we considered in this work and if the *linear* code $\mathcal{C}$ satisfies the condition we presented for that channel, then we obtain the same equivalence relation between the ML and the MD (or SMD) decoders.

7.3 Future Work

Future work may include extending the results to other channel models with memory, and particularly the M^{th}-order QBC and the GEC when the state vector is available at the decoder. By noticing a pattern on the most and least likely error patterns generated by the QBC with $M = 1$ and $M = 2$, we make a conjecture about the most and least likely error patterns for the general M^{th}-order QBC.

Conjecture 7.1. *We make the following two conjectures:*

- *The error patter* $\mathbf{a}_1^n$ *of Hamming weight* $0 < m < n$ *where all zeros and ones* *are consecutive (e.g.,* $\mathbf{z}_1^n = 00...011...11$*) is the most likely among all other error patterns of the same length and weight generated by the* M^{th}*-order QBC.*

- *Consider the error words of length* n *having a Hamming weight* $0 < m < \frac{n}{M+1}$*. The pattern*

$$\mathbf{b}_1^n \quad = \quad \left(\underbrace{\underbrace{00...0}_{M}\,1\,\underbrace{00...0}_{M}\,...\,\underbrace{00...0}_{M}\,1}_{m}\,00...00 \right)$$

  is the least likely among all patterns of the same length and weight generated by the M^{th}*-order QBC.*

Conjecture 7.1 might prove very useful in deriving conditions on error patterns similar to the ones given in Theorems 5.1 and 6.4 and hence in deriving conditions on binary block codes under which SMD and ML decoding are equivalent and for which classical perfect codes are optimal under ML decoding over the M^{th}-order QBC.

Another interesting direction is to study optimal or sub-optimal structures of binary block codes over channels with memory as we have established that the Hamming distance is not necessarily the most important parameter in the code design.

There is also room for improvement in the results obtained for the GEC when the state vector is not available at the decoder. In fact, the derived condition for this case might be too loose and can be potentially tightened to include a wider class of block linear codes.

Finally, the construction of decoders that exploit the memory between blocks via the use of estimates of the previous channel noise samples is a worthwhile future direction. Such endeavour will improve the system's error performance vis-a-vis the (memoryless) block-by-block decoding considered in this work for a cost of increased complexity and delay.

Bibliography

[1] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2035-2041, Nov. 1994.

[2] H. Al-Lawati and F. Alajaji, "On decoding binary perfect and quasi-perfect codes over Markov noise channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 873-878, Apr. 2009.

[3] T. Ancheta, "Syndrome source coding and its universal generalization," *IEEE Trans. on Inform. Theory*, vol. 22, no. 4, pp. 432-436, July 1976.

[4] R. E. Blahut, *Algebraic Codes for Data Transmission.* Cambridge university press, 2003.

[5] A. Cohen, F. Alajaji, N. Kashyap, and G. Takahara, "LP decoding for joint source-channel codes and for the non-ergodic Polya channel," *IEEE Commun. Letters, IEEE*, vol. 12, no. 9, pp. 678-680, Sep. 2008.

[6] T. M. Cover and J. A. Thomas, "Elements of Information Theory." Wiley, 1991.

[7] A. Eckford, F. Kschischang, and S. Pasupathy, “Analysis of low-density parity-check codes for the Gilbert-Elliott channel,” *IEEE trans. Inform. Theory,*, vol. 51, pp. 3872–3889, Nov. 2005.

[8] M. Effros, A. Goldsmith and Y. Liang, “Generalizing capacity: New definitions and capacity theorems for composite channels,” *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3069-3087, July 2010.

[9] E. O. Elliott, “Estimates of error rates for codes on burst-noise channels, *Bell Syst. Tech. J.*, vol. 42, pp. 1977-1997, Sep. 1963.”

[10] R. G. Gallager, *Information Theory and Reliable Communication*, New York: Wiley, 1968.

[11] J. Garcia-Frias, “Decoding of low-density parity check codes over finite-state binary Markov channels,” in *IEEE trans. Commun.,*, vol. 52, pp. 1840–1843, Nov. 2004.

[12] E. N. Gilbert, “Capacity of a burst-noise channel,” *Bell Syst. Tech. J*, vol. 39, pp. 1253-1266, Sep. 1960.

[13] M. Hamada, “A sufficient condition for a code to achieve the minimum decoding error probability–generalization of perfect and quasi-perfect codes,” *IEICE Trans. Fundamentals Electronics, Commun. and Comp. Sciences*, vol. E83-A, no. 10, pp. 1870-1877, Oct. 2000.

[14] M. Hamada, “Near-optimal codes on the two-state Markovian additive channel,” in *Proc. IEEE Int. Symp. Inform. Theory*, 2001, p. 246.

[15] M. Mushkin and I. Bar-David,“Capacity and coding for the Gilbert-Elliott channel,” *IEEE Trans. Inform. Theory*, vol. 35, pp. 1277–1290, Nov. 1989.

[16] C. Pimentel, F. Alajaji and P. Melo, “A discrete queue-based model for capturing memory and soft-decision information in correlated fading channels,” *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1702-1711, June 2012..

[17] C. Nicola, F. Alajaji, and T. Linder, “Decoding LDPC codes over binary channels with additive Markov noise,” in *Proc. CWIT ’05.*, pp. 187-190, Montreal, June 2005.

[18] P. Sadeghi and P. Rapajic, “Capacity analysis for finite-state Markov mapping of Markov channels,” *IEEE Trans. Commun.*, vol. 53, pp. 833-840, May 2005.

[19] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379-423, 1948.

[20] C. H. Stapper, A. N. McLaren and M. Dreckmann, “Yield Model for productivity optimization of VLSI memory chips with redundancy and partially good product,” *IBM Journal of Research and Development*, vol. 24, pp. 398-409, 1980.

[21] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *Euro. Trans. Telecomm. (ETT)*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[22] A. Tietäväinen, “On the nonexistence of perfect codes over finite fields,” *SIAM Journal on Applied Mathematics*, vol. 24, pp. 88-96, 1973.

[23] J. H. Van Lint, *Introduction to Coding Theory.* Springer Verlag, 1999.

[24] J. H. Van Lint, "Nonexistence theorems for perfect error-correcting codes," *American Mathematical Society*, 1971.

[25] J. L. Vasil'ev, "On nongroup close-packed codes," In *Probl. Kibernet.* , vol. 8, pp. 375-378, 1962.

[26] Z. Wang, M. G. Karpovsky and K. J. Kulikowski, "Replacing linear Hamming codes by robust nonlinear codes results in a reliability improvement of memories," *IEEE/IFIP Int. Conf. Dependable Systems & Networks*, pp. 514-523, 2009.

[27] X. Wu, *Syndrome Source Coding for Lossless Data Compression Based on Linear Block Codes,* M.Sc. thesis, Dept. Mathematics and Statistics, Queen's University, Sep. 2008.

[28] A. D. Wyner and J. Ziv, "The rate-distorion function for source coding with side information at the decoder," *IEEE Trans. Inform. Theory,* vol. 22, pp. 1-10, Jan. 1976.

[29] L. Zhong, F. Alajaji and G. Takahara, "A binary communication channel with memory based on a finite queue," *IEEE Trans. Inform. Theory,* vol. 53, pp. 2815-2840, Aug. 2007.

[30] L. Zhong, F. Alajaji, and G. Takahara, "A model for correlated Rician fading channels based on a finite queue," *IEEE Trans. Veh. Technology*, vol. 57, no. 1, pp. 79-89, Jan. 2008.