Syndrome Source Coding for Lossless Data Compression Based on Linear Block Codes

by

Xiaoyan Wu

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

Queen's University
Kingston, Ontario, Canada
Sep. 2008

Abstract

In this work we focus on syndrome source coding for lossless data compression with and without side information based on fixed-length linear block codes.

For syndrome source coding without side information, we prove that this scheme can achieve the source entropy rate when the source is stationary and ergodic. Then we consider applying short-length perfect or quasi-perfect codes for syndrome source coding to reduce the system's delay. We describe the minimum distance (MD), strict minimum distance (SMD), maximum likelihood (ML) and strict maximum likelihood (SML) decoding methods for this scheme, and examine their relationship for Markov sources using perfect and quasi-perfect codes. For Markov sources, we also use a modified MD decoding method - so called the MD+ decoding method. Moreover we provide simulation results using Hamming, BCH and Golay codes under the different decoding methods.

For syndrome source coding with side information, we prove that this scheme can achieve the Slepian-Wolf limit when the correlation channel is an additive noise channel with a stationary and ergodic noise process. We also consider employing short-length perfect and quasi-perfect codes for this scheme. We describe the MD, SMD, ML, SML decoding methods for this scheme, and examine their relationship for sources with Markov noise correlation channel using perfect and quasi-perfect codes. The MD+ decoding method is described for sources with Markov noise correlation channel. Furthermore, we introduce a more common model with the side information as output and describe the algorithm for optimal MAP decoding. Finally, we provide simulation results using

Hamming, BCH and Golay codes under the different decoding methods.

Acknowledgments

I would like to express my sincere thank to my supervisors, Dr. Fady Alajiaji and Dr. Tamas Linder for their helpful guidance and significant contribution to my work. I appreciate their time and patience for supervising my research.

Contents

List of Figures

Chapter 1

Introduction

1.1 Problem Description

Lossless data compression has numerous applications in information and communications technologies. Shannon [16] laid out the basics of information theory and proved that the entropy rate is the ultimate limit of the rate for lossless data compression for identically independent distributed sources.

The design of low-complexity fixed-length source coding algorithms to losslessly compress a Markov source $\{X_i\}_{i=1}^{\infty}$ at rates close to its entropy rate remains a subject of intensive research. In this work we study syndrome source coding based on linear block codes and prove that this scheme can asymptotically achieve the entropy rates of Markov sources.

Low density parity check codes (LDPC) [10] [13] are good candidates for syndrome source coding as they can perform very close to the channel capacity. However, they may cause significant delay since they require very large block length to achieve good performance. However, in many practical applications there are strict constrains on the coding and decoding delay of the system. So in this work we consider using certain perfect or quasi-perfect codes with short block lengths.

In recent years, sensor networks have been a very active research area. One of the enabling technologies for sensor networks is distributed source coding (DSC) [20], where separate encoders compress the data without communicating with each other. One of the theoretical foundations for DSC is the Slepian-Wolf theorem [17], which shows that separate encoding is as efficient as joint encoding for lossless compression of independent, identically distributed (i.i.d.) sources. Wyner [18] proposed a scheme for data compression with side information and proved that if the correlated channel representing the side information is a binary symmetric channel (BSC), the scheme can achieve the Slepian-Wolf limit. Our goal is to extend this scheme to a more general situation with additive noise correlation channels, where the noise is stationary and ergodic.

Many wireless sensor network applications have real-time requirements where the sensor data must be sent to a base station within a very short time. For example, a sensor network that monitors temperature would require the sensors to report the temperature to a base station within very limited periods of time to detect rapid increases in the temperature. To reduce the delay caused by long coding block lengths, we also consider using short-length perfect and quasi-perfect block codes for source coding with side information.

1.2 Literature Review

For lossless data compression (without side information), in his landmark paper, ”A Mathematical Theory of Communication,” [16] Shannon established that there is a fundamental limit called entropy, which is the smallest possible rate for lossless data compression.

Elias [8] proposed an algorithm for fixed-length lossless data compression which uses the parity-check matrix of a linear block code to compress data via syndromes. Specifically, the encoder is characterized by an $m \times n$ matrix $\mathbf{H}$, which maps the source $x^n = (x_1, ..., x_n)$ to the syndrome $s^m = \mathbf{H}x^n$. The optimal decoding method is maximum likelihood (ML) decoding. This decoding method maps the received syndrome s^m to the most likely $\hat{x}^n$ which maximizes $\Pr(x^n)$ over all x^n such that $\mathbf{H}x^n = s^m$. He also proved that for i.i.d. sources under ML decoding, there exists a parity-check matrix to compress the data at a rate arbitrarily close to the source entropy with an asymptotically vanishing error probability. The main idea of the proof is using random codebooks and the fact that the probability that each coset has one and only one element in common with the set of typical sequences tends to 1 when the block length n tends to infinity.

In a related work [3], Ancheta proposed a similar syndrome source coding scheme based on linear block codes. The main idea is treating the source as the noise in an associated additive noise channel. He proved that the error probability for syndrome source coding using a linear block code on a given binary source is the same as the error probability when this code is used on the associated additive channel, where the noise statistics of the noise are the same as the source. He also proved that for binary i.i.d. sources, this scheme can asymptotically achieve the source entropy .

Caire et al. [4] proposed a data compression algorithm using LDPC codes to compress sources with or without memory. The algorithm is based on the concatenation of a syndrome source coding scheme using LDPC codes with the Burrows-Wheeler block sorting transform. For the syndrome source coding scheme, they adapted the Belief-Propagation decoding algorithm to decode the received syndrome.

For data compression with side information, it is known via the Slepian-Wolf theorem [17] that for two correlated i.i.d. sources $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$, we can compress $\{X_i\}_{i=1}^{\infty}$ losslessly at a rate arbitrarily close to conditional entropy $H(X|Y)$, where $\{Y_i\}_{i=1}^{\infty}$ is only available at the decoder. Cover [6] extended this theorem to stationary and ergodic sources $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$, and proved that the conditional entropy rate $H(\mathfrak{X}|\mathfrak{Y})$ is the limit for the data compression of $\{X_i\}_{i=1}^{\infty}$ with side information $\{Y_i\}_{i=1}^{\infty}$ at the decoder.

Wyner proposed in [18] a syndrome source coding scheme to compress data with side information. The correlation between the two sources can be modeled via a channel. At the encoder, one uses the parity check matrix of a linear code to compress the data using a syndrome. At the decoder, one estimates the source input from the received syndrome and the side information using the error pattern estimator of the linear code. Wyner also proved that this scheme can achieve the Slepian-Wolf limit when the correlation channel is a BSC.

Liveris et al. [12] applied LDPC codes for Wyner's syndrome source coding scheme for the case where the correlation channel is modeled as a BSC. Their simulation results show that the scheme's performance is very close to the Slepian-Wolf limit.

Garcia-Frias et al. [9] proposed the use of LDPC codes for syndrome source coding for the case where the correlation channel between the sources is a hidden Markov noise channel (a special case is Gilbert-Elliott channel (GEC)).

1.3 Overview of this Work

This work consists of four chapters.

In Chapter 1, we introduce basic definitions and results, including basic information measures, the source coding theorem, the Slepian-Wolf theorem and the channel coding theorem.

In Chapter 2, we focus on syndrome source coding without side information. The scheme for syndrome source coding is introduced. We prove that this scheme can achieve the entropy rate of the source asymptotically when the source is stationary and ergodic. Then we consider applying perfect and quasi-perfect codes with short block length in this scheme for (first order) Markov sources. We describe five decoding methods, namely minimum distance (MD) decoding, strict minimum distance (SMD) decoding, maximum likelihood (ML) decoding, strict maximum likelihood (SML) decoding and a modification of MD decoding (MD+). We examine the relationship among these decoding methods for the compression of Markov sources using perfect and quasi-perfect codes. Finally, we present simulation results using Hamming, BCH and Golay codes under these different decoding methods.

In Chapter 3, we focus on syndrome source coding with side information. First we introduce Wyner's syndrome source coding scheme for data compression with side information. Then we prove that this scheme can achieve the Slepian-Wolf limit asymptotically for sources with an additive noise correlation channel where the noise is stationary and ergodic. As in Chapter 2, we consider applying short-length perfect and quasi-perfect codes for this scheme. We describe MD, SMD, ML, SML, MD+ decoding methods for this scheme and analyze the relationship among them for sources with a Markov noise correlation channel. Furthermore, we study a more natural correlation model where the side information is the channel output. Finally, we provide simulation results for sources with a Markov noise correlation channel and sources with a GEC correlation channel using Hamming, BCH and Golay codes under these different decoding methods.

In Chapter 4, we state our conclusions and discuss directions for future work.

Chapter 2

Preliminaries

2.1 Information Measures

We start with the definition of the basic measures of information proposed by Shannon [16]. From now on, the logarithm is to the base 2 unless otherwise specified.

Definition 1 *The entropy of a random variable X with discrete alphabet $\mathcal{X}$ and probability distribution $p(x) = \Pr(X = x)$ is given by*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Definition 2 *Let X, Y be two discrete random variables with joint probability distribution $p(x, y)$, then the joint entropy of X given Y is given by*

$$H(X, Y) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y).$$

Definition 3 *Let X, Y be two discrete random variables with joint probability distribu-*

tion $p(x, y)$, then the conditional entropy of X given Y is given by

$$H(X|Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y)\log p(x|y).$$

Definition 4 *The mutual information between random variables X and Y defined over alphabet $\mathcal{X}$ and $\mathcal{Y}$, respectively, is defined by*

$$I(X; Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x, y)\log\frac{p(x, y)}{p(x)p(y)}.$$

These concepts can be extended to random processes.

Definition 5 *The entropy rate of the random process $\{X_i\}_{i=1}^{\infty}$ is given by*

$$H(\mathfrak{X}) = \lim_{n\to\infty}\frac{1}{n}H(X_1, X_2, ..., X_n).$$

The entropy rate may not exist for all random processes, but for a stationary source $\{X_i\}_{i=1}^{\infty}$, its entropy rate $H(\mathfrak{X})$ always exists and is equal to $H(X_n|X_{n-1}, X_{n-2}, ..., X_1)$.

Definition 6 *The joint entropy rate of the random processes $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ is given by*

$$H(\mathfrak{X}, \mathfrak{Y}) = \lim_{n\to\infty}\frac{1}{n}H(X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_n).$$

Definition 7 *The conditional entropy rate of the random processes $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ is given by*

$$H(\mathfrak{X}|\mathfrak{Y}) = \lim_{n\to\infty}\frac{1}{n}H(X_1, X_2, ..., X_n|Y_1, Y_2, ..., Y_n).$$

2.2 Source Coding

2.2.1 Shannon's Source Coding Theorem

Definition 8 *A discrete memoryless source (DMS) consists of a sequence of i.i.d. random variables* $\{X_i\}_{i=1}^{\infty}$.

Definition 9 *A* (k, n) *D-ary block (fixed-length) code for a discrete source defined on the alphabet* $\mathcal{X}$ *consists of*

- *encoder* $f : \mathcal{X}^n \rightarrow \{0, 1, 2, ..., D-1\}^k$*;*
- *decoder* $g : \{0, 1, 2, ..., D-1\}^k \rightarrow \mathcal{X}^n$.

Definition 10 *The rate of the block code is defined as* $R = \frac{k}{n}$ *D-ary code symbol/source symbol.*

Definition 11 *The error probability of a block code is defined as*

$$P_e = \Pr\{g(f(x_1, x_2, ..., x_n)) \neq (x_1, x_2, .., x_n)\}.$$

Theorem 1 *Asymptotic Equipartition Property (AEP) If* $X_1, X_2, ...$ *are i.i.d. with distribution function* $p(\cdot)$*, then for any* $\varepsilon > 0$,

$$\lim_{n\rightarrow\infty} Pr(| - \frac{1}{n} \log p(x_1, ..., x_n) - H(X)| > \varepsilon) = 0.$$

Definition 12 *Fix* $\varepsilon > 0$*, the typical set* $A_\varepsilon^{(n)}$ *is the set of sequence* $x^n = (x_1, ..., x_n)$ *where* x^n *are i.i.d. from random variable* X *with distribution function* $p(\cdot)$ *such that*

$$A_\varepsilon^{(n)} = \{x^n \in \mathcal{X}^n : 2^{-n(H(x)+\varepsilon)} \leq p(x^n) \leq \ 2^{-n(H(x)-\varepsilon)}\}$$

Theorem 2 *Consequence of AEP: Consider a DMS with distribution function* $p(x)$, $x \in \mathcal{X}$*, then*

1. $Pr(A_\varepsilon^{(n)}) > 1-\varepsilon$ for n sufficiently large.

2. $(1-\varepsilon)2^{n(H(x)-\varepsilon)} \leq |A_\varepsilon^{(n)}| \leq 2^{n(H(x)+\varepsilon)}$ for n sufficiently large.

Theorem 3 *Shannon's Lossless Fixed-Length Source Coding Theorem for Discrete Memoryless Sources(DMS). Consider a DMS with alphabet $\mathcal{X}$, then the following hold.*

Forward part: For any $\varepsilon \in (0,1)$ and any $\delta > 0$, there exists a sequence of D-ary block codes (k,n) for the source such that for n sufficiently large, $\frac{k}{n} \leq H(X)/\log D + \delta$ and $P_e < \varepsilon$.

Converse Part: For any $\varepsilon \in (0,1)$ and any sequence of D-ary block codes (k,n) with $R = \frac{k}{n} < H(X)/\log D$ and sufficiently large n, $P_e > \varepsilon$.

AEP and Shannon's lossless fixed-length source coding theorem also hold for stationary and ergodic source $\{X_i\}_{i=1}^{\infty}$ if we replace entropy $H(X)$ with the entropy rate $H(\mathfrak{X})$.

2.2.2 Slepian-Wolf Coding Theory

Let $(X_1, Y_1), (X_2, Y_2), ...$ be an i.i.d. sequence of jointly distributed random variables X and Y with joint distribution function $p(x,y)$. Assume that X^n and Y^n are encoded separately without knowledge of each other and the compressed outputs are sent to a joint decoder for reconstruction. This problem is called the distributed source coding problem and is illustrated in Fig. 2.1.

Definition 13 *A $(2^{nR_1}, 2^{nR_2}, n)$ distributed source code for the joint source (X,Y) con-*

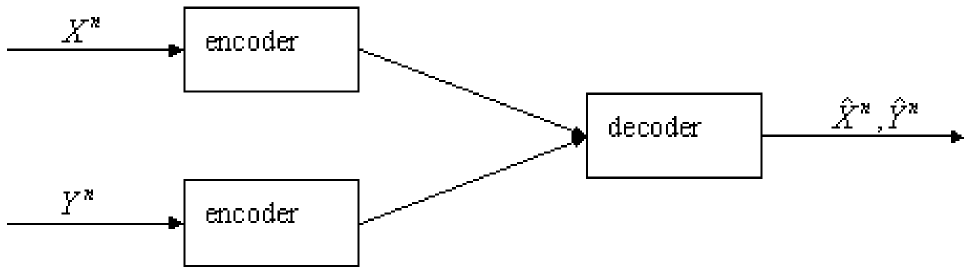

Figure 2.1: Model for DSC with two sources

sists of two encoder maps,

$$f_1 : \mathcal{X}^n \to \{1, 2, ..., 2^{nR_1}\},$$

$$f_2 : \mathcal{Y}^n \to \{1, 2, ..., 2^{nR_2}\}$$

and a decoder map

$$g : \{1, 2, ..., 2^{nR_1}\} \times \{1, 2, ..., 2^{nR_2}\} \to \mathcal{X}^n \times \mathcal{Y}^n,$$

where (R_1, R_2) is called the rate pair of the code.

Definition 14 *The probability of error for a distributed source code is defined as*

$$P_e^{(n)} = P(g(f_1(X^n), f_2(Y^n)) \neq (X^n, Y^n)).$$

Definition 15 *A rate pair (R_1, R_2) is said to be achievable for a source pair $\{(X_i, Y_i)\}_{i=1}^{\infty}$ if there exists a sequence of $(2^{nR_1}, 2^{nR_1}, n)$ distributed source codes with $P_e^{(n)} \to 0$. The achievable region is the closure of the set of achievable rates.*

Theorem 4 *Slepian-Wolf Source Coding Theorem. For the distributed source coding* *problem for the source (X,Y) drawn i.i.d. from $p(x,y)$, the achievable region is given by*

$$R_1 \geq H(X|Y),$$

$$R_2 \geq H(Y|X),$$

$$R_1 + R_2 \geq H(X,Y).$$

The achievable region of Slepian-Wolf theory is depicted in Fig. 2.2.

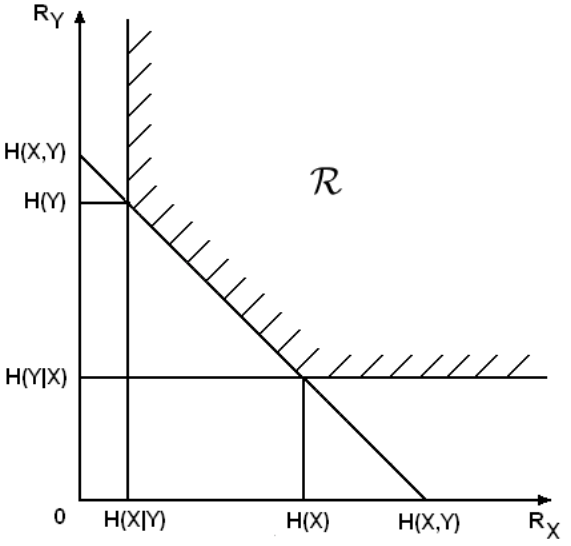

Figure 2.2: Slepian-Wolf achievability region

Cover proved that this theorem also holds for stationary and ergodic sources if we replace entropies with entropy rates and replace conditional entropies with the conditional entropy rates.

Now consider achieving the corner point in the Slepian-Wolf rate region: $R_x = H(X|Y), R_y = H(Y)$. It is already known how to compress Y losslessly at rate $H(Y)$; so we will focus on compressing X at rate $H(X|Y)$ with side information Y only at the decoding end. This is exactly the problem of data compression with side information and the model of this problem is shown in Fig. 2.3.

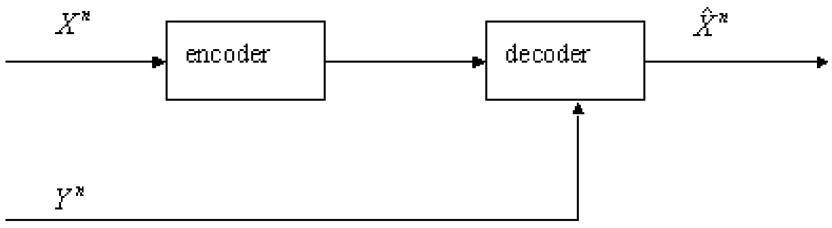

Figure 2.3: Model for data compression with side information

2.3 Channel Coding

2.3.1 Shannon's Channel Coding Theorem

Definition 16 *A discrete channel is characterized by:*

- *A finite input alphabet* $\mathcal{X}$
- *A finite output alphabet* $\mathcal{Y}$
- *n-dimensional conditional distribution:* $p(y^n|x^n) = \Pr(Y^n = y^n|X^n = x^n)$

where $x^n = (x_1, ..., x_n) \in \mathcal{X}^n$ *and* $y^n = (y_1, ..., y_n) \in \mathcal{Y}^n$*, such that* $\sum_{y^n} p(y^n|x^n) = 1$

Definition 17 *Discrete Memoryless Channel(DMC): The DMC is a channel with the property that for all* $n \geq 1$*,*

$$\Pr(Y^n = y^n|X^n = x^n) = \prod_{i=1}^{n} \Pr(Y_i = y_i|X_i = x_i).$$

Definition 18 *Given a discrete memoryless channel, its information capacity is defined by*

$$C = \max_{p(x)} I(X;Y)$$

where the maximum is taken over all input distributions $p(x)$.

Definition 19 *Block codes for discrete channels: Consider a discrete channel with input alphabet $\mathcal{X}$, output alphabet $\mathcal{Y}$ and transition probabilities $\{P(y^n|x^n)\}_{n=1}^{\infty}$. Given a block length n and a set of messages $\{1, 2, ..., M\}$, then an (M, n) block code for the channel consists of:*

- *Encoding function: $f : \{1, 2, ..., M\} \to \mathcal{X}^n$;*
- *Decoding function: $g : \mathcal{X}^n \to \{1, 2, ..., M\}$.*

Definition 20 *The rate of the block code is $R = \frac{\log|M|}{n}$ code symbol/message symbol.*

Definition 21 *Given message i is sent ($i \in \{1, 2, ..., M\}$), the conditional probability of decoding error of an (M, n) code is given by*

$$\lambda_i = \Pr\{g(y^n) \neq i | x^n = f(i)\}.$$

Definition 22 *Average probability of error of an (M, n) code is given by $P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$.*

Definition 23 *Achievable rate. $R > 0$ is said to be achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ codes for the channel such that the average probability of error $P_e^{(n)} \to 0$ as $n \to \infty$. $\lceil \cdot \rceil$ denotes rounding up to the next integer value.*

Theorem 5 *Shannon's Coding Theorem for Discrete Memoryless Channels(DMC).*

Consider a DMC with transition distribution $p(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and information capacity C, then the following hold.

Forward Part: All code rates below C are achievable, i.e.,$\forall \varepsilon > 0$, and any positive $R < C$, $\exists$ a sequence of $(2^{nR}, n)$ codes for the channel with rate R and block length n such that $P_e^{(n)} < \varepsilon$ for n sufficiently large.

Converse Part: If $R > C$, then it is not achievable.

Theorem 6 *[[11]] For a discrete-time binary additive noise channel $Y_i = X_i \bigoplus Z_i$ with stationary and ergodic noise,*

$$C = \lim_{n\to\infty} \max_{p(x)} I(X^n; Y^n) = 1 - H(\mathcal{Z})$$

2.3.2 Channel Models

2.3.2.1 Binary Symmetric Channel (BSC)

The binary symmetric channel is a binary additive noise channel $Y = X \oplus Z$, where $\bigoplus$ is modulo-2 addition and the noise Z is drawn from an i.i.d source and is independent of input X such that $\Pr(Z = 1) = p$. The structure of the BSC is shown in Fig. 2.4.

The capacity of the BSC is $C = 1 - h_b(p)$.

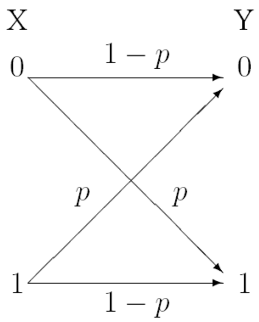

Figure 2.4: BSC transition diagram

2.3.2.2 Binary Markov Noise Channel (BMNC)

The binary Markov noise channel (BMNC) is an additive noise channel $Y_i = X_i \bigoplus Z_i$, where noise $\{Z_i\}_{i=1}^{\infty}$ is a first-order Markov process with transition matrix given by

$$Q = \begin{bmatrix} \varepsilon_Z + (1-\varepsilon_Z)(1-p_Z) & (1-\varepsilon_Z)p_Z \\ (1-\varepsilon_Z)(1-p_Z) & \varepsilon_Z + (1-\varepsilon_Z)p_Z \end{bmatrix}.$$

where p_Z is the channel bit error rate and ε_Z is the correlation coefficient of the noise process. We assume that $0 < \varepsilon_Z \leq 1$ and $0 < p < 1/2$ to ensure that the noise process is irreducible.

The n-fold distribution of the noise if given by

$$P(z^n) = p_Z^{z_1}(1-p_Z)^{1-z_1} \prod_{i=2}^{n} [z_{i-1}\varepsilon_Z + (1-\varepsilon_Z)p_Z]^{z_i}[(1-z_{i-1})\varepsilon_Z + (1-\varepsilon_Z)(1-p_Z)]^{1-z_i}$$

2.3.2.3 Gilbert-Elliott Channel(GEC)

The Gilbert-Elliott channel is also binary additive noise channel but its noise is a hidden Markov process. It has two states: good state G or 0 and bad state B or 1. The sequence of states is a Markov process characterized by the transition matrix P. Each state corresponds to a BSC with a crossover probability P_G or P_B. The model of the GEC is shown in Fig. 2.5.

The transition matrix of the Markov state process is given by

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

Let S_k be the state at time k and Z_k be the noise out at time k. Define the matrix $P(z_k)$, whose ij^{th} entry is given by $Pr(Z_k = z_k, S_k = j | S_{k-1} = i)$, $i, j \in \{0, 1\}$. Then for

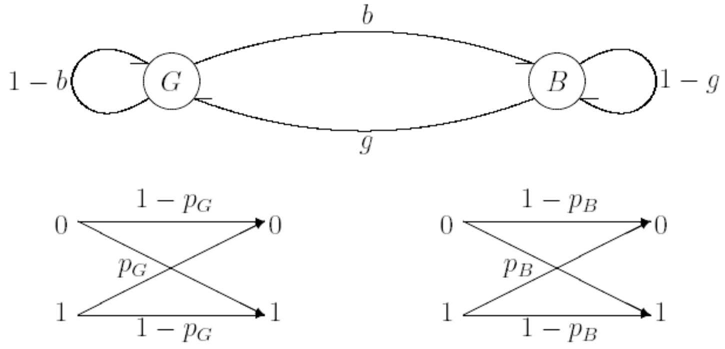

Figure 2.5: GEC model.

the GEC,

$$P(0) = \begin{pmatrix} (1-b)(1-p_1) & b(1-p_2) \\ g(1-p_1) & (1-g)(1-p_2) \end{pmatrix},$$

$$P(1) = \begin{pmatrix} (1-b)p_1 & bp_2 \\ gp_1 & (1-g)p_2 \end{pmatrix}.$$

The n-fold noise distribution is given by

$$P(z^n) = \mathbf{\Pi}^T (\prod_{i=1}^{n} P(z_i))\mathbf{1}$$

where $\mathbf{1}$ is the all-one column and $\mathbf{\Pi}$ is the state stationary distribution given by

$$\mathbf{\Pi} = \left(\frac{g}{b+g}, \frac{b}{b+g}\right).$$

2.3.3 Linear Block Codes

Let GF(q) denotes the Galois field of size q such that $q = p^m$, where p is a prime and m is an integer.

Definition 24 *Let $\mathbb{F} = GF(q)$, a q-ary (n, k) linear code $\mathcal{C}$ is a k-dimensional linear subspace of $\mathbb{F}^n$.*

Definition 25 *Any $k \times n$ matrix whose rows form a basis of $\mathcal{C}$ is called a generator matrix of $\mathcal{C}$ and denoted by $\mathbf{G}$. The parity-check matrix $\mathbf{H}$ of the linear code $\mathcal{C}$, is an $(n - k) \times n$ matrix with the property that $\mathbf{G} \cdot \mathbf{H}^T = \mathbf{0}$.*

Definition 26 *A binary linear code $\mathcal{C}$ is a linear subspace of $\{0, 1\}^n$. The dimension of the code k is the size of the basis of $\mathcal{C}$, which is given by $k = \log |\mathcal{C}|$.*

Definition 27 *Hamming Weight. Let $\mathcal{C}$ be a binary linear code. Then, for any codeword $x^n \triangleq (x_1, ..., x_n) \in \mathcal{C}$, the Hamming weight of x^n, denoted by $w_H(x^n)$, is the number of ones in x^n.*

Definition 28 *Hamming Distance. The Hamming distance between two binary words x^n and y^n in $\{0, 1\}^n$ is the Hamming weight of their difference, i.e., $d_H(x^n, y^n) = w_H(x^n - y^n) = w_H(x^n \bigoplus y^n)$.*

Definition 29 *The minimum distance d of the binary linear code $\mathcal{C}$ is the smallest Hamming weight among its non-zero codewords.*

We also designate a binary linear block code of length n, dimension k and minimum distance d as an (n, k, d) code or simply an (n, k) code if the minimum distance is not of interest.

2.3.4 Decoding Methods for Linear Block Codes

2.3.4.1 MD and ML Decoding

MD and ML decoding are two common decoding methods for linear block codes. MD decoding has less computational complexity than ML decoding but ML decoding is optimal only in terms of minimizing the error probability when channel input is uniformly distributed.

Both MD and ML decoding have large computational complexity, so they can only be applied to decode short-length linear block codes.

MD decoding: Received channel output y^n is decoded into codeword $c_0 \in \mathcal{C}$ if $w(c_0 \bigoplus y^n) \leq w(c \bigoplus y^n)$ for all $c \in \mathcal{C}$. If two or more codewords satisfy the inequality, randomly choose one of them.

SMD decoding: Received channel output y^n is decoded into codeword $c_0 \in \mathcal{C}$ if $w(c_0 \bigoplus y^n) < w(c \bigoplus y^n)$ for all $c \in \mathcal{C}$. If no codeword satisfies the strict inequality, report a decoding failure.

ML decoding: Received channel output y^n is decoded into codeword $c_0 \in \mathcal{C}$ if $\Pr(Y^n = y^n | X^n = c_0) \geq \Pr(Y^n = y^n | X^n = c)$ for all $c \in \mathcal{C}$. If two or more codewords satisfy the inequality, randomly choose one of them.

SML decoding: Received channel output y^n is decoded into codeword $c_0 \in \mathcal{C}$ if $\Pr(Y^n = y^n | X^n = c_0) > \Pr(Y^n = y^n | X^n = c)$ for all $c \in \mathcal{C}$. If no vector satisfies the strict inequality, report a decoding failure.

For any sequence $x^n \in \mathcal{X}^n$, $s = x^n \mathbf{H}^T$ is called the syndrome of x^n. If x^n is a codeword, $s = 0$. The coset indexed by s is defined as the set $\{x^n \in \mathcal{X}^n : x^n\mathbf{H}^T = s\}$. Coset leaders corresponding to MD decoding are the minimum weight vector in each coset.

For MD decoding, we can calculate the syndrome of the channel output first, then decode it to the coset leader indexed by the syndrome. Coset leader is stored in advance to avoid the exhaustive search and reduce the computational complexity.

Definition 30 *Perfect Code. A linear code $\mathcal{C}$ is said to be perfect code if for some non-negative integer t, it has all patterns of Hamming weight t or less and no others as coset leaders.*

Definition 31 *Quasi-Perfect Code. A linear code $\mathcal{C}$ is said to be quasi-perfect if, for some non-negative integer t, it has all patterns of Hamming weight t or less, some of weight $t+1$ and none of greater weight as coset leaders.*

2.3.4.2 MD+ decoding

For a BMNC channel, MD+ decoding is a compromise decoding method between MD decoding and ML decoding [2] . It is based on MD decoding and has more computational complexity but better performance than MD decoding. It mainly deals with the ties among the vectors with minimum Hamming weights in a coset under MD decoding.

For a BMNC channel $Y_i = X_i \bigoplus Z_i$, let $t_{ij}(z^n)$ denote the number of times two consecutive bits in z^n are equal to (i, j), where $i, j \in \{0, 1\}$, then

$$t_{00}(z^n) = \sum_{k=1}^{n-1} (1 - z_k)(1 - z_{k+1})$$

and

$$t_{11}(z^n) = \sum_{k=1}^{n-1} z_k z_{k+1}.$$

MD+ decoding: [**2**] Assume y^n is received at the channel output. Suppose the decoder outputs the codeword c_0 satisfying the MD decoding condition. If there is more than one such codeword, then the decoder chooses c_0 from them that maximizes $t_{00}(c_0 \oplus y^n) + t_{11}(c_0 \oplus y^n)$. If there is still a tie, then the decoder chooses c_0 from the tying codewords that maximizes $t_{11}(c_0 \oplus y^n)$. Finally, if there is still a tie, then the codeword c_0 is picked at random.

The advantage of the MD+ decoding over the ML decoding is the computational complexity. Whereas ML decoding requires an exhaustive search, MD+ can be implemented using syndrome decoding where the coset leaders are chosen according to the MD+ criteria.

2.3.5 Hamming Codes

Hamming codes are perfect codes and can correct one and only one error under MD decoding. The binary r^{th} Hamming code is a $(2^r - 1, 2^r - 1 - r, 3)$ code.

The parity check matrix of a (7,4,3) Hamming code with $r = 3$ is given by

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{bmatrix}.$$

2.3.6 BCH codes

For an (n, k) binary linear block code $\mathcal{C}$, a codeword $c^n = (c_0, c_1, ..., c_{n-1})$ can be represented in a polynomial form $c(x) = c_0 + c_1 x + ... + c_{n-1}x^{n-1}$. A code is cyclic if every cyclic shift of every codeword is also a codeword. The generator polynomial for a cyclic linear block code is the polynomial $g(x)$ with the property that every codeword in $\mathcal{C}$ can be written as the multiplication $a(x)g(x) \mod (x^n - 1)$ for some $a(x)$. For cyclic linear block code, the generator polynomial always exists and is unique.

Definition 32 *If $a \in GF(2^n)$, the Galois field of size 2^n, with the property that the smallest integer l satisfying $a^l = 1$ is $l = 2^n - 1$, then a is called primitive. l is called the order of a.*

Let $a \in GF(2^n)$. Let $\phi(x)$ be a non-zero polynomial over $GF(2) = \{0, 1\}$ with the smallest degree such that $\phi(a) = 0$. Then $\phi(x)$ is called the minimal polynomial of a.

For any $r \geq 3$ and $t < 2^{r-1}$ there exists a t-error-correcting binary BCH code with the properties $n = 2^m - 1$, $k \geq n - rt$ and $d \geq 2t + 1$. The double-error correcting BCH codes are quasi-perfect. The generator polynomial for such code is given by

$$g(x) = LCM\{\phi_1(x), \phi_3(x), ..., \phi_{2t-1}(x)\}$$

where LCM is the least common multiple, $\phi_i(x)$ is the minimal polynomial of a^i and a is a primitive element in $GF(2^n)$.

The minimal polynomial of an $(15, 7)$ BCH code is given by

$$g(x) = x^8 + x^7 + x^6 + x^4 + 1$$

2.3.7 Golay codes

The binary $(23, 12, 7)$ Golay code is the only other nontrivial binary perfect code besides Hamming codes. Due to its rich structure, the Golay code has been used in many communication applications. The $(23, 12, 7)$ Golay code is also a cyclic code, where the generator polynomial is given by

$$g(x) = x^{11} + x^{10} + x^6 + x^5 + x^4 + x^2 + 1.$$

Chapter 3

Syndrome Source Coding (without Side Information)

In this chapter, we study syndrome source coding for lossless data compression without side information. In the remaining part of this chapter we only refer to it as syndrome source coding.

3.1 Syndrome Source Coding Scheme

Consider compressing a source $\{X_i\}_{i=1}^{\infty}$ defined over alphabet $GF(q)$. We associate an additive noise channel $Y_i' = X_i' + Z_i'$ (In this chapter, when '$+$' or '$\times$' is taken between two elements in $GF(q)$, it is the '$+$' or '$\times$' operation in the finite field $GF(q)$, respectively), where the statistics of the noise $\{Z_i'\}_{i=1}^{\infty}$ is same as that of the source $\{X_i\}_{i=1}^{\infty}$. Assume that $\mathcal{C}$ is an q-ary (n, k) linear block code for this associated channel and has parity check matrix $\mathbf{H}$ of size $m \times n$ and the decoder of $\mathcal{C}$ consists of an error pattern estimator whose input is the syndrome of the channel output and whose output is the estimate of the noise. The syndrome source coding scheme employs this code to compress source $\{X_i\}_{i=1}^{\infty}$ as follows:

Encoder: The encoder compresses each source word x^n to the syndrome $s^m = \mathbf{H}x^n$.

Decoder: The decoder decodes s^m using the error pattern estimator of code $\mathcal{C}$.

When the source is binary, stationary and ergodic, if the rate of this code for channel coding is close to the associated channel capacity, the rate of syndrome source coding is close to $H(\mathfrak{X})$. Since

$$R_s = \frac{n-k}{n} = 1 - \frac{k}{n} = 1 - R_c$$

where R_s is the compression rate for syndrome source coding and R_c is the rate for channel coding. For any $\varepsilon > 0$, if $R_c > C - \varepsilon$, by theorem 6, $C = 1 - H(\mathfrak{X})$, thus we have $R_s < H(\mathfrak{X}) + \varepsilon$.

Next we examine the existence of the codes for the syndrome source coding to achieve the source entropy with arbitrary small error probability.

Theorem 7 *(Elias [8]) For DMS $\{X_i\}_{i=1}^{\infty}$ with alphabet $\mathcal{X} = GF(q)$, we use a matrix $\mathbf{H}$ (over $\mathcal{X}$) of size $m \times n$ to compress the source via $s^m = \mathbf{H}x^n$ and use ML decoder: $\hat{x}^n = \arg\max_{x^n}\{\Pr(x^n), \mathbf{H}x^n = s^m\}$. For any $\varepsilon' > \varepsilon > 0$, assume that $\frac{H(X)}{\log|\mathcal{X}|} + \varepsilon' > R > \frac{H(X)}{\log|\mathcal{X}|} + \varepsilon$ and $m = \lceil nR \rceil$, where $|\cdot|$ is the number of elements in the set. If the entries of $\mathbf{H}$ are independent, equiprobable on the alphabet $\mathcal{X}$, then the average block error probability $\bar{P}_e$ (over the ensemble of $\mathbf{H}$) tends to 0 as $n \to \infty$.*

For an stationary and ergodic source, we can prove a similar result.

Theorem 8 *For stationary and ergodic source* $\{X_i\}_{i=1}^{\infty}$ *with alphabet* $\mathcal{X} = GF(q)$*, we use a matrix* $\mathbf{H}$ *(over* $\mathcal{X}$*) of size* $m \times n$ *to compress the source by* $s^m = \mathbf{H}x^n$ *and use ML decoding:* $\hat{x}^n = \arg\max_{x^n}\{\Pr(x^n), \mathbf{H}x^n = s^m\}$*. For any* $\varepsilon' > \varepsilon > 0$*, Assume that* $\frac{H(\mathfrak{X})}{\log|\mathcal{X}|} + \varepsilon' > R > \frac{H(\mathfrak{X})}{\log|\mathcal{X}|} + \varepsilon$ *and* $m = \lceil nR \rceil$*, where* $H(\mathfrak{X})$ *is the entropy rate for the source. If the entries of* $\mathbf{H}$ *are independent, equiprobable on the alphabet* $\mathcal{X}$*, the average block error probability* $\bar{P}_e$ *(over the ensemble of* $\mathbf{H}$*) tends to* 0 *as* $n \to \infty$*.*

Proof:

Let $P_e(\mathbf{H})$ be the block error probability of syndrome source coding based on $\mathbf{H}$ under ML decoding. Then

$$\begin{aligned}
\bar{P}_e &= \sum_{\mathbf{H}} P_{\mathbf{H}}(\mathbf{H}) P_e(\mathbf{H}) \\
&= \sum_{\mathbf{H}} P_{\mathbf{H}}(\mathbf{H}) \sum_{x^n} P_{X^n}(x^n) 1(\hat{x}^n \neq x^n)) \\
&= \sum_{x^n} P_{X^n}(x^n) \sum_{\mathbf{H}} P_{\mathbf{H}}(\mathbf{H}) 1(\hat{x}^n \neq x^n)) \\
&= \sum_{x^n} P_{X^n}(x^n) \underbrace{P_{\mathbf{H}}(\mathbf{H} : \exists x_0^n, \mathbf{H}x_0^n = \mathbf{H}x^n, P_{X^n}(x_0^n) > P_{X^n}(x^n))}_{A}
\end{aligned}$$

where $1(\cdot)$ denotes the indicator function.

For fixed x^n and x_0^n with $P_{X^n}(x_0^n) > P_{X^n}(x^n)$, if $\mathbf{H}x_0^n = \mathbf{H}x^n$, then $\mathbf{H}(x_0^n - x^n) = 0$. Let $(t_1, t_2, ..., t_n) = x_0^n - x^n$. Assume that $t_j \neq 0$, for each row of $\mathbf{H}$, $(h_{i1}, h_{i2}, ..., h_{in})$, we have $h_{ij}t_j = -h_{i1}t_1 - .... - h_{i(j-1)}t_{j-1} - h_{i(j+1)}t_{j+1} - ... - h_{in}t_n$.

We have $|\mathcal{X}|^{(n-1)}$ ways to choose $h_{i1}, ...h_{i(j-1)}, h_{i(j+1)}, ..., h_{in}$, then h_{ij} is determined.

Thus we have $|\mathcal{X}|^{m(n-1)}$ different matrices $\mathbf{H}$ such that $\mathbf{H}(x_0^n - x^n) = 0$.

Then

$$\begin{aligned} A &< \frac{|\{x_0^n : P_{X^n}(x_0^n) > P_{X^n}(x^n)\}| \cdot |\mathcal{X}|^{m(n-1)}}{|\mathcal{X}|^{mn}} \\ &= |\{x_0^n : P_{X^n}(x_0^n) > P_{X^n}(x^n)\}| \cdot |\mathcal{X}|^{-m}. \end{aligned}$$

For each x^n in the typical set T, $P_{X^n}(x^n) \geq |\mathcal{X}|^{-n\left(\frac{H(p)}{\log|\mathcal{X}|}+\delta\right)}$ for $\delta < \varepsilon$ and sufficiently large n. Hence $|\{x_0^n : P_{X^n}(x_0^n) > P_{X^n}(x^n)\}| \leq |\mathcal{X}|^{n\left(\frac{H(X)}{\log|X|}+\delta\right)}$, and

$$\bar{P}_e < \sum_{x^n \in T} P_{X^n}(x^n) \underbrace{|\mathcal{X}|^{-m+n\left(\frac{H(X)}{\log|\mathcal{X}|}+\delta\right)}}_{B} + \sum_{x^n \notin T} P_{X^n}(x^n).$$

Since $\frac{m}{n} > \frac{H(X)}{\log|\mathcal{X}|} + \varepsilon$, we obtain that $B \to 0$ as $n \to \infty$. Thus $P_e \to 0$ as $n \to \infty$.

□

Note: This theorem also holds for non-stationary homogeneous irreducible Markov sources since the AEP also holds for such Markov sources.

Corollary 1 *For a stationary and ergodic source $\{X_i\}_{i=1}^{\infty}$ with finite field alphabet $\mathcal{X}$, we can find a linear code for syndrome source coding under ML decoding such that the compression rate is arbitrarily close to $H(\mathfrak{X})$ and the error probability is arbitrarily small.*

Proof. For any $\varepsilon' > 0$ and $\delta > 0$, choose ε and R such that $0 < \varepsilon < \varepsilon'$ and $\frac{H(\mathfrak{X})}{\log|\mathcal{X}|} + \varepsilon' > R > \frac{H(\mathfrak{X})}{\log|\mathcal{X}|} + \varepsilon$. Let $m = \lceil nR \rceil$.

By theorem 8, The average error probability of the syndrome source coding scheme is less than δ over the ensemble of H when n is sufficiently large, then there exists a special $\mathbf{H}$ such that the error probability for the syndrome source coding based on this matrix is less than δ, while the rate R is less than ε.

□

3.2 Syndrome Source Coding Using Short-Length Block Codes

In this section, we consider employing short-length perfect and quasi-perfect codes for syndrome source coding for binary sources to reduce the system delay. We have the following decoding methods for syndrome source coding:

MD decoding: A received s^m is decoded into $\hat{x}^n$ in the coset indexed by s^m if $w(\hat{x}^n) \leq w(x^n)$ for all x^n in the coset indexed by s^m. If two or more vectors satisfy the inequality, randomly choose one of them.

SMD decoding: A received s^m is decoded into $\hat{x}^n$ in the coset indexed by s^m if $w(\hat{x}^n) < w(x^n)$ for all x^n in the coset indexed by s^m. If no vector satisfies the strict inequality, report a decoding failure.

ML decoding: A received s^m is decoded into $\hat{x}^n$ in the coset indexed by s^m if $Pr(\hat{x}^n) \geq Pr(x^n)$ for all x^n in the coset indexed by s^m. If two or more vectors satisfy the inequality, randomly choose one of them.

SML decoding: A received s^m is decoded into $\hat{x}^n$ in the coset indexed by s^m if $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m. If no vector satisfies the strict inequality, report a decoding failure.

For Markov sources $\{X_i\}_{i=1}^{\infty}$, we also have the MD+ decoding method.

MD+ decoding: s^m is decoded into $\hat{x}^n$ in the coset indexed by s^m if $w(\hat{x}^n) \leq w(x^n)$ for all x^n in the coset indexed by s^m. If two or more vectors satisfy the inequality, the decoder chooses from them the $\hat{x}^n$ that maximizes $t_{00}(x^n) + t_{11}(x^n)$. If there is still a tie, then the decoder chooses from these vectors in the tie the $\hat{x}^n$ that maximizes $t_{11}(x^n)$. Finally, if there is still a tie, then $\hat{x}^n$ is picked from the vectors in the tie at random.

ML decoding is the optimal decoding method in terms of minimizing the error probability.

Next we will explore the relationship among these decoding methods for binary Markov sources.

Lemma 1 *[1][2] For a binary stationary (first order) Markov process $\{X_i\}_{i=1}^{\infty}$ with $Pr(X_i = 1) = p$ and correlation coefficient ε, suppose that*

$$t^* = \frac{\ln\left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p}\right] + \ln\left[\frac{1-p}{p}\right]}{\ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)}\right] + \ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)p}\right]}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}$$

Let x^n be a sequence such that $w(x^n) \leq \min\{t^*, n/2\}$. Then if $w(\bar{x}^n) > w(x^n)$, $\Pr(\bar{x}^n) < \Pr(x^n)$.

Theorem 9 *Let $\mathcal{C}$ be an (n, M, d) perfect code to be used for syndrome source coding for a binary stationary (first order) Markov source $\{X_i\}_{i=1}^{\infty}$ with $Pr(X_i = 1) = p$ and correlation coefficient ε. Assume that*

$$\left\lfloor \frac{d-1}{2} \right\rfloor < \frac{\ln\left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p}\right] + \ln\left[\frac{1-p}{p}\right]}{\ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)}\right] + \ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)p}\right]} \tag{3.1}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)} \tag{3.2}$$

Then MD, SMD, ML and SML decoding are equivalent.

Proof. For perfect codes, the leader with minimum weight in each coset is unique, and the weight of the coset leader is less than $\lfloor \frac{d-1}{2} \rfloor \leq n/2$.

For received syndrome s^m, the MD decoding and SMD decoding result $\hat{x}^n$ is the leader of the coset indexed by s^m where $w(\hat{x}^n) < w(x^n)$ for all x^n in the coset indexed by s^m. By Lemma 1, $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m, this $\hat{x}^n$ is also the decoding result for ML decoding and SML decoding.

□

Theorem 10 *Let $\mathcal{C}$ be an (n, M) quasi-perfect code to be used for syndrome source coding for a binary stationary (first order) Markov source $\{X_i\}_{i=1}^{\infty}$. Assume that*

$$\left\lfloor \frac{d-1}{2} \right\rfloor + 1 < \frac{\ln\left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p}\right] + \ln\left[\frac{1-p}{p}\right]}{\ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)}\right] + \ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)p}\right]}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}$$

.

Then for the syndrome s^m, the following hold.

(a) If there exists $\hat{x}^n$ in the coset indexed by s^m such that $w(\hat{x}^n) < w(x^n)$ for all x^n in the coset indexed by s^m, then $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m.

(b) If there exists $\hat{x}^n$ in the coset indexed by s^m such that $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m, then $w(\hat{x}^n) \leq w(x^n)$ for all x^n in the coset indexed by s^m.

Proof: (a) for quasi-perfect codes, the weight of the leader with minimum weight in each coset is less than $\lfloor \frac{d-1}{2} \rfloor + 1 \leq n/2$.

For a received syndrome s^m, if $\exists \hat{x}^n$ in the coset indexed by s^m such that $w(\hat{x}^n) < w(x^n)$ for all x^n in the coset indexed by s^m, this $\hat{x}^n$ must be the leader in this coset, so its weight is less than $\left\lfloor \frac{d-1}{2} \right\rfloor + 1 \leq n/2$. By Lemma 1 , $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m.

(b) If $\exists \hat{x}^n$ such that $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m, then if $w(\hat{x}^n) \leq w(x^n)$ doesn't hold for all x^n in the coset indexed by s^m , let $\bar{x}^n$ be the leader in the coset indexed by s^m such that $w(\bar{x}^n) < w(\hat{x}^n)$. By Lemma 1 we have $\Pr(\bar{x}^n) > \Pr(\hat{x}^n)$, contradicting $\Pr(\hat{x}^n) > \Pr(x^n)$ for all x^n in the coset indexed by s^m. Then $w(\hat{x}^n) \leq w(x^n)$ for all x^n in the coset indexed by s^m.

□

3.3 Simulation Results

Throughout the simulations, we will measure the performance of the different source coding schemes in terms of probability of codeword error on frame error rate (FER).

Simulation results are shown in Fig. 3.1 - 3.6, where we use (15,7) BCH, (23,12) Golay and (7,4) Hamming codes, under MD, MD+ and ML decoding to compress a binary Markov source $\{X_i\}_{i=1}^{\infty}$, with $Pr(X_i = 1) = p$ and correlation coefficient is ε.

Fig. 3.1 and Fig. 3.2 indicate that for the $(7,4)$ Hamming code, MD, MD+ and ML decoding are identical for the cases $\varepsilon = 0.1$ and $\varepsilon = 0.25$, which is expected by Theorem 9 since the (7,4) Hamming code is perfect code and equations (3.1) and (3.2) are satisfied.

In Fig. 3.3 and Fig. 3.4, for the $(15, 7)$ BCH code, MD+ decoding improves on MD decoding. By comparing simulations with $\varepsilon = 0.1$ and $\varepsilon = 0.25$, it shows that when source correlation is smaller, MD+ decoding performs closer to ML decoding. In Fig. 3.3, MD+ decoding is very close to ML decoding when $0.05 < p < 0.1$.

Fig. 3.5 and Fig. 3.6 show that for the $(23, 12)$ Golay code, when MD+ decoding is implemented it does not show any improvement over MD decoding since the (23,12) Golay code is a perfect code and there are no ties in MD decoding for perfect code. Because $d = 7$, 3.1 does not hold for parameters used in these two simulations, so MD and ML decoding are not identical.

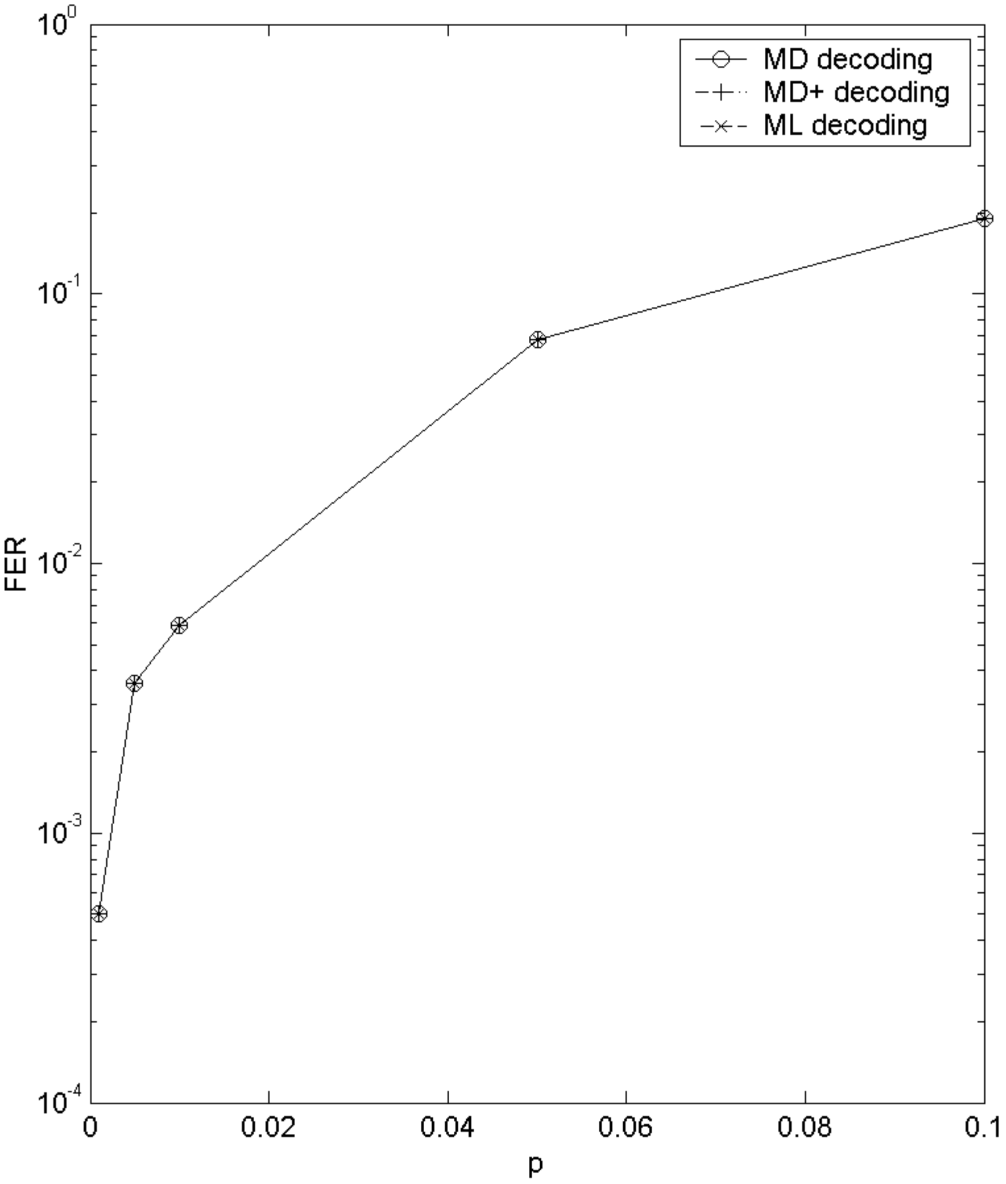

Figure 3.1: FER vs. p under different decoding methods for syndrome source coding based on (7,4) Hamming code for Markov sources with correlation coefficient $\varepsilon = 0.1$

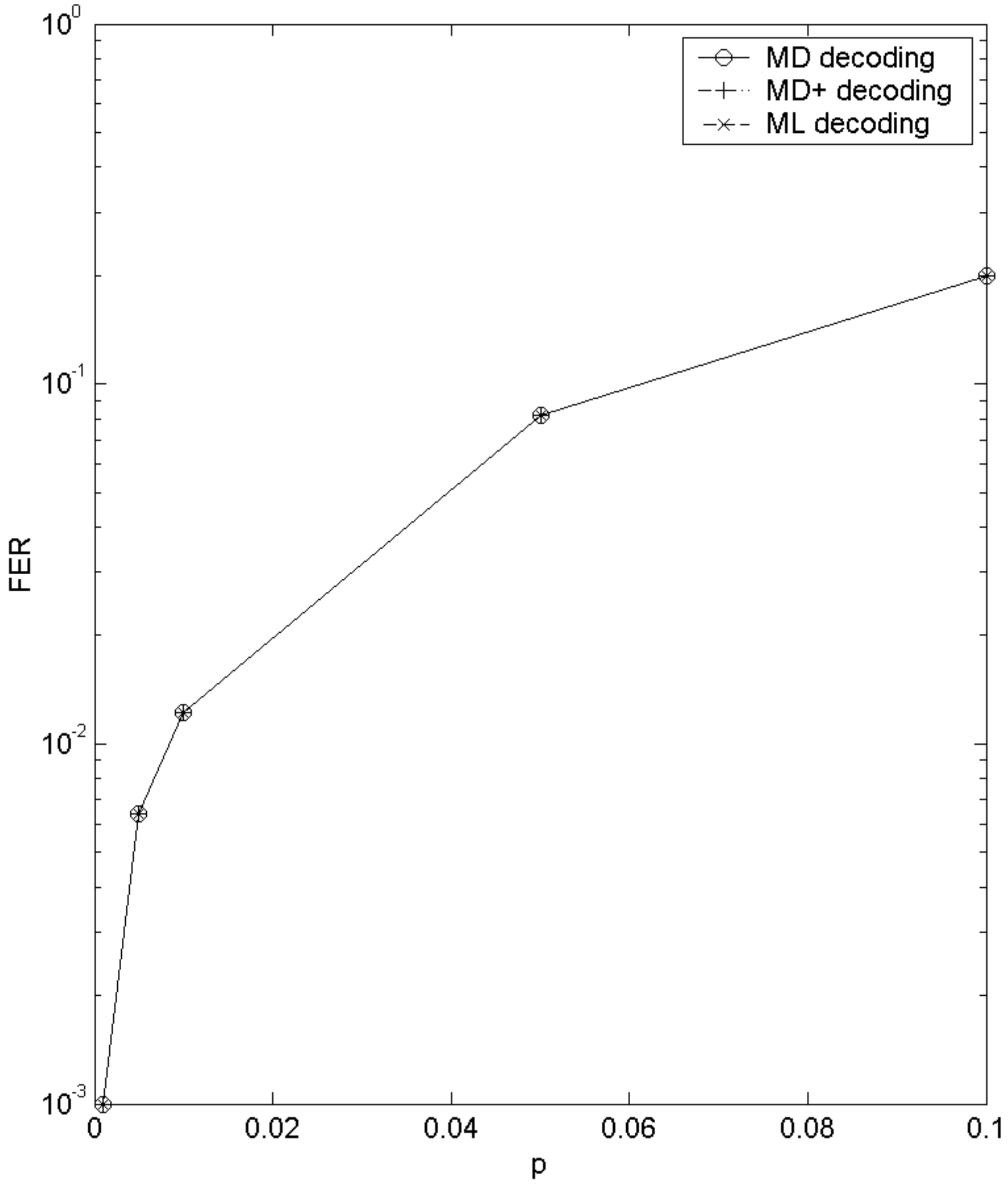

Figure 3.2: FER vs. p for syndrome source coding based on (7,4) Hamming code for Markov sources with correlation coefficient $\varepsilon = 0.25$

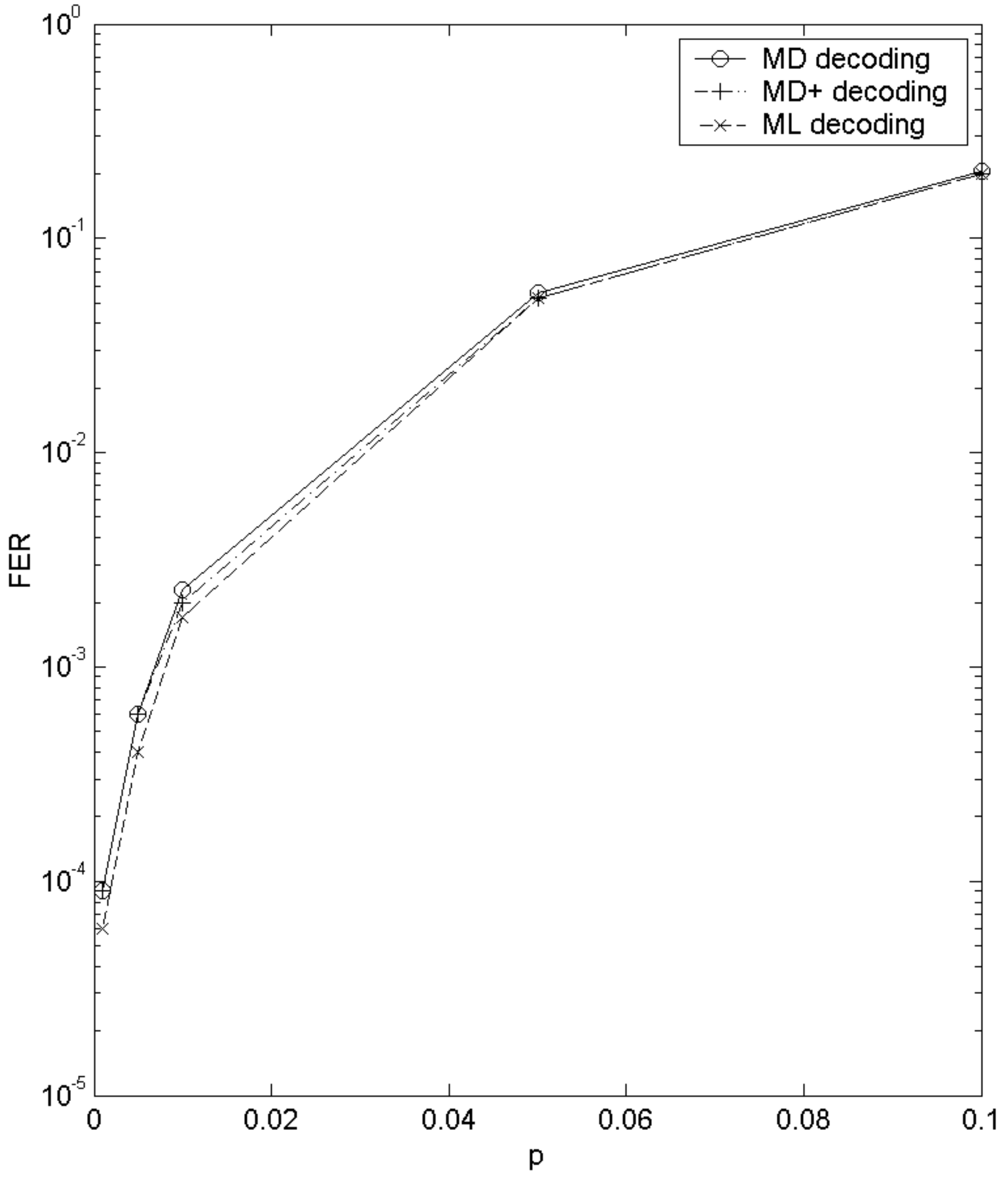

Figure 3.3: FER vs. p for syndrome source coding based on (15,7) BCH code for Markov sources with correlation coefficient $\varepsilon = 0.1$

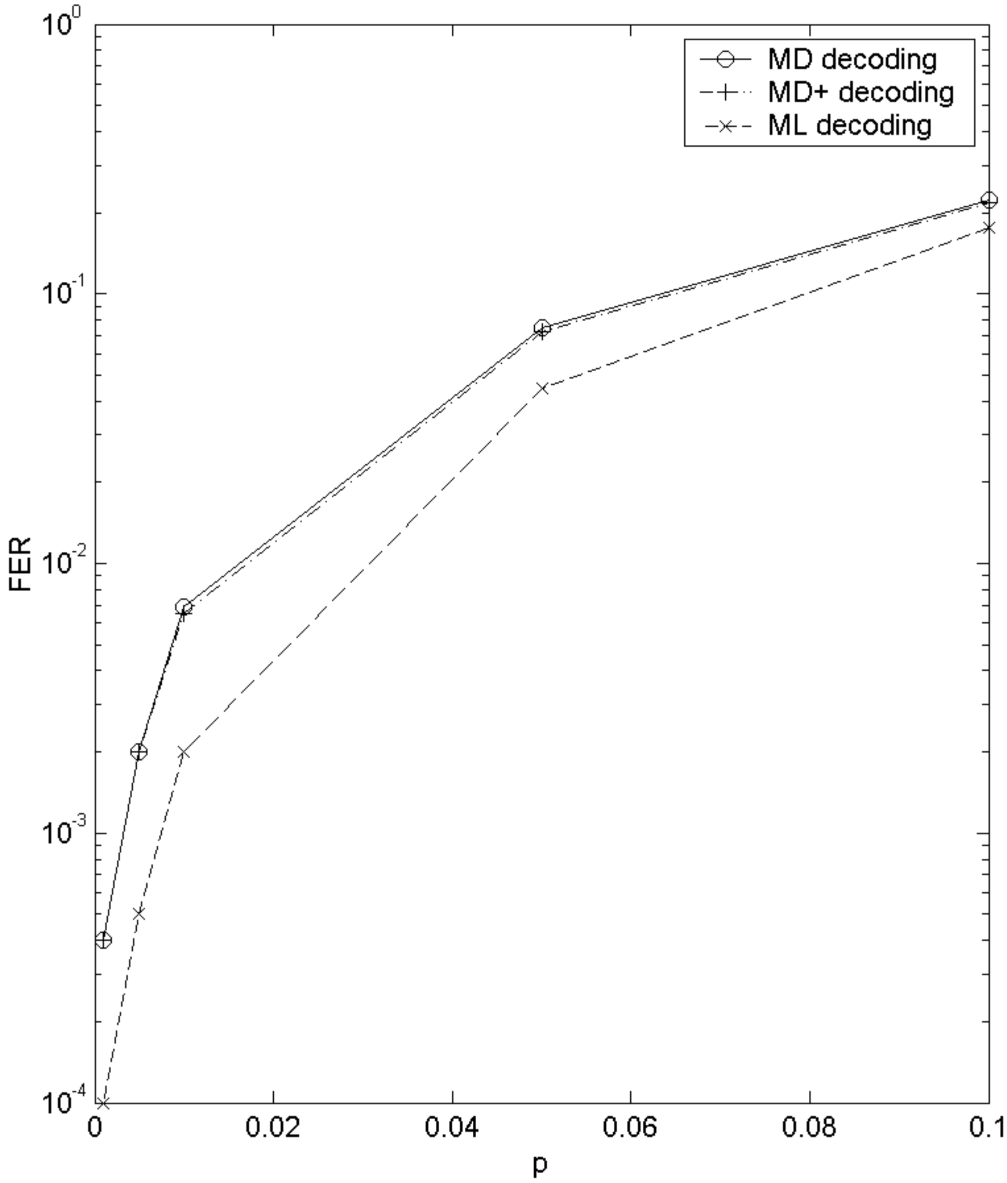

Figure 3.4: FER vs. p for syndrome source coding based on (15,7) BCH code for Markov sources with correlation coefficient $\varepsilon = 0.25$

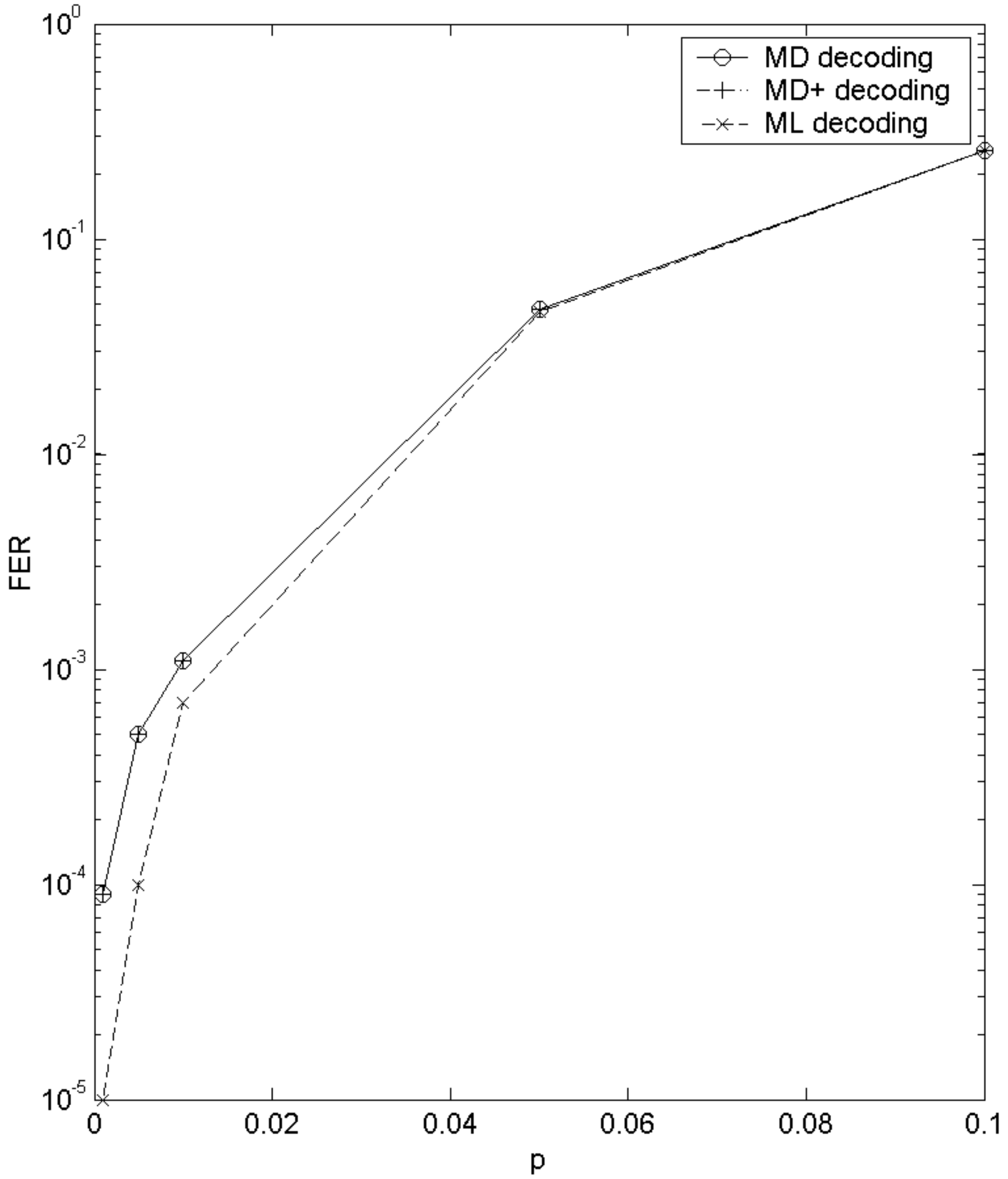

Figure 3.5: FER vs. p for syndrome source coding based on (23,12) Golay code for Markov sources with correlation coefficient $\varepsilon = 0.1$

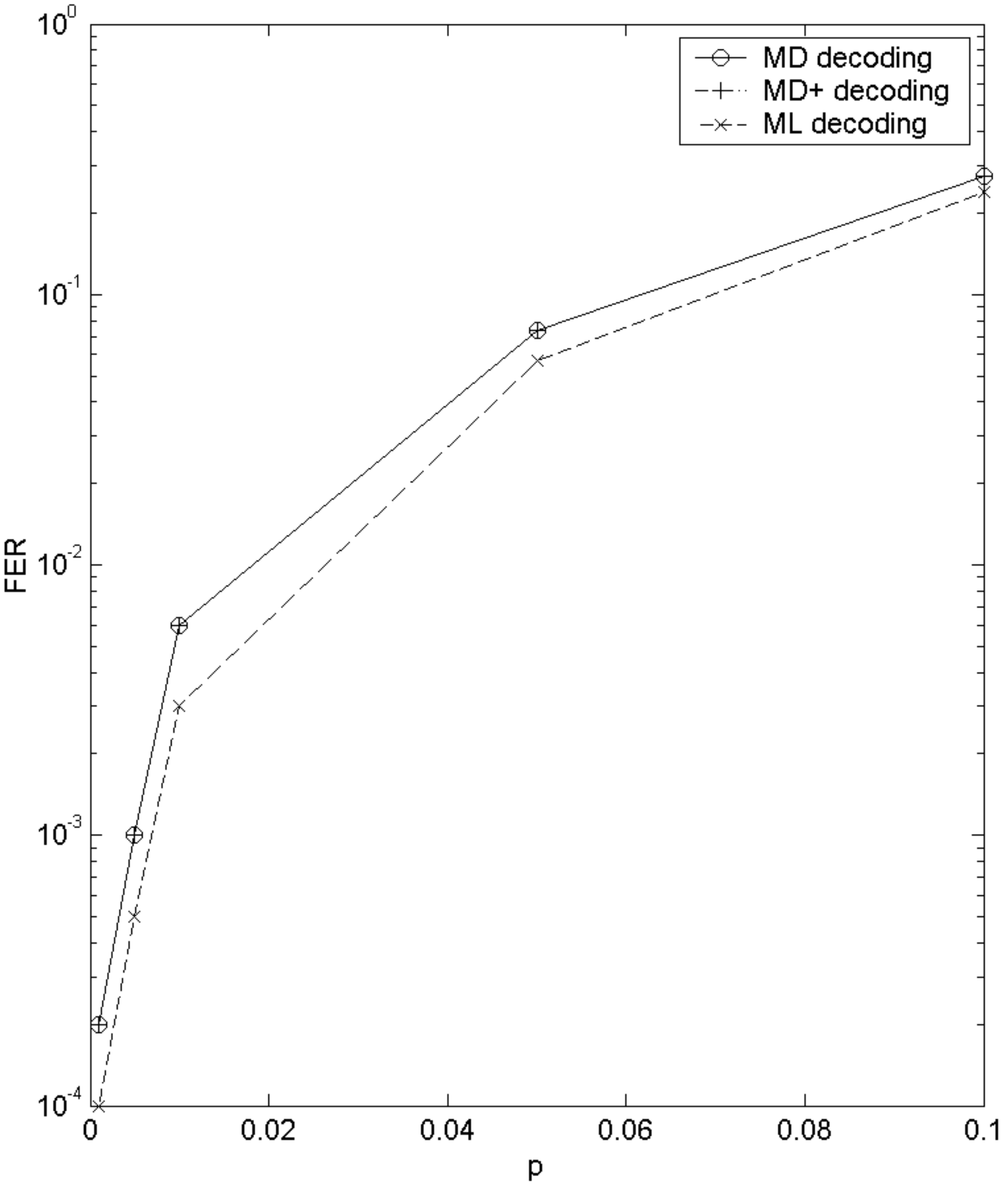

Figure 3.6: FER vs. p for syndrome source coding based on (23,12) Golay code for Markov sources with correlation coefficient $\varepsilon = 0.25$

Chapter 4

Syndrome Source Coding with Side Information

4.1 Syndrome Source Coding Scheme with Side Information

For source pairs $\{(X_i, Y_i)\}_{i=1}^{\infty}$, where X_i and Y_i are defined over alphabet $GF(q)$, assume that the correlation between $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$ can be modeled as $X_i = Y_i + Z_i$, where $\{Z_i\}_{i=1}^{\infty}$ defined over alphabet $GF(q)$ is independent from $\{Y_i\}_{i=1}^{\infty}$, stationary and ergodic. We compress $\{X_i\}_{i=1}^{\infty}$ with $\{Y_i\}_{i=1}^{\infty}$ as side information only at the decoder.

For a linear block code $\mathcal{C}$ with parity check matrix $\mathbf{H}$ defined over alphabet $GF(q)$, assume the decoder of $\mathcal{C}$ consists of an error pattern estimator whose input is the syndrome of the channel output and whose output is the estimate of the noise, Wyner's syndrome source coding scheme to compress $\{X_i\}_{i=1}^{\infty}$ with $\{Y_i\}_{i=1}^{\infty}$ as side information is as follows: (In this chapter, when '+' or '×' is taken between two elements in $GF(q)$, it is the '+' or '×' operation in the finite field $GF(q)$, respectively; for a and b in $GF(q)$, $a - b$ denotes $a + b^{-1}$, where b^{-1} is the inverse of b in the finite field $GF(q)$.)

Encoder: Compress x^n to the syndrome $s_X^m = \mathbf{H}x^n$.

Decoder: First calculate the syndrome of z^n: $s_Z^m = s_X^m - \mathbf{H}y^n$. Then we can estimate z^n by the error pattern estimator of $\mathcal{C}$. Finally we estimate x^n via $\hat{x}^n = y^n + \hat{z}^n$.

When the channel is a binary channel (X_i, Y_i and Z_i are all over $GF(2)$), and noise in the correlation channel is stationary and ergodic, if the rate of this code for the channel coding is close to the correlation channel capacity, the rate of syndrome source coding is close to $H(\mathfrak{X}|\mathfrak{Y})$. Since

$$R_s = \frac{n-k}{n} = 1 - \frac{k}{n} = 1 - R_c$$

where R_s is the compression rate for syndrome source coding and R_c is the rate for channel coding. For any $\varepsilon > 0$, if $R_c > C - \varepsilon$, by theorem 6, $C = 1 - H(\mathfrak{Z})$, hence we have $R_s < H(\mathfrak{Z}) + \varepsilon = H(\mathfrak{X}|\mathfrak{Y}) + \varepsilon$.

The following corollary shows the existence of syndrome source coding to achieve the Slepian-Wolf limit.

Corollary 2 *For two sources $\mathfrak{X} = \{X_i\}_{i=1}^{\infty}$ defined over alphabet $GF(q)$ and $\mathfrak{Y} = \{Y_i\}_{i=1}^{\infty}$ defined over alphabet $GF(q)$, if the correlation between them is modeled as an additive noise channel $X_i = Y_i + Z_i$ where the noise $\mathfrak{Z} = \{Z_i\}_{i=1}^{\infty}$ defined over alphabet $GF(q)$ is stationary and ergodic, and is independent of $\mathfrak{Y}$, we can find a q-ary linear code for*

Wyner's syndrome source coding scheme such that the compression rate of $\mathfrak{X}$ is arbitrarily close to $H(\mathfrak{X}|\mathfrak{Y})$ and the error probability is arbitrarily small.

Proof: By Theorem 8, for $\{Z_i\}_{i=1}^{\infty}$ we can find an $m \times n$ matrix $\mathbf{H}$ to compress z^n to the syndrome $\mathbf{H}z^n$ under ML decoding, such that $\frac{m}{n}$ is arbitrarily close to $H(\mathfrak{Z})$ and the error probability is arbitrarily small for n sufficiently large.

Then we use this $\mathbf{H}$ for Wyner's syndrome source coding scheme. At the decoder, after we calculate s_Z^m, decode s_Z^m using the ML decoding: $\hat{z^n}$ is the vector in the coset indexed by s_Z^m which maximizes $\Pr(z^n)$. Finally let $\hat{x}^n = y^n + \hat{z^n}$.

Then $Pr(\hat{x}^n \neq x^n) = Pr(\hat{z}^n \neq z^n)$, which is arbitrarily small, and the rate is arbitrarily close to $H(\mathfrak{Z}) = H(\mathfrak{X}|\mathfrak{Y})$ when n is sufficiently large.

□

4.2 Syndrome Source Coding with Side Information Using Short-Length Block Codes

In this section, we consider using short-length perfect and quasi-perfect codes for binary syndrome source coding with side information to avoid the significant delay that can be caused by long-length block codes such as LDPC codes. This low-complexity solution comes at the price that the code rate may not be close to the Slepian-Wolf limit. For these short-length codes, the following decoding methods are provided to decode short-length linear block codes. We assume a side information correlation model given by

$X_i = Y_i \bigoplus Z_i$.

MD decoding: For received s_X^m and y^n, first calculate the syndrome of z^n: $s_Z^m = s_X^m \bigoplus \mathbf{H} y^n$. Then estimate z^n: s_Z^m is decoded into $\hat{z}^n$ in the coset indexed by s_Z^m if $w(\hat{z}^n) \leq w(z^n)$ for all z^n in the coset indexed by s_Z^m. If two or more vectors satisfy the inequality, randomly choose one of them. Finally estimate x^n by $\hat{x}^n = y^n \bigoplus \hat{z}^n$.

SMD decoding: For received s_X^m and y^n, first calculate the syndrome of z^n: $s_Z^m = s_X^m \bigoplus \mathbf{H} y^n$. Then estimate z^n: s_Z^m is decoded into $\hat{z}^n$ in the coset indexed by s_Z^m if $w(\hat{z}^n) < w(z^n)$ for all z^n in the coset indexed by s_Z^m. If no vector satisfies the strict inequality, report a decoding failure. Finally estimate x^n by $\hat{x}^n = y^n \bigoplus \hat{z}^n$ if there is no decoding failure.

ML decoding: For received s_X^m and y^n, first calculate the syndrome of z^n: $s_Z^m = s_X^m \bigoplus \mathbf{H} y^n$. Then estimate z^n: s_Z^m is decoded into $\hat{z}^n$ in the coset indexed by s_Z^m if $Pr(\hat{z}^n) \geq Pr(z^n)$ for all z^n in the coset indexed by s_Z^m. If two or more vectors satisfy the inequality, randomly choose one. Finally estimate x^n by $\hat{x}^n = y^n \bigoplus \hat{z}^n$.

SML decoding: For received s_X^m and y^n, first calculate the syndrome of z^n: $s_Z^m = s_X^m \bigoplus \mathbf{H} y^n$. Then estimate z^n: s_Z^m is decoded into $\hat{z}^n$ in the coset indexed by s_Z^m if $\Pr(\hat{z}^n) > \Pr(z^n)$ for all z^n in the coset indexed by s_Z^m. If no vector satisfies the strict inequality, report a decoding failure. Finally estimate x^n by $\hat{x}^n = y^n \bigoplus \hat{z}^n$ if there is no decoding failure.

Obviously the ML decoding method is optimal in terms of minimizing the error probability since the optimal decoding maximizes $Pr(x^n|y^n, s_X)$, and

$$\begin{aligned} & P(x^n|y^n, s_X) \\ = \; & P(x^n|y^n, s_Z) \\ = \; & P(z^n|s_Z) \end{aligned}$$

For sources with binary Markov noise correlation channels, the MD+ decoding is formulated as follows:

MD+ decoding: For received s_X^m and y^n, first calculate the syndrome of z^n: $s_Z^m = s_X^m \bigoplus \mathbf{H}y^n$. Then estimate z^n: s^m is decoded into $\hat{z}^n$ in the coset indexed by s_Z^m if $w(\hat{z}^n) \leq w(z^n)$ for all z^n in the coset indexed by s_Z^m. If two or more vectors satisfy the inequality, the decoder chooses from them the $\hat{z}^n$ that maximizes $t_{00}(z^n) + t_{11}(z^n)$. If there is still a tie, then the decoder chooses from the tying vectors the $\hat{z}^n$ that maximizes $t_{11}(z^n)$. If there is still a tie, then $\hat{z}^n$ is picked from the tying vectors at random. Finally estimate x^n by $\hat{x}^n = y^n \bigoplus \hat{z}^n$.

For these decoders, we can store the coset leaders in advance and the computing complexity will be substantially lower than that of using exhaustive search.

Next we examine the relationship among these decoding methods.

Theorem 11 *For two binary sources $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$, assume that the correlation channel between $\{X_i\}_{i=1}^{\infty}$ and side information $\{Y_i\}_{i=1}^{\infty}$ is $X_i = Y_i \bigoplus Z_i$ where $\{Z_i\}_{i=1}^{\infty}$ is a binary stationary Markov process and independent of input $\{Y_i\}_{i=1}^{\infty}$. The distribution of Y^n can be arbitrary. Let $\mathcal{C}$ be an (n, M, d) perfect code to be used on syndrome source* *coding for* $\{X_i\}_{i=1}^{\infty}$ *with side information* $\{Y_i\}_{i=1}^{\infty}$*. Assume that*

$$\left\lfloor \frac{d-1}{2} \right\rfloor < \frac{\ln\left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p}\right] + \ln\left[\frac{1-p}{p}\right]}{\ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)}\right] + \ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)p}\right]} \tag{4.1}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)} \tag{4.2}$$

Then MD, SMD, ML and SML decoding are equivalent.

The proof is similar to that of Theorem 9.

Theorem 12 *For two binary sources* $\{X_i\}_{i=1}^{\infty}$ *and* $\{Y_i\}_{i=1}^{\infty}$*, assume the correlation channel between* $\{X_i\}_{i=1}^{\infty}$ *and side information* $\{Y_i\}_{i=1}^{\infty}$ *is* $X_i = Y_i \bigoplus Z_i$ *where* $\{Z_i\}_{i=1}^{\infty}$ *is binary stationary Markov process and is independent of input* $\{Y_i\}_{i=1}^{\infty}$*. The distribution of* Y^n *can be arbitrary. Let* $\mathcal{C}$ *be an* (n, M, d) *quasi-perfect code to be used for syndrome source coding for* $\{X_i\}_{i=1}^{\infty}$ *with side information* $\{Y_i\}_{i=1}^{\infty}$*. Assume that*

$$\left\lfloor \frac{d-1}{2} \right\rfloor + 1 < \frac{\ln\left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p}\right] + \ln\left[\frac{1-p}{p}\right]}{\ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)}\right] + \ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)p}\right]}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}$$

If $\hat{x}^n$ *is the decoding output of SMD (not the decoding failure), it's also the decoding output of SML.*

The proof is similar to that of Theorem 10.

4.3 Syndrome Source Coding Scheme for Model $Y_i = X_i + Z_i$

For two sources $\{X_i\}_{i=1}^{\infty}$ and $\{Y_i\}_{i=1}^{\infty}$, since Y^n is known at the decoder and decoder needs to estimate x^n, it is more natural to model the correlation channel as an additive noise channel $Y_i = X_i + Z_i$ where $\{Z_i\}_{i=1}^{\infty}$ defined over alphabet $GF(q)$ is independent of $\{X_i\}_{i=1}^{\infty}$. Based on a linear block code $\mathcal{C}$ with parity check matrix $\mathbf{H}$, for the correlation channel $Y_i = X_i + Z_i$, Wyner's scheme can be adjusted as follows.

Encoder: Compress x^n to the syndrome $s_X^m = \mathbf{H}x^n$.

Decoder: First calculate the syndrome of z^n: $s_Z^m = \mathbf{H}y^n - s_X^m$. Next estimate z^n by the error pattern estimator of $\mathcal{C}$. Finally, estimate x^n by $\hat{x}^n = y^n - \hat{z}^n$.

However, we may get some rate loss with this correlation model and may not achieve the Slepian-Wolf limit. Since $Y^n = X^n + Z^n$, by [5], $H(Y^n) \geq H(X^n)$; then $H(\mathfrak{Y}) \geq H(\mathfrak{X})$. Thus we have that $H(\mathfrak{X}|\mathfrak{Y}) = H(\mathfrak{X}, \mathfrak{Y}) - H(\mathfrak{Y}) \leq H(\mathfrak{X}, \mathfrak{Y}) - H(\mathfrak{X}) = H(\mathfrak{Y}|\mathfrak{X})$. Using this model and syndrome source coding, we can only achieve a rate of $H(\mathfrak{Y}|\mathfrak{X})$, which is greater than or equal to $H(\mathfrak{X}|\mathfrak{Y})$.

Next we consider applying perfect and quasi-perfect codes for syndrome source coding based on this new model to compress binary sources. For this new model, MD, SMD, ML, SML, MD+ are the same as described in Section 4.2. Moreover we propose an optimal decoding method for this model.

MAP decoding: For received s_X^m and y^n, decode them into $\hat{x}^n$ in the coset indexed by s_X^m which maximize $P(x^n)P(z^n)$, where $z^n = y^n \bigoplus x^n$.

The MAP decoding method is optimal. Since for any given y^n and s_X^m, the optimal decoder finds $\hat{X}$ to maximize $P(\hat{x}^n|y^n, s_X)$,

$$\begin{aligned} P(x^n|y^n, s_X) &= \frac{P(x^n, y^n, s_X)}{P(y^n, s_X)} \\ &= \frac{P(x^n)P(y^n|x^n)P(s_X|x^n, y^n)}{P(y^n, s_X)} \\ &= \frac{P(x^n)P(z^n)P(s_X|x^n)}{P(y^n, s_X)} \end{aligned}$$

If x^n is i.i.d. and uniformly distributed maximizing $P(x^n|y^n, s_X)$ is equivalent to maximizing $P(z^n)$ in coset indexed by s_Z, and hence ML decoding is optimal.

However, usually x^n is not uniformly distributed. In this case we can search all x^n in the coset indexed by s_X^m to find $\hat{x}_n$ which maximize $P(x^n)P(z^n)$, $z^n = y^n \bigoplus x^n$ to maximizes $P(x^n|y^n, s_X)$. This is MAP decoding.

For this new model we have similar results as Theorem 11 and Theorem 12.

Let the noise $\{Z_i\}_{i=1}^{\infty}$ is binary stationary Markov process which has channel bit error rate p and correlation coefficient ε.

For a (n, M, d) linear block code $\mathcal{C}$, assume

$$\left\lfloor \frac{d-1}{2} \right\rfloor < \frac{\ln\left[\frac{\varepsilon+(1-\varepsilon)p}{(1-\varepsilon)p}\right] + \ln\left[\frac{1-p}{p}\right]}{\ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)(1-p)}\right] + \ln\left[\frac{\varepsilon+(1-\varepsilon)(1-p)}{(1-\varepsilon)p}\right]} \tag{4.3}$$

and

$$0 < \varepsilon < \frac{1-2p}{2(1-p)}$$

If $\mathcal{C}$ is perfect code, then MD, SMD, ML and SML decoding are equivalent.

If $\mathcal{C}$ is quasi-perfect code, then the decoding output (not decoding failure) of SMD is also the decoding output of SML.

4.4 Simulation Results

For the simulations shown in Fig. 4.1 - 4.5, the correlation between $\mathfrak{X}$ and $\mathfrak{Y}$ is modeled as an BMNC $X_i = Y_i + Z_i$. The noise $\mathfrak{Z}$ is a first order stationary Markov process with $P(Z = 1) = p_Z$ and correlation coefficient is ε_Z. The channel input Y^n is uniformly distributed.

For the simulations shown in Fig. 4.6 - 4.10, the correlation of $\mathfrak{X}$ and $\mathfrak{Y}$ is modeled as BMNC $Y_i = X_i \bigoplus Z_i$ where the noise $\mathfrak{Z}$ is a first order stationary Markov process where $P(Z = 1) = p_Z$ and correlation coefficient is ε_Z. $\mathfrak{X}$ is a first order stationary Markov source with $P(X = 1) = p_X$ and correlation coefficient is ε_X. In these simulations, $\varepsilon_X = 0.5$, $p_X = 0.1$.

For the simulations shown in Fig. 4.11 - 4.15, the correlation between $\mathfrak{X}$ and $\mathfrak{Y}$ is modeled as a GEC $Y_i = X_i + Z_i$, where the noise $\mathfrak{Z}$ is a hidden Markov process with two states S_1 and S_2. The probability that S_2 is current state is p_Z and correlation coefficient is ε_Z. p_1 and p_2 are crossover probabilities for these two states S_1 and S_2, respectively. $\mathfrak{X}$ is a first order stationary Markov source where $P(X = 1) = p_X$ and correlation coefficient is ε_X. In these simulations, $\varepsilon_X = 0.5$, $p_X = 0.1$, $p_1 = 0.9$, $p_2 = 0.01$.

In Figs. 4.1, 4.2, 4.6, 4.7, we observe that for the $(7, 4)$ Hamming code, MD, MD+ and ML decoding are identical for Wyner's syndrome source coding scheme for the cases

$\varepsilon_Z = 0.1$ and $\varepsilon_Z = 0.25$. These simulations coincide with Theorem 11 since the (7,4) Hamming code is a perfect code and equations (3.1) and (3.2) are satisfied. In Figs 4.11 and 4.12, the correlation channel is a GEC. Although we do not have a theorem to show that MD, MD+ and ML decoding methods are equivalent for perfect codes under certain condition when the correlation channel is GEC, these three decoding methods are still identical in these simulations.

In Figs. 4.3, 4.4, 4.8, 4.9, 4.13 and 4.14, we remark that for the $(15, 7)$ BCH code, MD+ is improved from MD decoding. In Fig. 4.13, MD+ is almost identical to ML decoding. By comparing those cases when $\varepsilon = 0.1$ and $\varepsilon = 0.25$, we note that when source correlation is smaller, MD+ decoding performs closer to ML decoding.

Figs. 4.5, 4.10, 4.15 indicates that for the $(23, 12)$ Golay code, when MD+ decoding is implemented it does not show any improvement over MD decoding. This is expected since there are no ties in MD decoding for the $(23, 12)$ Golay code, which is a perfect code. Note that (4.3) does not hold for the parameters used in these simulations.

In Figs. 4.6-4.15, MAP decoding outperforms than MD, MD+ and ML decoding as expected since MAP decoding is optimal. This comes however at a cost of substantial increase in decoding complexity.

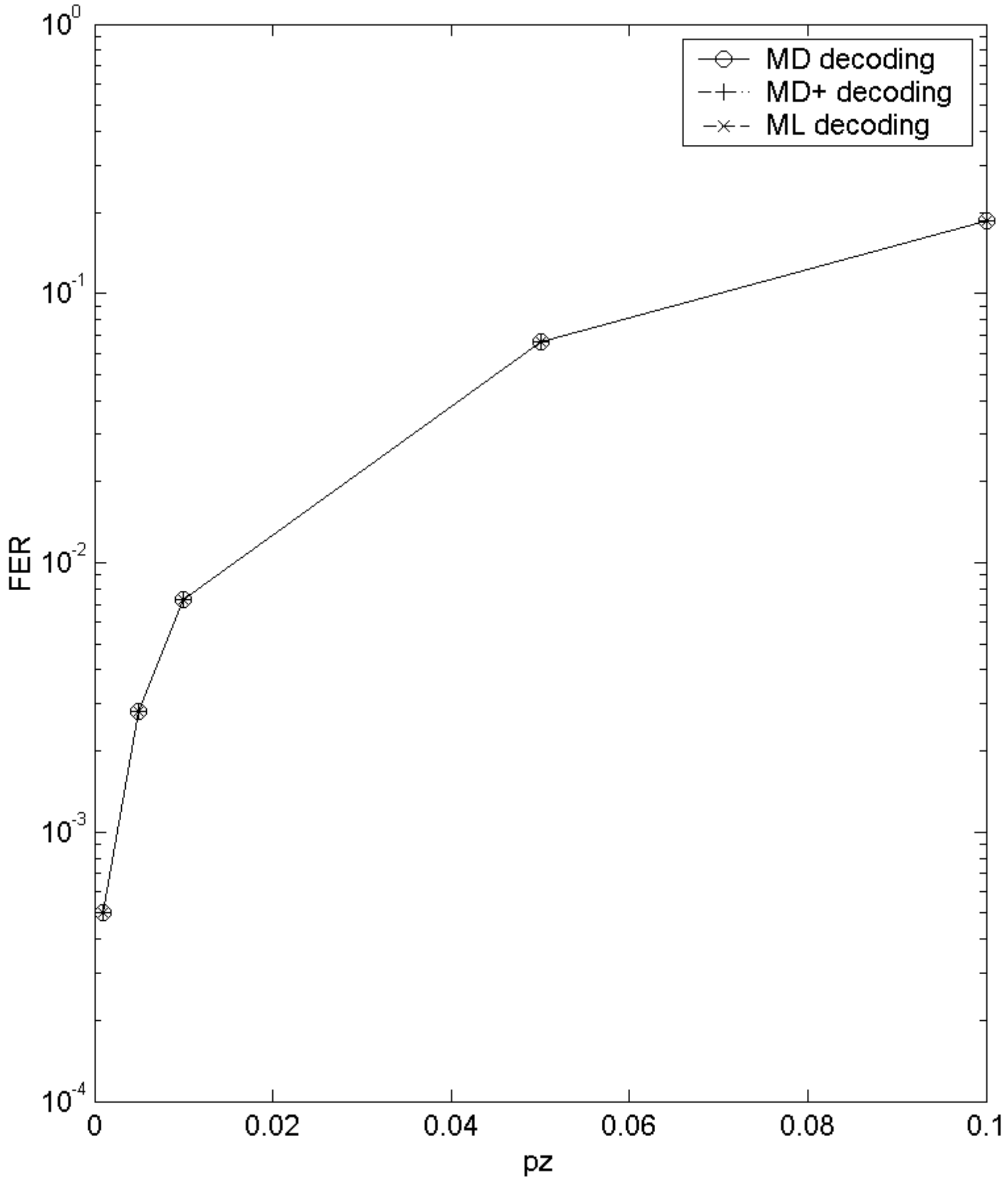

Figure 4.1: FER vs. p_z under different decoding methods for Wyner's syndrome source coding based on (7,4) Hamming code for BMNC correlation channel $X_i = Y_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

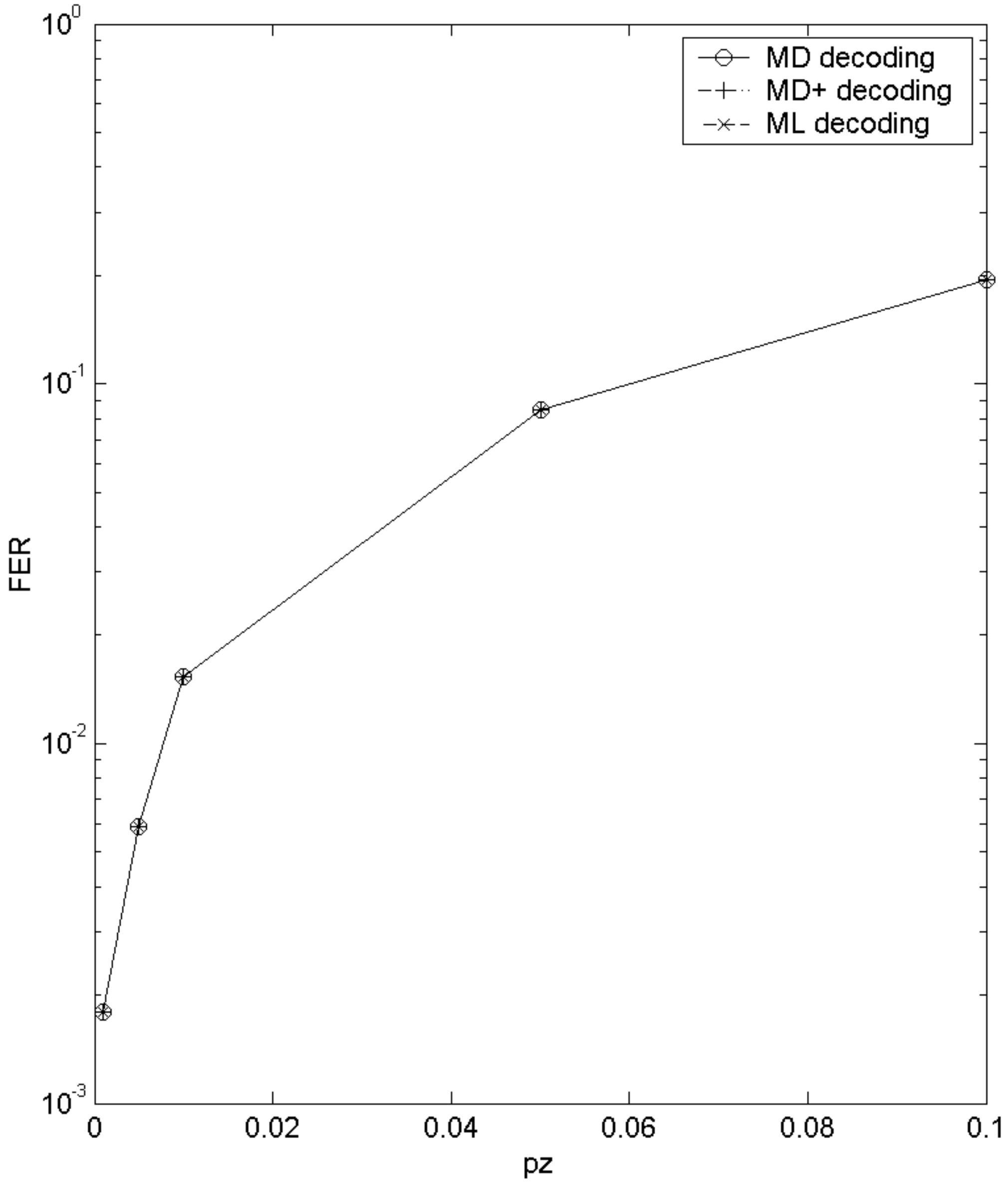

Figure 4.2: FER vs. p_z for Wyner's syndrome source coding based on (7,4) Hamming code for BMNC correlation channel $X_i = Y_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.25$

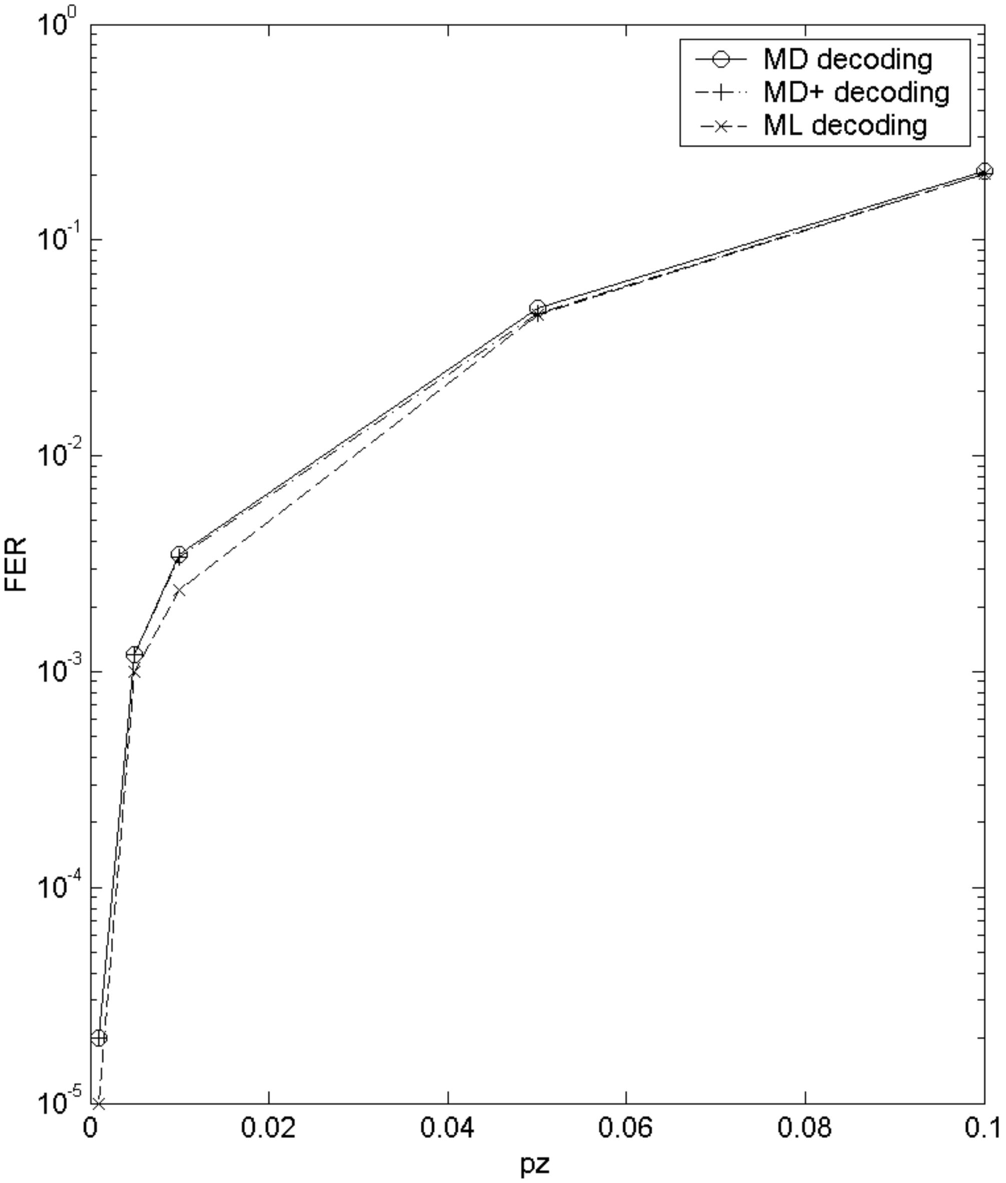

Figure 4.3: FER vs. p_z for Wyner's syndrome source coding based on (15,7) BCH code for BMNC correlation channel $X_i = Y_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

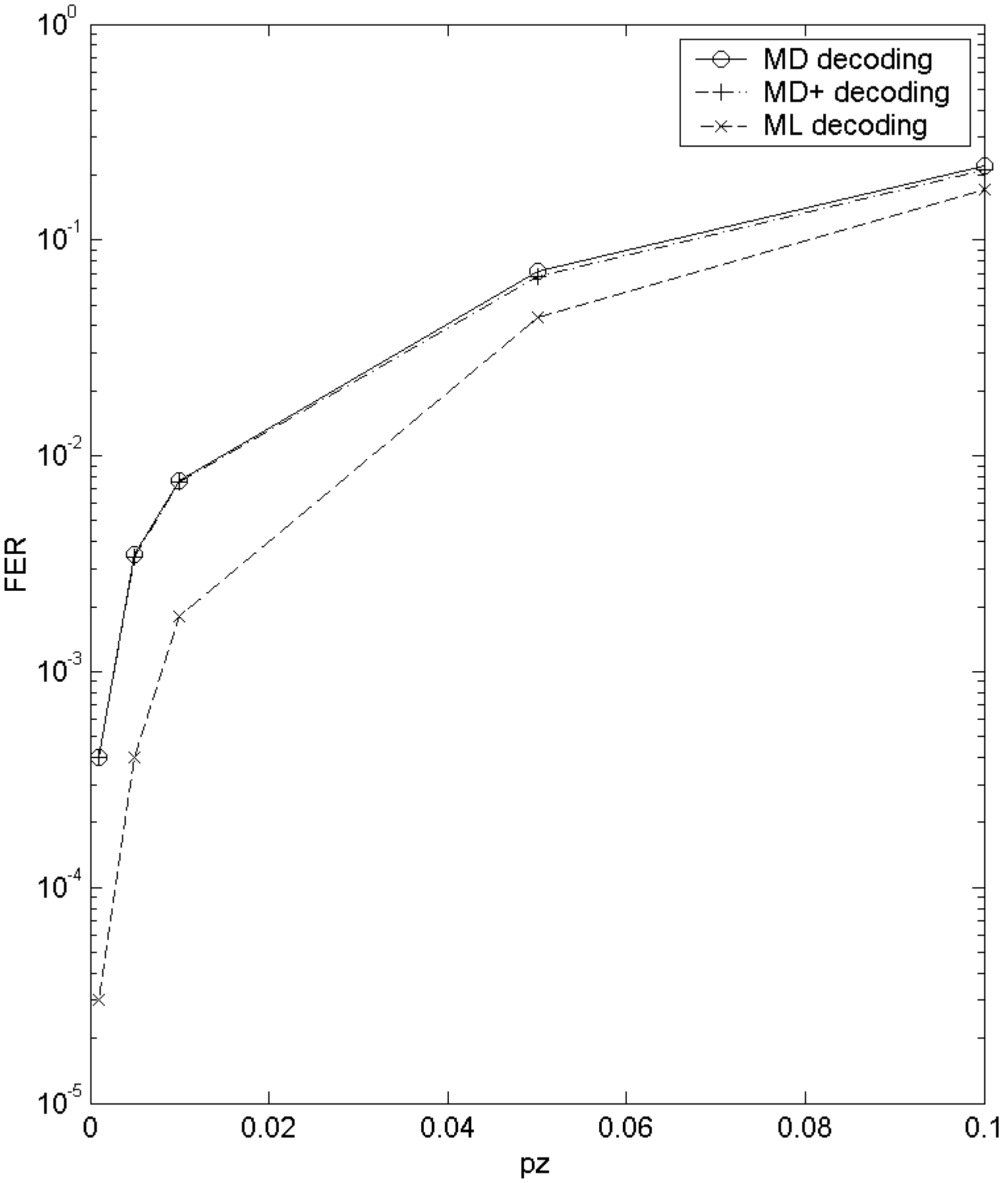

Figure 4.4: FER vs. p_z for Wyner's syndrome source coding based on (15,7) BCH code for BMNC correlation channel $X_i = Y_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon = 0.25$

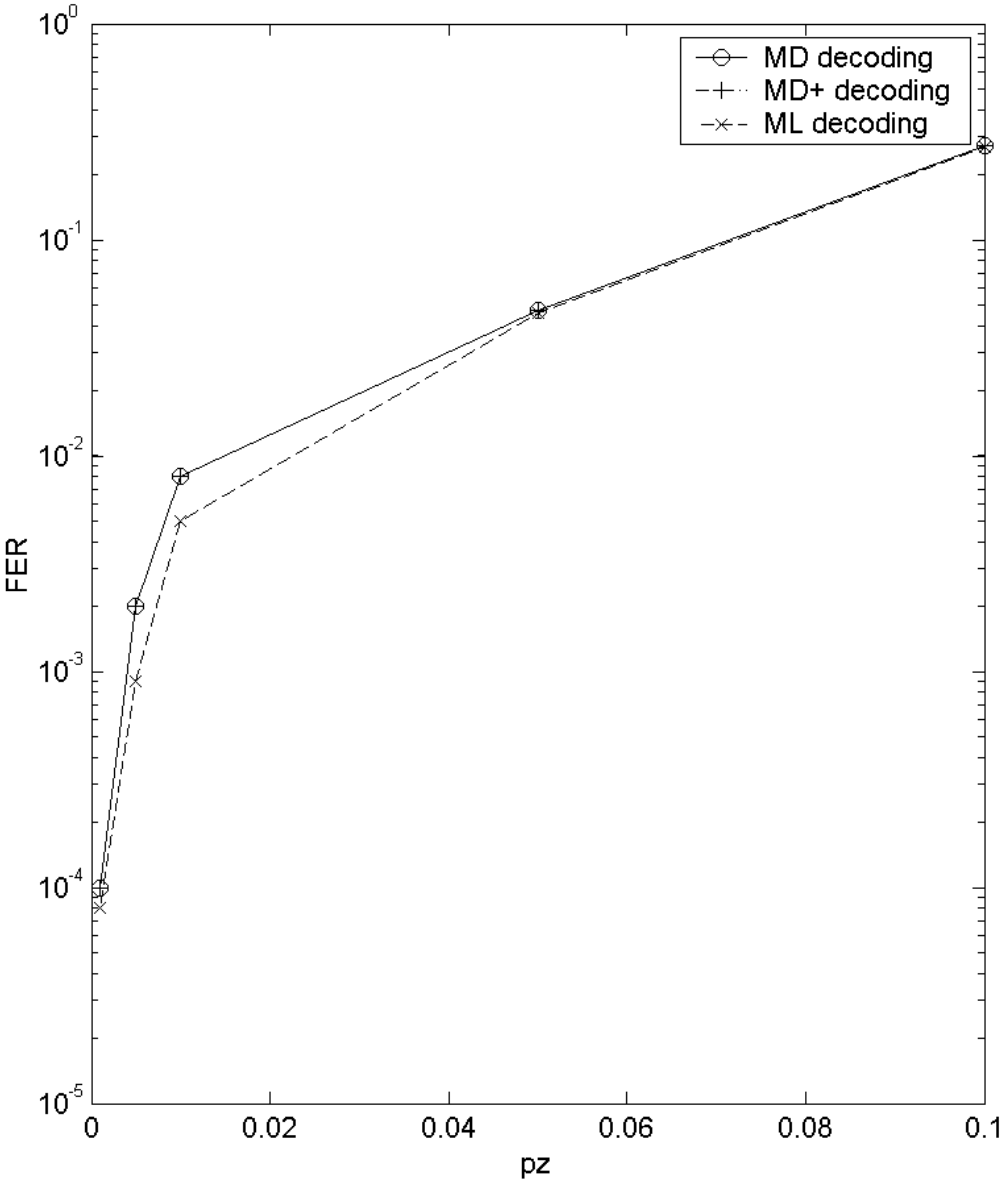

Figure 4.5: FER vs. p_z for Wyner's syndrome source coding based on (23,12) Golay code for BMNC correlation channel $X_i = Y_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon = 0.1$

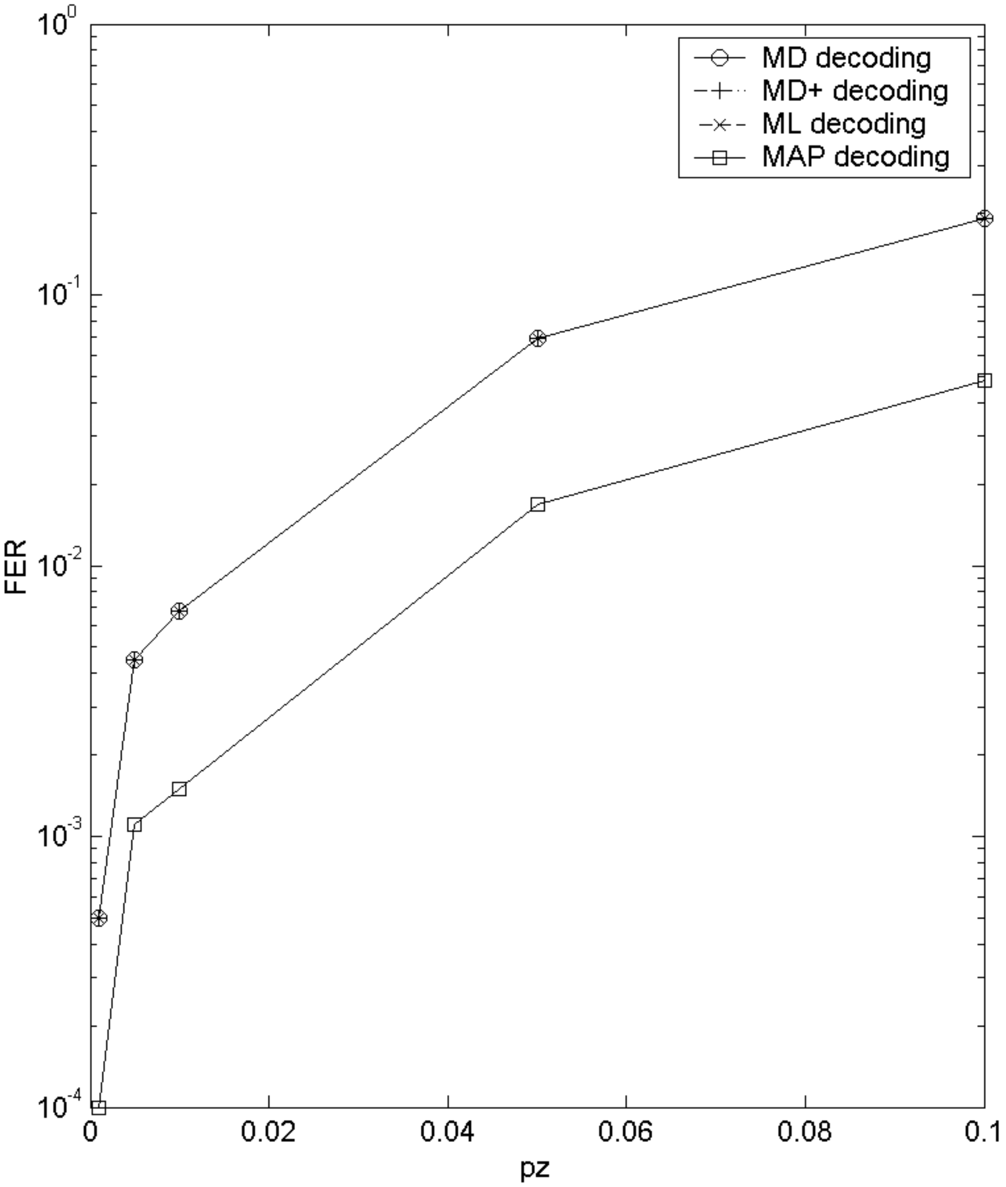

Figure 4.6: FER vs. p_z for Wyner's syndrome source coding based on (7,4) Hamming code for BMNC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

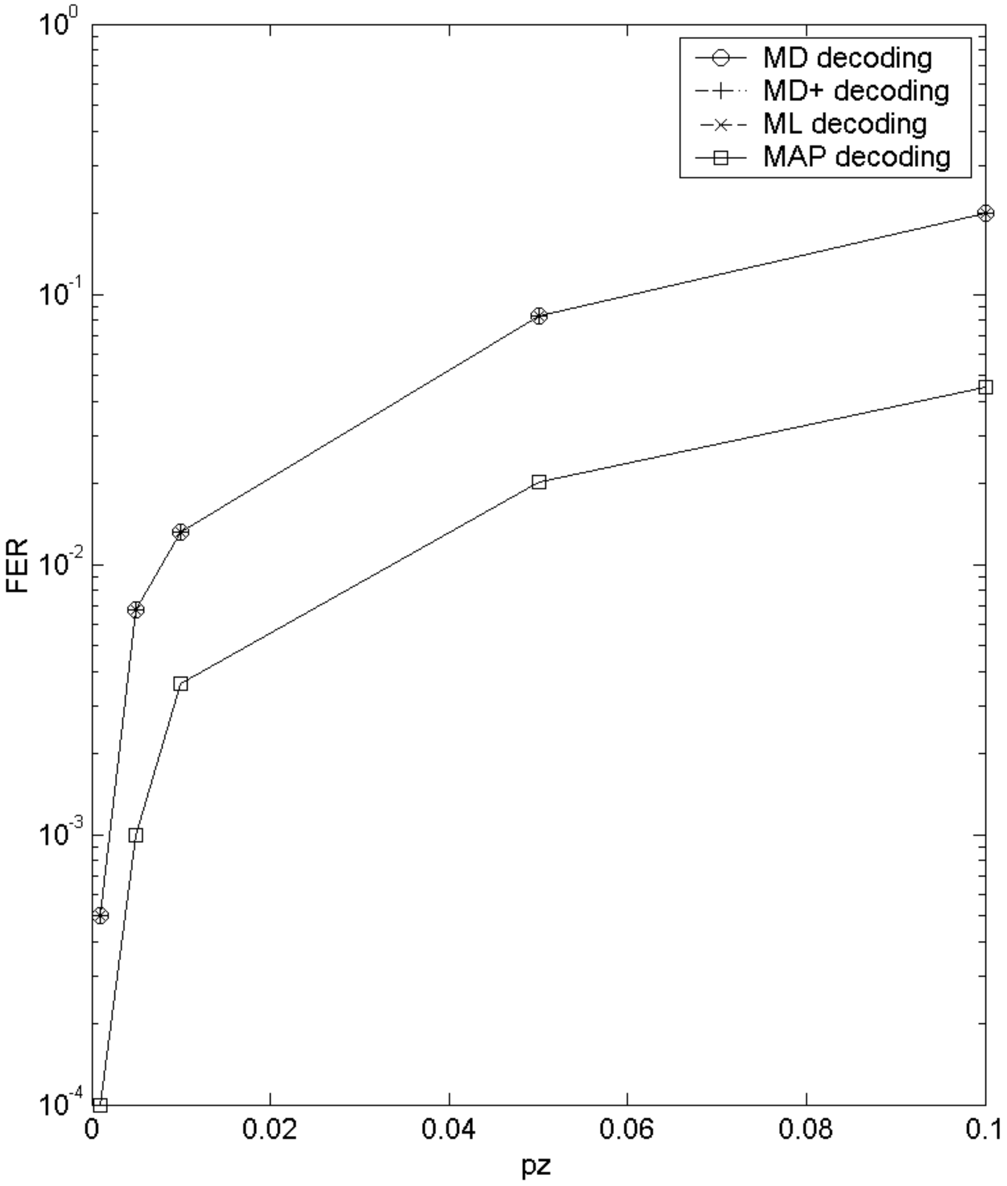

Figure 4.7: FER vs. p_z for Wyner's syndrome source coding based on (7,4) Hamming code for BMNC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.25$

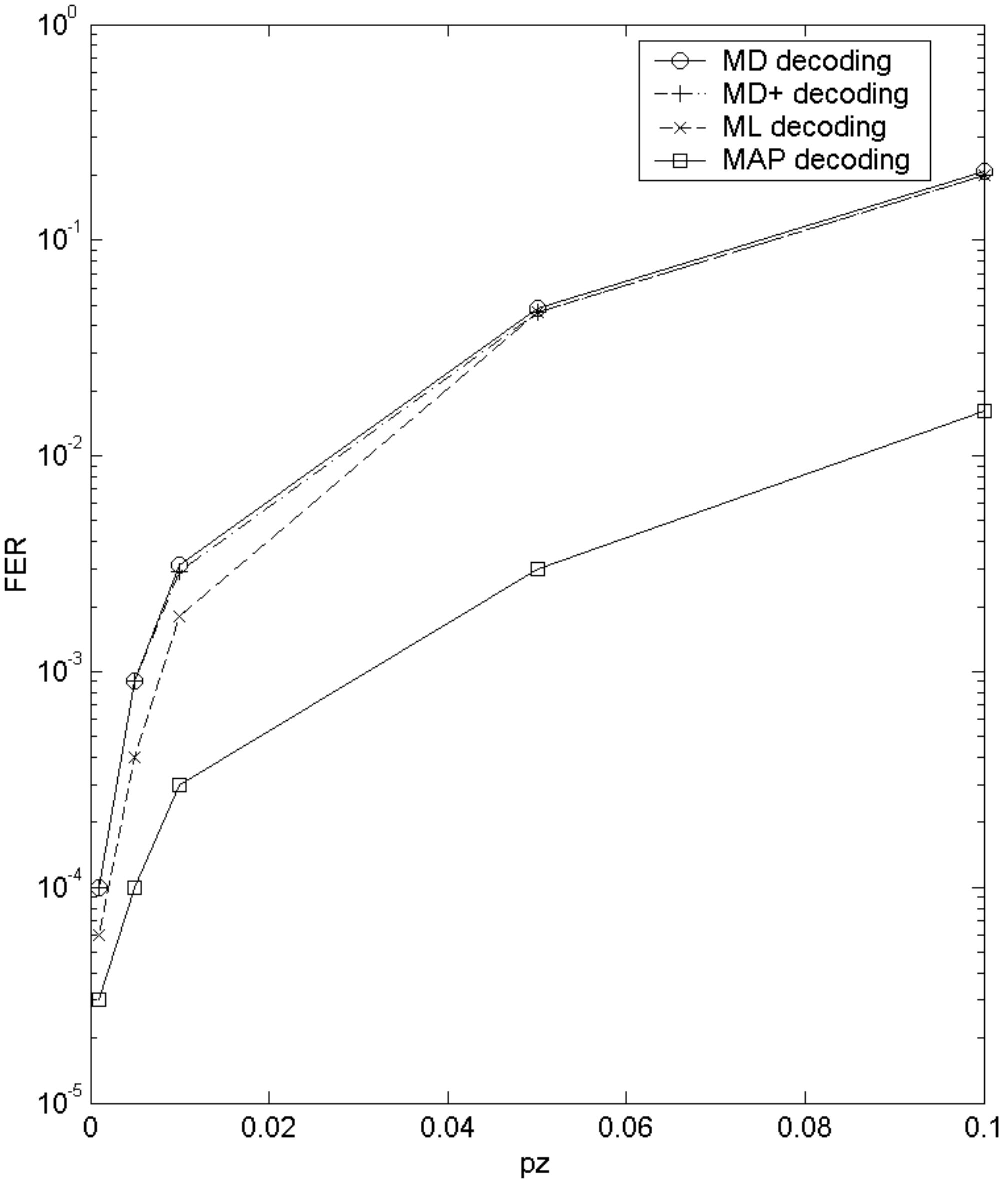

Figure 4.8: FER vs. p_z for Wyner's syndrome source coding based on (15,7) BCH code for BMNC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

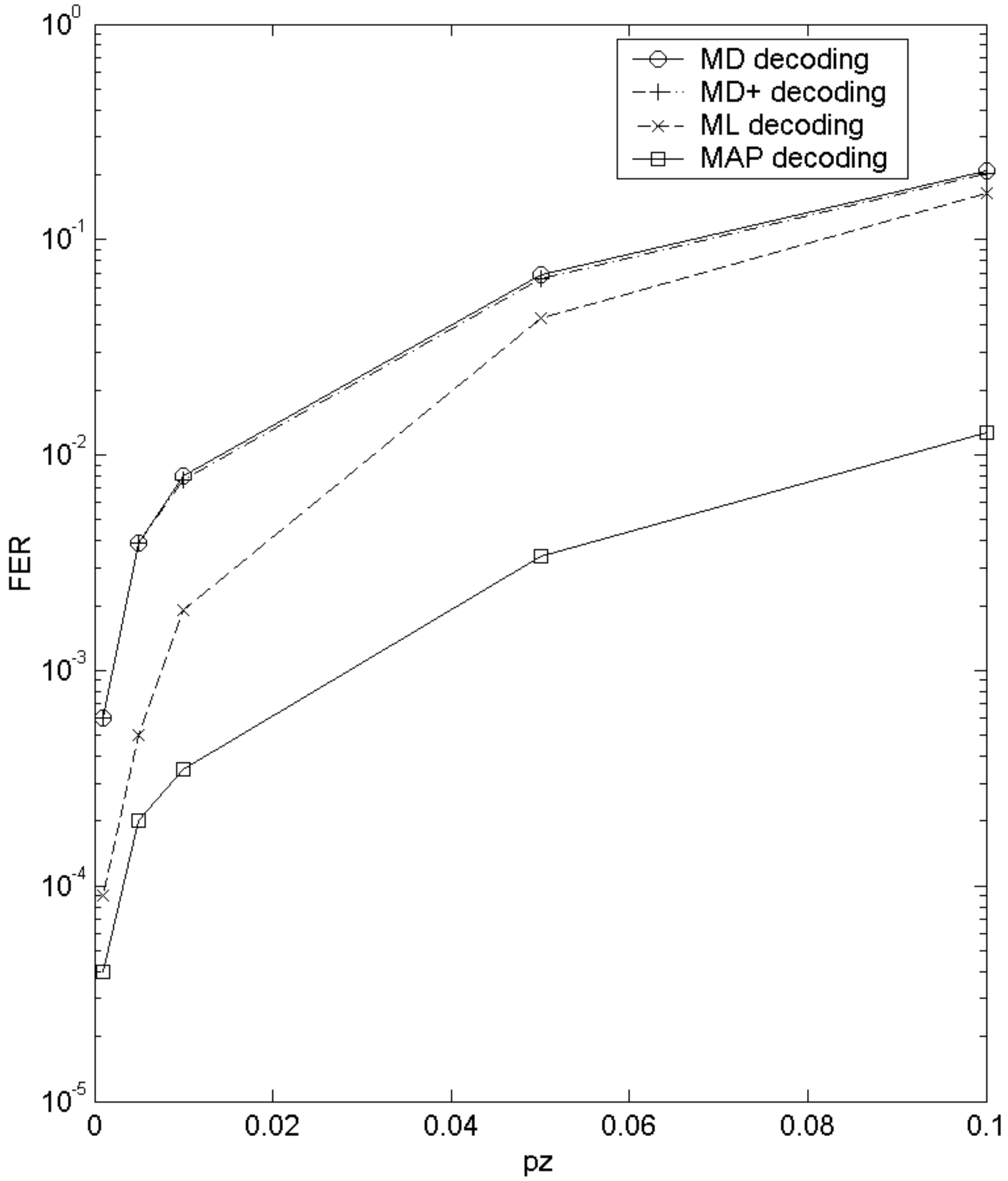

Figure 4.9: FER vs. p_z for Wyner's syndrome source coding based on (15,7) BCH code for BMNC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.25$

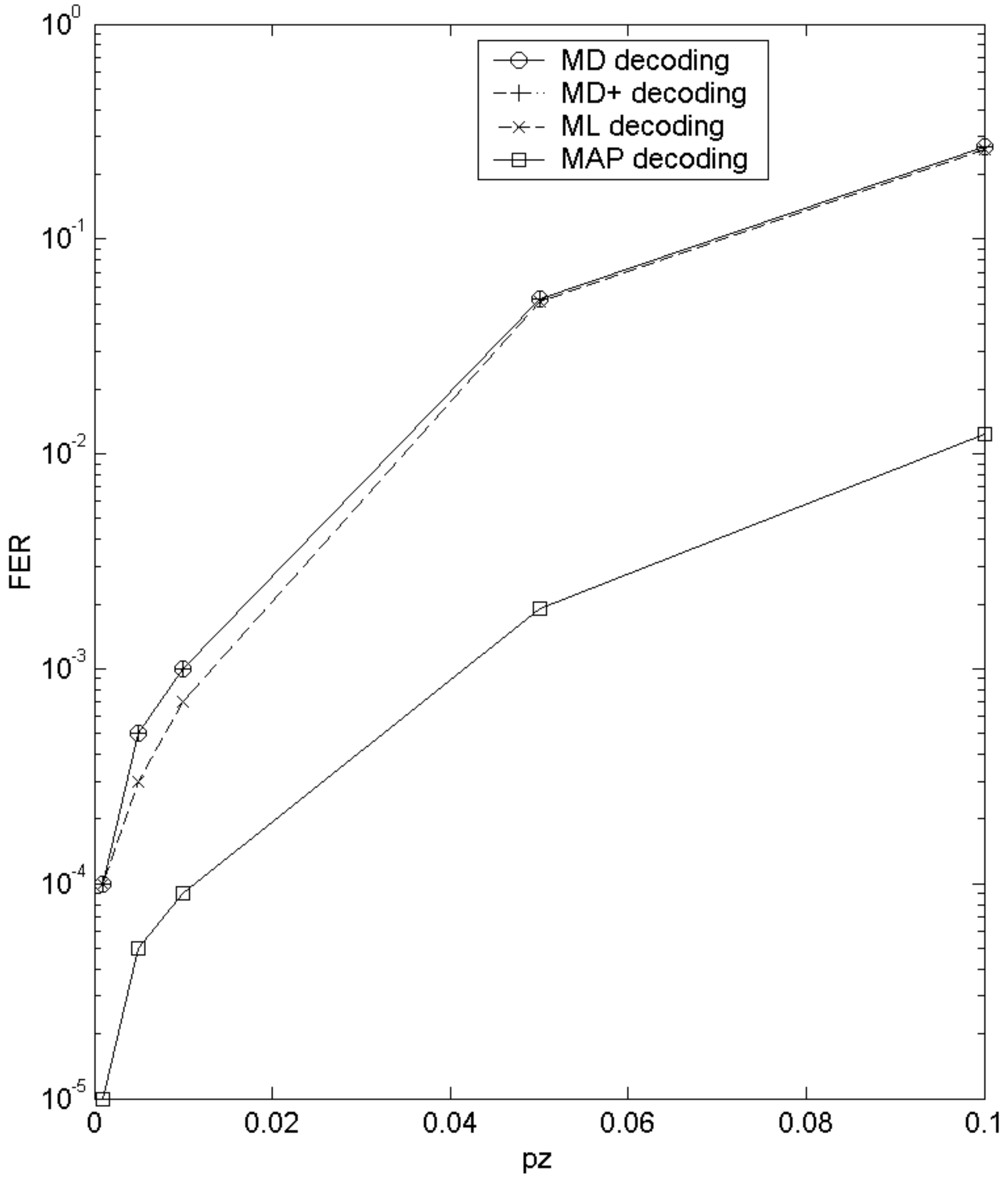

Figure 4.10: FER vs. p_z for Wyner's syndrome source coding based on (23,12) Golay code for BMNC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

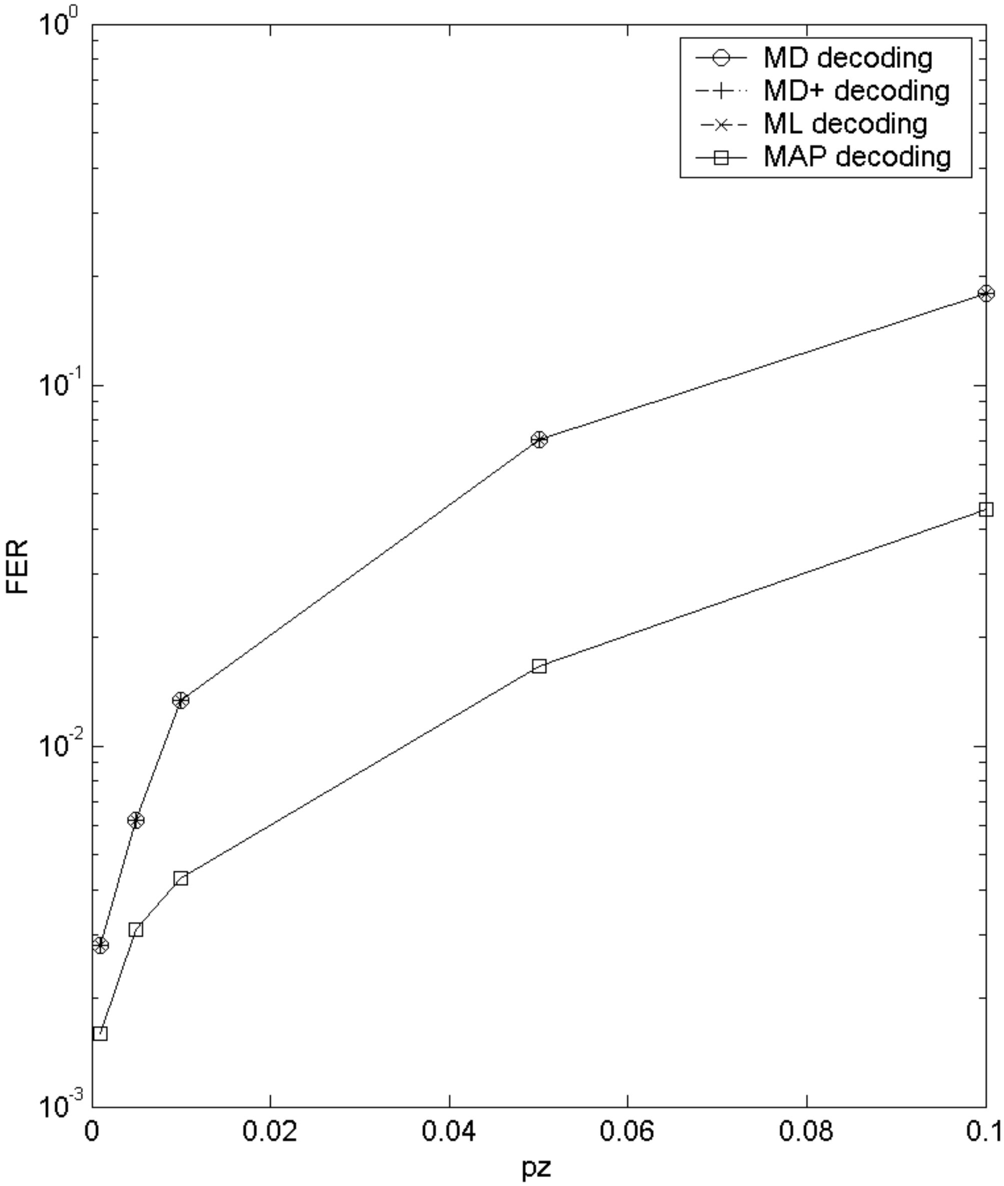

Figure 4.11: FER vs. p_z for Wyner's syndrome source coding based on (7,4) Hamming code for GEC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient$\varepsilon_Z = 0.1$

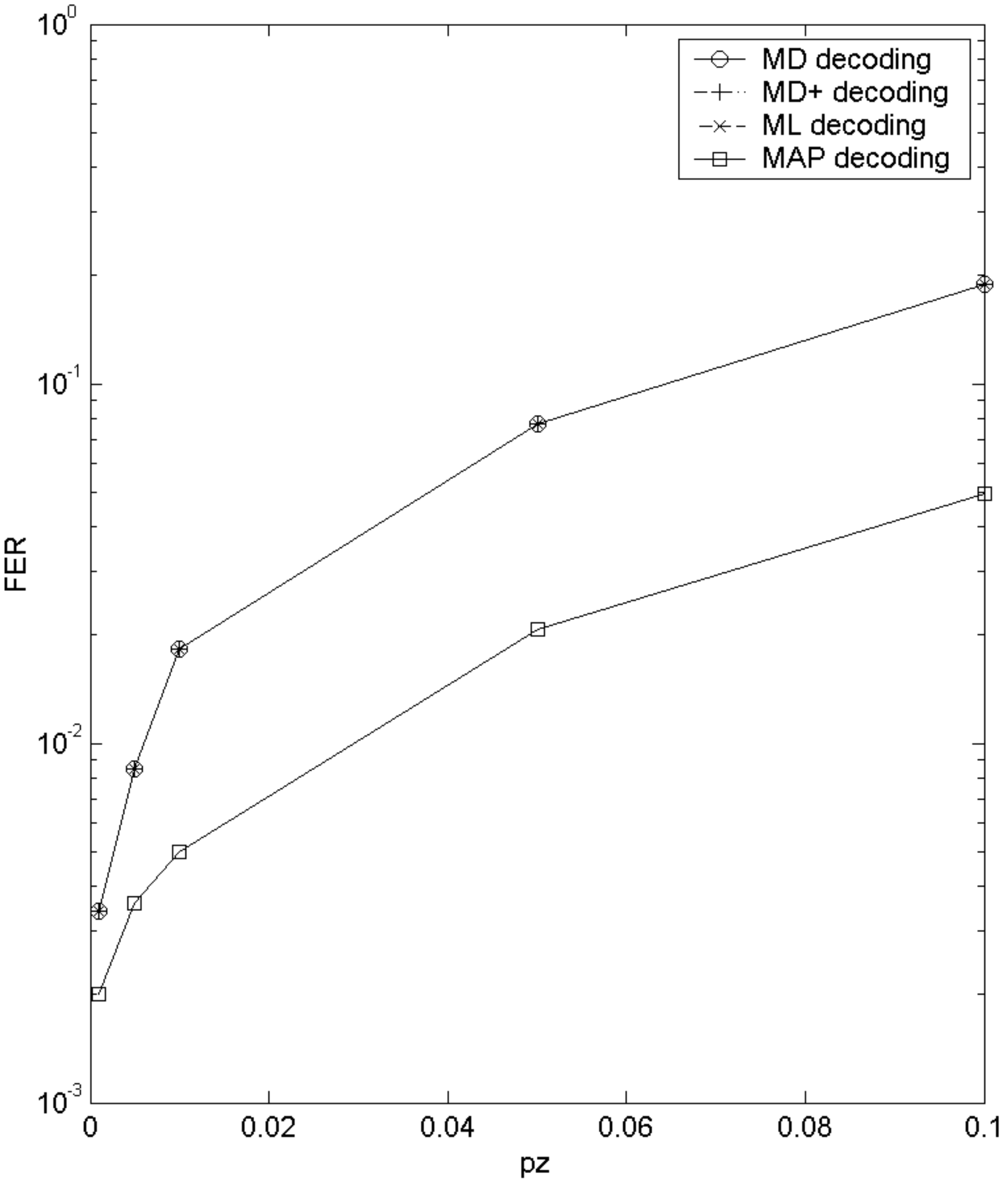

Figure 4.12: FER vs. p_z for Wyner's syndrome source coding based on (7,4) Hamming code for GEC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.25$

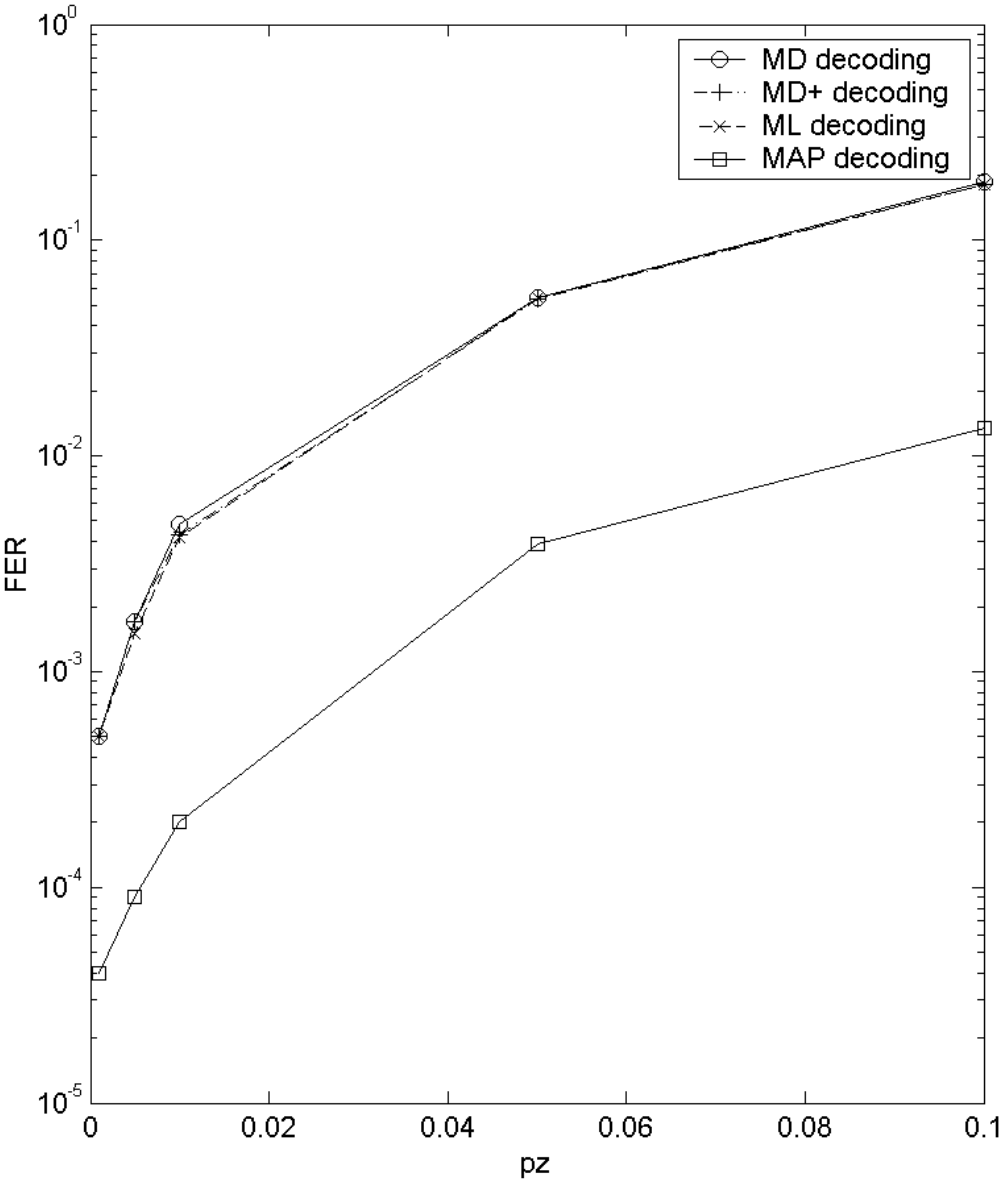

Figure 4.13: FER vs. p_z for Wyner's syndrome source coding based on (15,7) BCH code for GEC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

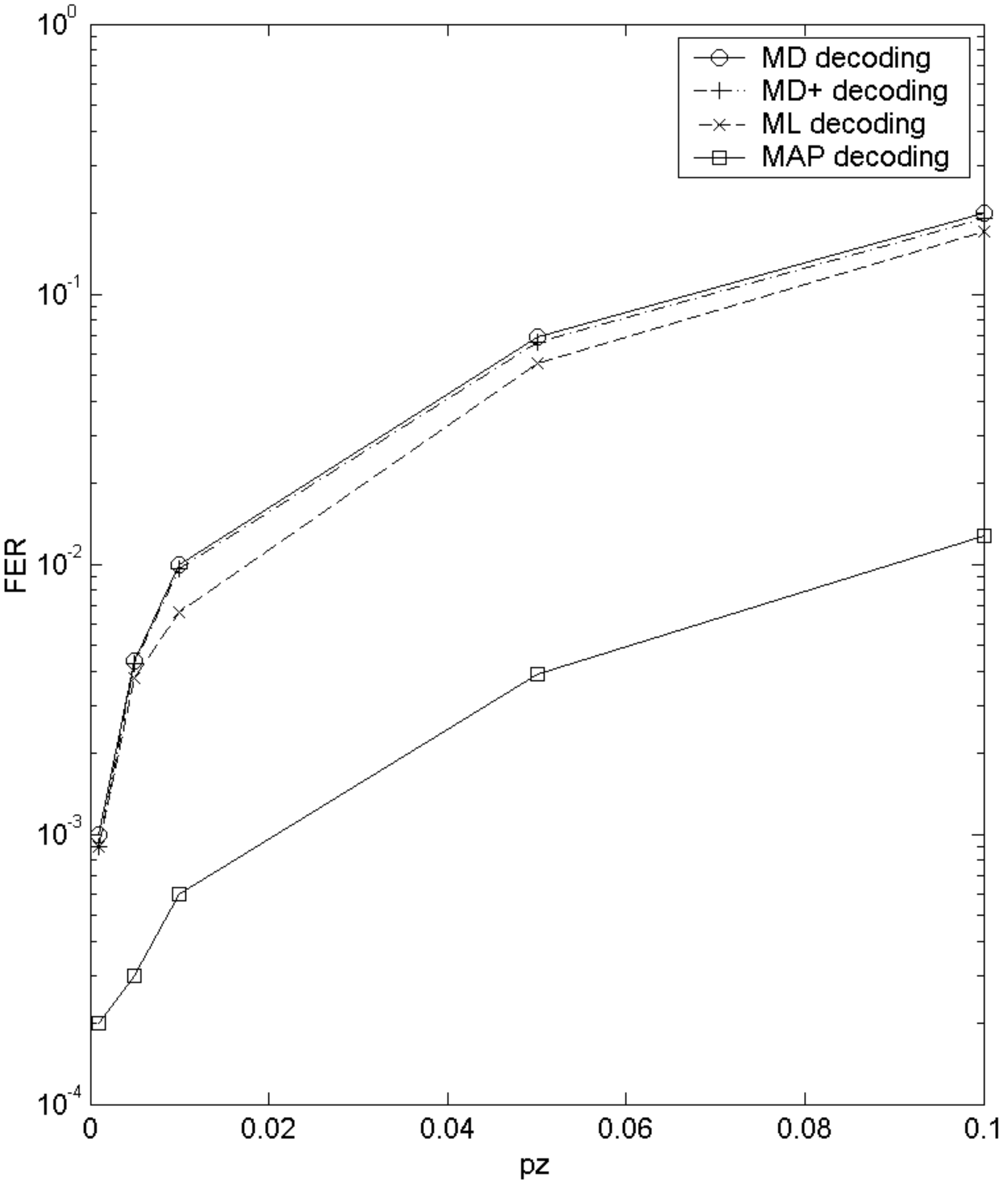

Figure 4.14: FER vs. p_z for Wyner's syndrome source coding based on (15,7) BCH code for GEC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.25$

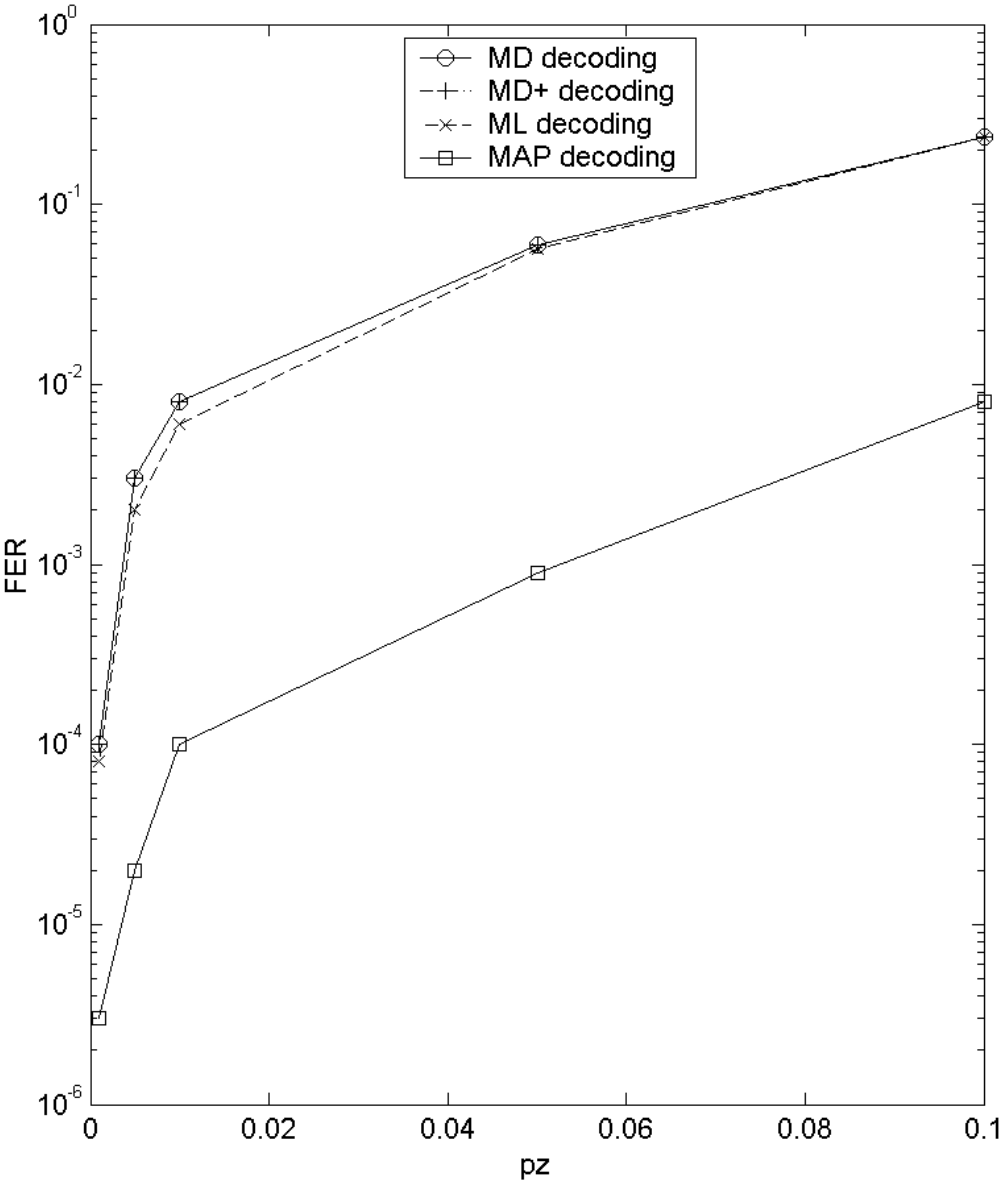

Figure 4.15: FER vs. p_z for Wyner's syndrome source coding based on (23,12) Golay code for GEC correlation channel $Y_i = X_i \bigoplus Z_i$ with noise correlation coefficient $\varepsilon_Z = 0.1$

Chapter 5

Conclusion and Future Work

In this work, we studied syndrome source coding for two cases: data compression without side information and data compression with side information.

For data compression without side information, we proved that syndrome source coding scheme can achieve asymptotically the entropy rate of stationary and ergodic sources. The MD, SMD, ML, SML, MD+ decoding methods were provided for this scheme, and the relationships among these decoding methods with perfect or quasi-perfect codes for Markov sources were explained. We also provided simulation results using Hamming, BCH and Golay codes under the different decoding methods to show the effectiveness of syndrome source coding based on these codes, and the good performance of MD+ decoding method.

For data compression with side information, we proved that syndrome source coding scheme can achieve the Slepian-Wolf limit asymptotically for sources with additive noise correlation channel, where the noise is stationary and ergodic. The MD, SMD, ML, SML, MD+ decoding methods were provided for this scheme, and the relationships among these decoding methods with perfect or quasi-perfect codes for Markov sources were also presented. Furthermore we studied another more natural model with the side information as the output, and the MAP decoding was presented, which is the optimal decoding method. We also provided simulation results using Hamming, BCH and Golay codes under these decoding methods.

Lossless data compression with side information is only a special case of the Wyner-Ziv Theorem [19], which examines lossy data compression with side information. In future work, one can study lossy data compression using linear block codes with or without side information to achieve the Wyner-Ziv rate distortion function. Some progress for this problem was reported by Matsunaga et al. [14] and Miyake [15].

Bibliography

[1] H. AI-Lawati, "Performance analysis of linear block codes over the queue-based channel," M.Sc. thesis, Sep. 2007.

[2] H. Al-Lawati and F. Alajaji, "On decododing binary perfect and quasi-perfect codes over Markov noise channels," *IEEE Trans. on Commun.*, to appear.

[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] G. Caire, S. Shamai and S. Verdu, "A new data compression algorithm for sources with memory based on error correcting codes," *ITW 2003*, pp. 291-295, March 31-April 4, 2003.

[5] T. M. Cover and J.A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[6] T. M. Cover, "A Proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Information Theory*, vol.21 pp.226-228, March 1975.

[7] I. Csiszar and J. Corner, *Infroamtion Theory: Coding Theorems for Discrete Memoryless Systems*, Academic, New Yrok, 1981.

[8] P. Elias,"Coding for noisy channels," *IRE Convetion Record*, vol.4, pp.37-46, 1955.

[9] J. Garcia-Frias and W. Zhong, "LDPC codes for compression of muti-termial sources with hidden Markov correlation," *IEEE Commun. Letters*, vol.7, No.3, pp.115-117, March 2003.

[10] R. G. Gallager, *Low-Density Parity-Check Codes*, M.I.T Press, 1963.

[11] R. G. Gallager, *Infromation Theory and Reliable Communication*, Wiley, New York, 1968.

[12] A. Liveris, Z. Xiong and C. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Letters*, vol.6, no.10, pp. 440-442, 2002.

[13] D. J. C. Mackay, "Good error-correcting codes based on very sparse matrices," *IEEE Irans. on Inform. Theory*, vol.45, no.2, pp. 399-41, March 1999.

[14] Y. Matsunaga and H. Yamamoto, "A coding theorem for lossy data compression by LDPC codes," *IEEE Trans. Inform. Theory*, vol.49, pp.2225-2229, Sep. 2003.

[15] S. Miyake, "Lossy Data Compression over Z_q by LDPC code," ISIT 2006, pp.813-816, Seattle, USA, July, 2006.

[16] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol.27, pp.379-423,623-656, July and Oct. 1948.

[17] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. on Information Theory*, vol. 19, pp. 471-480, July 1973.

[18] A. D. Wyner,“Recent results in the shannon theory,” *IEEE Trans. on Inform. Theory*, vol.IT-20, No.1, Jan. 1974.

[19] 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.

[20] Z. Xiong, A. D. Liveris, and S. Cheng, “Distributed source coding for sensor networks,” *IEEE Signal Processing*, Vol. 21, Issue 5, pp.80 - 94, Sept. 2004.