ABSTRACT

OBLIVIOUS DATA HIDING: A PRACTICAL APPROACH

by
Husrev T. Sencar

This dissertation presents an in-depth study of oblivious data hiding with the emphasis on quantization based schemes. Three main issues are specifically addressed:

1. Theoretical and practical aspects of embedder-detector design.
2. Performance evaluation, and analysis of performance vs. complexity tradeoffs.
3. Some application specific implementations.

A communications framework based on channel adaptive encoding and channel independent decoding is proposed and interpreted in terms of oblivious data hiding problem. The duality between the suggested encoding-decoding scheme and practical embedding-detection schemes are examined. With this perspective, a formal treatment of the "processing" employed in quantization based hiding methods is presented. In accordance with these results, the key aspects of embedder-detector design problem for practical methods are laid out, and various embedding-detection schemes are compared in terms of probability of error, normalized correlation, and hiding rate performance merits assuming AWGN attack scenarios and using mean squared error distortion measure.

The performance-complexity tradeoffs available for large and small embedding signal size (availability of high bandwidth and limitation of low bandwidth) cases are examined and some novel insights are offered. A new codeword generation scheme is proposed to enhance the performance of low-bandwidth applications. Embedding-detection schemes are devised for watermarking application of data hiding, where robustness against the attacks is the main concern rather than the hiding rate or payload. In particular, cropping-resampling and lossy compression types of non-invertible attacks are considered in this dissertation work.

OBLIVIOUS DATA HIDING: A PRACTICAL APPROACH

by
Husrev T. Sencar

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering

January 2004

APPROVAL PAGE

OBLIVIOUS DATA HIDING: A PRACTICAL APPROACH

Husrev T. Sencar

Dr. Ali N. Akansu, Dissertation Advisor — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Richard A. Haddad, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Yun Q. Shi, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Nasir Memon, Committee Member — Date
Associate Professor of Computer Science, Polytechnic University

---

Dr. Mahalingam Ramkumar, Committee Member — Date
Assistant Professor of Computer Science and Engineering, Mississippi State University

BIOGRAPHICAL SKETCH

Author: Husrev T. Sencar

Degree: Doctor of Philosophy

Date: January 2004

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2003
- Master of Science in Electrical Engineering,
  Baskent University, Ankara, Turkey, 1998
- Bachelor of Science in Electrical Engineering,
  Middle East Technical University, Ankara, Turkey, 1996

Major: Electrical Engineering

Presentations and Publications:

H. T. Sencar, M. Ramkumar, A. N. Akansu, Data Hiding Fundamentals and Applications, Academic Press, 2003.

A. N. Akansu, H. T. Sencar "Orthogonal Transmultiplexers: A Time Frequency Perspective," The Wiley Encyclopedia of Telecommunications, John Wiley & Sons, Inc., 2003.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "An Overview of Scalar Quantization Based Data Hiding Methods," *submitted to IEEE Transactions on Signal Processing*, 2003.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "An Embedding-Detection Technique for Data Hiding with Small Host Signal Sizes," *submitted to IEEE Transactions on Signal Processing*, 2003.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "An Analysis of Quantization Based Embedding-Detection Techniques," *submitted to IEEE ICASSP*, 2004.

H. T. Sencar, M. Ramkumar, A. N. Akansu, " A New Perspective for Embedding-Detection Methods with Distortion Compensation and Thresholding processing techniques," *Proceedings of IEEE ICIP*, 2003.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "Multiple Codebook Information Hiding Based on Minimum Distortion Criterion," *Proceedings of CISS*, 2003.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "A Robust Type-III Data Hiding Technique Against Cropping and Resizing Attacks," *Proceedings of IEEE ISCAS*, 2002.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "Improvements on Data Hiding for Lossy Compression," *Proceedings of IEEE ICASSP*, 2002.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "Multiple Codebook Information Hiding," *Proceedings of CISS*, 2002.

H. T. Sencar, M. Ramkumar, A. N. Akansu, "Efficient Codebook Structures for Practical Information Hiding Systems," *Proceedings of CISS*, 2001.

T. Sencar, G. Bozdagi, "Video segmentation based on MPEG bitstream," *Proceedings of CBMI*, 1999.

G. Bozdagi, T. Sencar, "Preprocessing Tool for Compressed Video Editing," *Proceedings of IEEE MMSP*, 1999.

R. de Queiroz, G. Bozdagi, T. Sencar, "Fast Video Segmentation Using Encoding Cost Data," *Proceedings of SPIE*, 1999.

To my family

ACKNOWLEDGMENT

I express my sincere gratitude to my advisor, Professor Ali N. Akansu, for his guidance, supervision, and moral support which made this work possible. He taught me what it means to be thorough, diligent, and professional. I am also thankful to Professor Mahalingam Ramkumar, my predecessor in the same field, for many insightful and fruitful discussions. His work has been an inspiration for this work. Thanks are also due to Professor Richard Haddad, Professor Yun Q. Shi, and Professor Nasir Memon for serving in my committee and for their time.

During my study, I have enjoyed the company of many friends and lab-mates without whom my stay in the school would have been very lonely. I am especially happy to have been associated with Ramkumar, Burak, Sebnem, Zafer, Kadir, Anil, Litao, Xuefei, Chris, Amer, Zoran, Jordi, Jingdi, Zhen, Woo-Jin, Seekhyun, Ozgur, and others.

The staff of ECE Department Ms. Brenda Walker and Ms. Barbara Faltz were always there for me. Their help would not go unnoticed.

The acknowledgment would not be complete without mentioning the ones to whom I owe the most. I am very grateful to my parents and my brother for their everlasting love and unconditional support. Finally, I wish to express the deep and heartfelt gratitude to my wife, Yelda, who went through this experience along with me, for her forbearance and unwavering encouragement. It would not have been possible without them.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

The study of data hiding (information hiding, watermarking) tries to establish the achievable limits and the design of methods for conveying a message data, embedded within a host (cover) signal, in an imperceptible and reliable way. Data hiding techniques aim at achieving three primary goals. These are:

- *Hiding rate:* The maximum amount of message data that can be embedded in a given host signal.
- *Robustness:* The level of resistance of the embedded signal (stego signal) against all forms of attacks so that the embedded message data can be reliably extracted by the receiver.
- *Transparency:* The degree of perceptual degradation in the host signal due to the embedding operation.

The design of optimum embedding and detection operations is the central issue in data hiding research.

Data hiding study provides tools that can be employed to serve a variety of purposes including, but not limited to, copyright control, ownership verification, secure media distribution, transaction tracking, authentication, captioning, and hybrid analog and digital communications. Ultimately, data hiding applications are classified based on how they make use of the tradeoff among the conflicting goals of hiding rate, transparency and robustness. Designing practical methods that will achieve wide acceptance depends on exploiting this tradeoff optimally. This requires an approach that incorporates the findings of many research areas, [1, and the references therein]. A significant number of researchers have introduced sophisticated

information hiding techniques that approach information theoretic limits of data hiding capacity, [2, 3, 4, 5, 6].

Performance of data hiding methods is usually restricted by the maximum amount of distortion that may be introduced to the host signal with no perceptual distortion. The embedding distortion is ideally derived from a perceptual distortion measure, and it is the resource of the communication between embedder and detector. The information hider needs to design the embedder-detector that makes the most effective use of this core resource.

One conservative assumption in data hiding is that the embedder has no access to the host signal (oblivious data hiding). Though, not all data hiding applications are necessarily oblivious, the focus in this dissertation is the oblivious one.

1.1 Data Hiding Framework

Let $\mathbf{C} \in \Re^N$ be some sampled real valued information signal, and $\mathbf{W} \in \Re^N$ the auxiliary message signal. An embedder $\mathcal{E}$ embeds the message signal $\mathbf{W}$ in the host signal $\mathbf{C}$ to yield the stego signal $\mathbf{S} \in \Re^N$ given as

$$\mathbf{S} = \mathcal{E}(\mathbf{C}, \mathbf{W}). \tag{1.1}$$

Let $d(.,.)$ be a predefined distortion metric suitable to information signal $\mathbf{C}$. In other words $d(\mathbf{S}, \mathbf{C})$, is the "distance" between $\mathbf{S}$ and $\mathbf{C}$. A commonly used metric or distance measure is the mean squared error given by

$$d(\mathbf{S}, \mathbf{C}) = \sum_{i=1}^{N} \frac{(\mathbf{S} - \mathbf{C})^2}{N} \tag{1.2}$$

The *embedding distortion*, $d(\mathbf{S}, \mathbf{C})$ is constrained to be less than a defined threshold P to ensure that the cover signal $\mathbf{C}$ and the stego signal $\mathbf{S}$ are perceptually the same or very similar.

The stego signal is corrupted by a noise signal $\mathbf{Z} \in \Re^N$ before it reaches the detector $\mathcal{D}$. At the detector, an estimate $\hat{\mathbf{W}} \in \Re^N$ of the message signal $\mathbf{W}$ is obtained from the received signal $\mathbf{Y}=\mathbf{S}+\mathbf{Z}$ as

$$\hat{\mathbf{W}} = \mathcal{D}(\mathbf{Y}). \tag{1.3}$$

The problem now boils down to the optimal design of embedder $\mathcal{E}$ and detector $\mathcal{D}$ to maximize the "fidelity" of $\hat{\mathbf{W}}$, subject to the distortion constraint $d(\mathbf{S}, \mathbf{C}) \leq P$.

The above setting can be equivalently translated into a classical communications problem. Consider a message letter m from an alphabet $\mathcal{M}$ with size M. (The message letter m can equivalently be considered as an index $1 \leq m \leq M$.) The encoder E is to transmit the message letter m to decoder D through N uses of a noisy channel with varying states at each transmission. The channel state vector $\mathbf{C}$ is also available at the encoder as a side information. The encoder uses a code with M codewords of length N and power P. At the decoder, the sent message is decoded from the received noisy codeword as $\hat{m}$. In this case, the objective is to find the optimal encoding and decoding so that reliable communication between E and D is possible for the given power constraint P and the side information $\mathbf{C}$. When the state vector $\mathbf{C}$ is additive to the sent codeword the two scenarios become identical. Consequently, the encoder-decoder pair, (E, D), in the communications framework becomes dual to the embedder-detector pair, $(\mathcal{E}, \mathcal{D})$, in data hiding framework with the inclusion of a mapping rule that maps a message index m to a message signal $\mathbf{W}$ and $\hat{\mathbf{W}}$ to $\hat{m}$ as

$$\mathcal{W} : m \to \mathbf{W} \in \Re^N, \quad \mathcal{W}^{-1} : \hat{\mathbf{W}} \to \hat{m} \in \mathcal{M}. \tag{1.4}$$

In the text following notation is used. Vectors are denoted by bold-faced characters. Random variables and their realizations are denoted by the capital and corresponding lower case letters, respectively, in italic typeface. The matrices are

denoted by 'blackboard bold' letters. For the general case all signals are assumed to be vectors of size N. However, in cases where the vector random variables are independent, identically distributed (*iid*), the analysis is simplified by using the individual random variables in derivations where the vector extensions are straightforward.

1.2 Review of Data Hiding Methods

The early works in the literature for data hiding mainly focused on heuristic approaches. As the similarities between the issues of data hiding and other fields become evident, a variety of approaches were made available by exploiting those similarities. Among these approaches the ones that generated a lot of attention are inspired from spread-spectrum communications and communication with side information [7, 8, 9, 10, 11].

Data hiding techniques are characterized by the embedding and detection techniques employed. Methodologically, the proposed embedder-detector designs can be categorized into two main groups: additive spread-spectrum based methods, and quantization based methods.

In additive spread spectrum methods, the watermark signal is generated by modulating the information symbols with a weighted unit energy spreading vector which is then added to the host signal [12, 13, 14, 15, 16]. By choosing an appropriate weighting factor, perceptual intactness of the host signal is retained. These methods are preferable due to their ease in processing, and their reliability under additive noise interference. With additive embedding, data hiding rate is uncompromisingly traded off against robustness to severe attacks while complying with the perceptual constraints. Major drawback of such methods is that host signal affects as a source of interference at the detector. As a result of this fact, satisfactory performance is not possible unless the host signal is available during detection or host signal interference

is negligibly smaller than the channel interference. In additive schemes, optimal decoding of the watermark signal depends on exact probabilistic characterization of the host signal at the detector.

The shortcomings of additive spread spectrum methods in suppressing the host signal interference are handled by adopting the results of communication with side information to data hiding applications. Costa in [17] introduced the notion that, in a communication channel, a side information *available to encoder but not to decoder* does not necessarily causes a reduction in the communication rate by making an analogy with a hypothetical case where a writer communicates to a reader by writing on a sheet of paper that is covered with *iid* Gaussian dirt spots. Costa showed that the two party can communicate at a rate as high as using a clean sheet of paper. His results, when evaluated within data hiding context, encouraged researchers in designing practical oblivious data hiding schemes that can achieve the hiding capacity.

To achieve the hiding rates that are closer to the upper capacity bound, several implementations that utilize this approach are proposed, in the literature [18, 8, 19, 20, 21, 22]. These techniques are characterized by the use of enhanced quantization procedures in order to design embedding-detection methods that approximate the performance of optimal encoding-decoding. In this class of methods, the optimal implementation requires higher dimensional quantization for embedding. However, a satisfactory performance is also achievable through scalar quantization. On the other hand, the extraction of the hidden message is achieved, most generally, by employing minimum distance decoding due to the use of lattice structures in embedding. As a consequence of such an embedding, these methods are vulnerable against signal scaling. Therefore, they perform well only if the attack is not severe. However, they are suitable for oblivious data hiding applications.

Chen, *et al.* in [23] provide a formal treatment of data hiding methods that use quantization to embed signals, that is called quantization index modulation (QIM).

In this type of methods, quantization is used to force the host signal coefficients to take desired values depending on the information signal to be embedded. Similarly, Chou, *et al.* in [10, 22], based on a duality with distributed source coding problem, implemented the exhaustive codeword generation for Costa's scheme by using a robust optimization method through the utilization of optimal quantizers. In this research direction, the most popular embedding technique is a low complexity implementation of QIM which relies on uniform scalar quantization, that is called dither modulation (DM) [24]. In fact the earliest data hiding methods [25, 26, 27, 28], which modified only 1 or 2 least significant bits (LSBs) of the host signal, are based on the same principle in rejecting the host signal interference, so called low bit modulation (LBM). For example, a method which modifies only 2 LSBs may be considered as a form of quantization index modulation where the step size of quantizer used is 4. Even-odd modulation is another embedding technique that operates similarly. In the data hiding scheme proposed by Wang, *et al.*, [29], the significant wavelet coefficients are modified such that they *quantize* to an even or odd value depending on the bit to be embedded. In [30], Wu, *et al.*, introduced a similar scheme based on JPEG quantizers by altering the DCT coefficients.

The additive spread spectrum and quantization based methods have poor performances for the "no attack" and "severe attack" cases, respectively. In the former, the performance becomes independent of the additive attack level. Whereas in the latter, the performance drops rapidly with the increase in the attack. These deficiencies point out to a non-optimal design procedure compared to Costa's scheme which can deliver perfect host signal interference rejection at all attack levels. The need for a class of practical methods where the hider has better control over the operating characteristics is immediately recognized by various researchers.

In quantization based data hiding methods, this effort resulted by incorporating a *processing* stage that follows the embedding quantization and by employing forms

of redundancy coding. In [8] and [31], Chen, *et al.* respectively introduced distortion compensated version of QIM (DC-QIM) (that can achieve the capacity under AWGN attacks), and spread transform (ST) technique for practical implementations (that embeds the message signal by spreading the resulting embedding distortion over many host signal coefficients). Ramkumar, *et al.* [20], considering scalar embedding, employed a thresholding type of processing at the embedder and, also, used a continuous triangular periodic function for extracting the embedded binary watermark signal. In [21], Eggers, *et al.* optimized the performance of DC-DM by a more careful optimization of embedding-detection parameters. They also combined multi-level signaling with binary coding techniques for low attack applications, and provided some performance results, [5, 32]. Perez-Gonzalez, *et al.* [33] proposed a probability density function (pdf) transformation type of processing for embedding. Furthermore, they provided a calculation of upper bound on the probability of error for multidimensional embedding case considering various noise distributions.

In order to improve the performance of additive spread spectrum methods, a similar approach to quantization based methods is also developed. Reference [33], inspired by ST-DM, proposed a decoding technique that integrates the underlying principles of quantization based methods with the additive schemes. In this method, watermark signal is selected such that when the linear correlation between the watermark signal and the undistorted stego signal is quantized, the resulting signal is a centroid of the lattice associated with the embedded signal. The probability of error performance of this method is improved by further processing. Consequently, the watermark signal is selected such that, rather than the quantized correlation metric itself, the properly scaled error due to quantization of the correlation metric is mapped to the desired centroid. Similarly, in [34], the watermark signal energy is properly shaped to compensate for the host signal interference at the detector. This is achieved by designing the weighting as a function of the projection of the

host signal onto the spreading sequence, so that at the detector, host signal's effect is diminished.

1.3 Dissertation Overview

This dissertation is a study of theory and practice of oblivious data hiding with the emphasis on efficient embedding and detection techniques. The dissertation is organized as follows. Chapter 2 starts with a discussion on the theory of *communication with side information* with reference to earlier works in the field. Then, an alternate communications framework is proposed from a data hiding perspective, and the duality between the communications and data hiding frameworks is elaborated from this point of view. Finally, codebook design and generation for data hiding methods is addressed.

In Chapter 3, the intricacies of the high performance embedder-detector design is explored in terms of the proposed framework assuming mean squared error distortion measure. The performance evaluation criteria needed for a fair comparison of those methods is laid out as: the type of post-processing, the type of demodulation, and the optimization criterion used to determine embedding-detection parameters. Various practical embedding-detection schemes are compared with respect to their rate, correlation, and probability of error performances under AWGN attacks.

Chapter 4 discusses and investigates the techniques for boosting the performance of embedding-detection techniques for the two extreme cases of large and small embedding signal sizes. These methods are the spread transforming and multiple codebook hiding, respectively, corresponding to cases where the embedding signal size is large and limited. General form of spread transforming for an arbitrary spreading gain is given with a transform domain embedding approach. Multiple codebook hiding method is introduced. The use of multiple codebooks offers freedom in the choice of the codeword that is more "friendly" with the host signal, especially when

the embedding signal size is small. In proposed scheme, each codebook is designed by the use of a real unitary transformation selected from a set of transformations that is known to both embedder and detector.

Chapter 5 proposes scalar quantization based embedding-detection methods against cropping and compression type of non-invertible attacks. Attacks on the stego signals can be classified into two main groups, namely, invertible and non-invertible attacks. Invertible attacks can be reversed by some intelligent and usually computationally intense manipulations. Therefore, hiding rate is not decreased. On the other hand, non-invertible attacks like cropping, AD-DA conversion, and compression may lead to insignificant hiding rates if they are not taken into account in advance by the designer. A true watermark embedding methodology should either be invariant to these attacks or include practical means of undoing and reducing the disturbing effects of them. In Section 5.1, a method to recover the message signal from a stego content that has undergone cropping and resampling consecutively is presented. The information loss due to the cropping is coped with by multiple embedding of the watermark signal, and the synchronization is restored by using cyclic autocorrelation features of the cropped-resampled signal and redundancy coding. In Section 5.2, embedder-detector operation is modified to make use of the compression scheme's quantization characteristics (*i.e.* quantization tables) assuming information hider has access to details of the compression algorithm prior to embedding. This is achieved by fine tuning the embedding-detection parameters to minimize the disturbing effects of the quantization noise.

Conclusions are given in Chapter 6.

CHAPTER 2

COMMUNICATION WITH SIDE INFORMATION AND DATA HIDING

Shannon [35], introduced the first analysis of discrete memoryless channels with side information, in the form of varying channel states from a finite set, causally known to the encoder. He proved that this channel is equivalent (in terms of capacity) to a usual memoryless channel that has the same output alphabet and an expanded input alphabet with no side information. Accordingly, each letter of the new input alphabet is generated as a mapping from the set of states into the input alphabet of the original channel. In [36], Kusnetsov *et al.* examined a practical version of the same problem where the errors in the channel are invariant, namely memory with defective cells. They offered an encoding scheme for reliable storage of information when the encoder is given the defect information, and they investigated the redundancy bounds for such codes. Gelfand, *et al.* in [37] considered a similar channel as in [35] by removing the causality condition on the encoder such that, at any transmission time, the encoder has the whole channel state information for all times. They proceeded to derive the capacity of this channel assuming an input alphabet $\mathcal{X}$, an output alphabet $\mathcal{Y}$, an auxiliary alphabet $\mathcal{U}$, and a finite set $\mathcal{C}$ of side information where $\mathcal{X}, \mathcal{Y}, \mathcal{U}, \mathcal{C}, \in \Re^N$. The channel capacity, C_0, is expressed in terms of random variables $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, $U \in \mathcal{U}$, and $C \in \mathcal{C}$ by a maximization over all conditional joint probability distributions $\mathrm{p}_C(c)\mathrm{p}_{U,X}(u,x|c)\mathrm{p}_Y(y|x,c)$ as

$$C_0 = \max_{\mathrm{p}(u,x|c)} (I(U,Y) - I(U,C)), \tag{2.1}$$

where $\mathrm{p}_X(x)$ is the probability mass function of a random variable X and $I(X,Y)$ is the mutual information between two random variables X and Y. Heegard, *et al.*

[38] also using this formulation, extended the idea to establish achievable storage rates for memory when defect information is given only to encoder or to decoder and completely to decoder but partially to encoder.

Costa [17], applied the results of [37] to memoryless channels with discrete time and continuous alphabets, and presented an information-theoretic analysis of a problem that also applies to oblivious data hiding. He studied a communications scenario where encoder transmits a message index to decoder in the presence of a side information and designed the auxiliary variable in Gelfand's formulation as $U = X + \alpha C$, where X is the power constrained input, C is the channel state information available at the encoder, and α is a scaling factor. Costa showed that for an additive white Gaussian noise (AWGN) channel with Gaussian input and side information, the channel capacity does not depend on the side information.

Later research gained considerable momentum first by reinterpreting these results in terms of oblivious data hiding, and later, by formulating the problem from a game theoretic perspective. References [39] and [40] assumed Gaussian distributed host signal and squared error distortion measure, and studied the problem as a data hiding game between the hider-extractor and attacker. In [39], Moulin, *et al.* introduced an information-theoretic model for data hiding considering memoryless attacks. In their model, the information hider determines the embedding strategy without knowing the attack, whereas the attacker uses the stego signal to design the attack. The extractor, on the other hand, is assumed to be in a position to learn the strategy of the attacker. It is shown that for squared error distortion measure and white Gaussian distributed host signal, Gaussian test channel is the optimal attack and the hiding capacity is the same as in the case when the host signal is known to the detector. They also showed that Costa's results are valid for this setting of the data hiding game under the small distortions scenario which assumes host signal power is much higher than that of the distortions introduced by the hider and attacker. Cohen,

et al. [40] presented a detailed discussion and the results of hiding capacity assuming Gaussian distributed host signal and squared error distortion measure, similar to [39], except the removal of the assumption that extractor knows the attack. They showed that independent, identically distributed (*iid*) Gaussian host signal maximizes the hiding capacity among all finite fourth moment distributions for the host signal. It is also discussed that additive attacks are sub-optimal. Furthermore, they extended Costa's results by considering non-white noise attacks and non-Gaussian embedding distortions.

These studies showed that the solution for the hiding capacity varies with the setting of the game, and Costa's framework yields the upper bound on the coding capacity among all versions of the game, since attacker has a fixed strategy (additive noise) that is known to both encoder and decoder. Therefore, Costa's framework and his results serve as a test-bed for comparing and evaluating the performances of various practical embedding-detection techniques.

2.1 Costa's Framework

Costa in [17], based on the results of [37], considered a power constrained AWGN channel with *iid* Gaussian input $\mathbf{X}$ and side information $\mathbf{C}$ (in the form of channel state) that is available *only* at the encoder in a non-causal manner. A message index m is transmitted to the receiver by properly selecting the codeword $\mathbf{X}$ that is distorted during transmission by the additive channel state $\mathbf{C}$ and the channel noise $\mathbf{Z}$. Consequently, the channel output is defined as $\mathbf{Y} = \mathbf{X} + \mathbf{C} + \mathbf{Z}$. Considering the design of $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$, $0 < \alpha < 1$, and assuming $\mathbf{X}$, $\mathbf{C}$, $\mathbf{Z}$ are *iid* length N sequences of random variables with zero covariance matrices and Gaussian marginal distributions (*i.e.* $X \sim \mathcal{N}(0, P)$, $C \sim \mathcal{N}(0, \sigma_C^2)$, $Z \sim \mathcal{N}(0, \sigma_Z^2)$), the communication rate is computed as [17]

$$R(\alpha) = I(U, Y) - I(U, C),$$

$$
\begin{aligned}
&= H(X+C+Z) - H(X+C+Z|X+\alpha C) \\
&\quad -H(X+\alpha C) + H(X+\alpha C|C), \\
&= H(X+C+Z) + H(X) - H(X+C+Z, X+\alpha C), \qquad (2.2)
\end{aligned}
$$

where $H(X)$ is defined as the entropy of random variable X. Since X, C and Z are assumed independent Gaussian random variables, $X+\alpha C$ and $X+C+Z$ are respectively distributed as $\mathcal{N}(0, P+\alpha^2\sigma_C^2)$ and $\mathcal{N}(0, P+\sigma_C^2+\sigma_Z^2)$. The joint distribution of $X+C+Z$ and $X+\alpha C$ is also Gaussian with the density function given as

$$
f_{X+C+Z,X+\alpha C}(x+c+z, x+\alpha c) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} P+\sigma_C^2+\sigma_Z^2 & P+\alpha\sigma_C^2 \\ P+\alpha\sigma_C^2 & P+\alpha^2\sigma_C^2 \end{bmatrix}\right).
$$

Hence, the rate in Equation (2.2) is obtained by calculating the entropies for the corresponding distributions as [41]

$$
R(\alpha) = \frac{1}{2}\log_2 \frac{P(P+\sigma_C^2+\sigma_Z^2)}{P\sigma_C^2(1-\alpha)^2 + \sigma_Z^2(P+\alpha^2\sigma_C^2)}. \qquad (2.3)
$$

Maximizing $R(\alpha)$ over α, Costa showed that communication rate achieves $\frac{1}{2}\log_2(1+\frac{P}{\sigma_Z^2})$ bits per transmission for $\alpha^* = \frac{P}{P+\sigma_Z^2}$ that is the capacity of the same AWGN channel with the side information available to both encoder and decoder. Thus, for a properly chosen α, the lack of side information at the decoder does not reduce the capacity.

The channel model for Costa's framework is displayed in Figure 2.1. In order to transmit message m, encoder E generates the codeword $\mathbf{X}$ that is additive to the channel state $\mathbf{C}$ at the given channel noise variance. Decoder D, not knowing the random channel state $\mathbf{C}$, detects the message $\hat{m}$ from the received signal $\mathbf{Y}$.

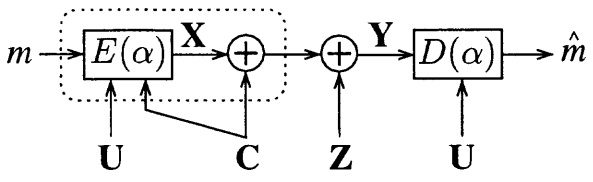


Figure 2.1 The channel model for Costa's framework corresponding to codebook design of $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$.

Costa outlined the capacity achieving encoding-decoding scheme based on random coding techniques. The optimal codebook has $M = \lfloor 2^{NR} \rfloor$ [1] codewords corresponding to M messages. Each message is transmitted in N uses of the channel. For optimal encoding and decoding, $2^{N(I(U,Y)-\epsilon)}$ (for an arbitrarily small ϵ) number of length N *iid* sequences with individual distributions $\mathcal{N}(0, P + \alpha^{*2}\sigma_C^2)$ are generated and then partitioned into 2^{NR} bins. Each bin is associated with the index of a message and points to $2^{N(I(U,C)+\epsilon)}$ number of sequences. This collection of sequences is made known to both encoder and decoder. In order to generate the codeword, the side information $\mathbf{C}$ is weighted by the proper α and subtracted from the sequences in the bin corresponding to the message to be conveyed. Among the resulting signals, the one that is orthogonal to $\mathbf{C}$ ($|(\mathbf{U}_j - \alpha^*\mathbf{C})^T\mathbf{C}| < \delta$, $j = 1, \ldots, 2^{N(I(U,C)+\epsilon)}$, for a proper δ value) and also satisfies the power constraint ($\frac{1}{N}||\mathbf{X}||^2 \leq P$) is the optimal codeword corresponding to message index being sent.

Encoder sends the codeword over the channel. Decoder receives the signal $\mathbf{Y}$ and searches over all $\mathbf{U}$ sequences for the jointly typical $(\mathbf{U}_j, \mathbf{Y})$ pair ($|(\mathbf{U}_j - \alpha\mathbf{Y})^T\mathbf{Y}| < \delta$, $j = 1, \ldots, 2^{N(I(U,Y)-\epsilon)}$). The sent message is decoded successfully from the $\mathbf{U}_j$ sequence and the received signal $\mathbf{Y}$, for $\alpha = \alpha^*$ and large N, as

$$|(\mathbf{U}_j - \alpha\mathbf{Y})^T\mathbf{Y}| = |(\mathbf{U}_j - \alpha^*\mathbf{C} - \alpha^*\mathbf{X} - \alpha^*\mathbf{Z})^T(\mathbf{X} + \mathbf{C} + \mathbf{Z})|, \quad (2.4)$$

$$= |(1 - \alpha^*)\mathbf{X}^T\mathbf{X} - \alpha^*\mathbf{Z}^T\mathbf{Z}|, \quad (2.5)$$

$$= (1 - \alpha^*)NE[X^2] - \alpha^*NE[Z^2], \quad (2.6)$$

[1] $\lfloor x \rfloor$ is the greatest integer smaller than or equal to x

$$= N\left(1 - \frac{P}{P+\sigma_Z^2}\right)P - \frac{NP}{P+\sigma_Z^2}\sigma_Z^2 = 0. \quad (2.7)$$

The message index associated with the bin that contains the sequence $\mathbf{U}_j$ is declared as the sent message. Such a code generation is asymptotically optimal as $N \to \infty$ [17].

2.2 An Alternate Framework Based on Channel Adaptive Encoding and Channel Independent Decoding (CAE-CID)

For the same communications scenario, let the channel model of Costa's framework be modified in two respects. First modification is by redefining the channel input as $\mathbf{X}_n = \mathbf{X} - \mathbf{X}_t$. The term $\mathbf{X}_t$ will be referred to as "processing distortion" since it is by nature, a "disturbance" to encoder output $\mathbf{X}$. The processing distortion $\mathbf{X}_t$ may be a function of the encoder output $\mathbf{X}$, and the correlation between $\mathbf{X}$ and $\mathbf{X}_t$ is denoted by ρ. Also, $\mathbf{X}_t$, like $\mathbf{X}$ is *iid* and independent of $\mathbf{C}$. In the CAE-CID framework, since the codeword transmitted by the encoder is $\mathbf{X}_n$, the power constraint that needs to be satisfied by the codeword $\mathbf{X}$ in Costa's framework, applies to $\mathbf{X}_n$, *viz.*, $\frac{1}{N}||\mathbf{X}_n||^2 \leq P$. Consequently, the received signal at the decoder is expressed as $\mathbf{Y} = \mathbf{X}_n + \mathbf{C} + \mathbf{Z}$. Second modification is by designing the shared variable as $\mathbf{U} = \mathbf{X} + \mathbf{C}$, where the α value employed in codebook generation is set to one regardless of the channel's noise level.

The transmission rate for the modified channel can now be computed for $U = X + C$, $X_n = X - X_t$, and $Y = X_n + C + Z$ as

$$\begin{aligned} R &= I(U,Y) - I(U,C), \\ &= I(X+C, X_n+C+Z) - I(X+C,C), \\ &= H(X_n+C+Z) - H(X_n+C+Z|X+C) - H(X+C) + H(X+C|C), \\ &= H(X_n+C+Z) - H(Z-X_t|X+C) - H(X+C) + H(X), \end{aligned}$$

$$= H(X) + H(X_n + C + Z) - H(Z - X_t, X + C). \tag{2.8}$$

The formulation given in Equation (2.8) can be solved for rate R assuming random variables X, X_t, C, and Z are mutually independent except for the known dependence between X and X_t, and they are distributed according to $\mathcal{N}(0, \sigma_X^2)$, $\mathcal{N}(0, \sigma_{X_t}^2)$, $\mathcal{N}(0, \sigma_C^2)$, and $\mathcal{N}(0, \sigma_Z^2)$, respectively. The normalized correlation between X and X_t is defined as

$$\rho = \frac{E[XX_t]}{\sqrt{E[X^2]E[X_t^2]}}. \tag{2.9}$$

On the other hand, X_n is a random variable with the second moment set to P and its distribution depends on how X_t is related to X. Furthermore, the random variables $Z - X_t$ and $X + C$ are jointly Gaussian with the probability density function given by

$$f_{Z-X_t, X+C}(z - x_t, x + c) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_Z^2 + \sigma_{X_t}^2 & E[XX_t] \\ E[XX_t] & \sigma_X^2 + \sigma_C^2 \end{bmatrix} \right) \tag{2.10}$$

Consequently, the rate in Equation (2.8) is derived by computing the entropies for the marginal and joint distributions as [41]

$$R(\sigma_X, \sigma_{X_t}, \rho) = \frac{1}{2} \log_2 \left(\frac{\sigma_X^2 (P + \sigma_C^2 + \sigma_Z^2)}{(\sigma_X^2 + \sigma_C^2)(\sigma_{X_t}^2 + \sigma_Z^2) - E[XX_t]^2} \right). \tag{2.11}$$

Using Equation (2.9), Equation (2.11) can be rewritten as

$$R(\sigma_X, \sigma_{X_t}, \rho) = \frac{1}{2} \log_2 \left(\frac{\sigma_X^2 (P + \sigma_C^2 + \sigma_Z^2)}{(\sigma_X^2 + \sigma_C^2)(\sigma_{X_t}^2 + \sigma_Z^2) - \rho^2 \sigma_X^2 \sigma_{X_t}^2} \right). \tag{2.12}$$

The achievable transmission rate for this channel can be found by maximizing the rate R over σ_X, σ_{X_t}, and ρ under the constraint $\frac{1}{N} \| \mathbf{X} - \mathbf{X}_t \|^2 = P$. Since ρ is a normalized variable, it does not depend on the variances of X and X_t. Hence, setting $\rho = 1$ (X_t is a linear function of X) will maximize Equation (2.12) in ρ. Moreover,

the power constraint on the input relates σ_X and σ_{X_t} as

$$\sigma_{X_t} = \begin{cases} \sigma_X - \sqrt{P}, \text{ if } \rho = 1 \\ \rho\sigma_X - \sqrt{\sigma_X^2(\rho^2 - 1) + P}, \text{ if } \rho \neq 1. \end{cases} \tag{2.13}$$

As a result, maximization of rate given in Equation (2.12) reduces to a maximization over σ_X for $\rho = 1$ and $\sigma_{X_t} = \sigma_X - \sqrt{P}$. Then,

$$\max_{\sigma_X} R(\sigma_X, \sigma_{X_t} = \sigma_X - \sqrt{P}, \rho = 1) = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_Z^2}\right)\Big|_{\sigma_X = \sigma_X^*} \tag{2.14}$$

which is maximized for

$$\sigma_X^* = \frac{P + \sigma_Z^2}{\sqrt{P}}, \quad \sigma_{X_t}^* = \frac{\sigma_Z^2}{\sqrt{P}}. \tag{2.15}$$

This is the capacity of the AWGN channel where the side information is also known to the decoder, as first derived by Costa [17]. The results above show that the optimal codebook design in Costa's framework based on a particular α^* can be equivalently achieved in the CAE-CID framework with the corresponding σ_X^* when $\rho = 1$. Therefore, the two frameworks are equivalent, and they can be translated into each other through $\sigma_X^* = \frac{\sqrt{P}}{\alpha^*}$ at the same transmission rate. The corresponding channel model for the proposed CAE-CID framework is displayed in Figure 2.2. When compared with Figure 2.1, main difference is that α dependency of (E, D) pair is replaced by the inclusion of $\mathbf{X}_t$ that is generated by the processing $\mathcal{P}$ at the encoder.

Optimal encoding-decoding scheme of the CAE-CID framework is similar to the one described in [17]. However, the encoding-decoding operations rely on the design of $\mathbf{U} = \mathbf{X} + \mathbf{C}$ as α is set to one. Correspondingly, the shared $\mathbf{U}$ sequences are *iid* with an underlying marginal distribution $\mathcal{N}(0, P + \sigma_C^2)$. The channel dependence, however, is reflected in the appropriate choice of processing that generates $\mathbf{X}_t$ from $\mathbf{X}$. At the encoder, for the given $\mathbf{C}$, the jointly typical $(\mathbf{U}, \mathbf{C})$ pair is searched in the bin corresponding to the message signal being sent. The codeword is generated from the $\mathbf{U}_j$ sequence that satisfies the orthogonality constraint ($|(\mathbf{U}_j - \mathbf{C})^T\mathbf{C}| <$

δ, $j = 1, \ldots, 2^{N(I(U,C)+\epsilon)}$) and yields codeword $\mathbf{X}_n$ such that the power constraint ($\frac{1}{N}||\mathbf{X}_n||^2 \leq P$) is satisfied. It should be noted that, in order to achieve capacity, $\mathbf{X}_t$ is a linear function of $\mathbf{X}$. Therefore, the codeword $\mathbf{X}_n$ is readily obtained from the encoder output $\mathbf{X}$ by the relation $\mathbf{X}_n = \frac{\sqrt{P}}{\sigma_X}\mathbf{X}$.

On the decoder side, the sent message is decoded as the index of the bin that contains the $\mathbf{U}$ sequence which is jointly typical with the received signal $\mathbf{Y}$. The particular sequence $\mathbf{U}_j$ is found, for large N, as

$$\begin{aligned} |(\mathbf{U}_j - \mathbf{Y})^T\mathbf{Y}| &= |\left(\mathbf{U}_j - (\mathbf{X} - \mathbf{X}_t + \mathbf{C} + \mathbf{Z})\right)^T (\mathbf{X} - \mathbf{X}_t + \mathbf{C} + \mathbf{Z})|, \\ &= |\mathbf{X}_t^T\mathbf{X} - \mathbf{X}_t^T\mathbf{X}_t - \mathbf{Z}^T\mathbf{Z}|, \qquad (2.16) \\ &= NE[XX_t] - NE[X_t^2] - NE[Z^2], \qquad (2.17) \\ &= N\frac{P + \sigma_Z^2}{\sqrt{P}}\frac{\sigma_Z^2}{\sqrt{P}} - N\frac{(\sigma_Z^2)^2}{P} - N\sigma_Z^2 = 0, \qquad (2.18) \end{aligned}$$

where $E[XX_t] = \sigma_X^*\sigma_{X_t}^*$, Equation (2.9) for $\rho = 1$, is used. The cancellation of the terms in Equation (2.18), completely relies on the choice of $\mathbf{X}$ and the corresponding $\mathbf{X}_t$ at the encoder.

In CAE-CID framework, since the design of the shared variable is fixed as $\mathbf{U} = \mathbf{X} + \mathbf{C}$, the optimal encoding and decoding merely relies on the proper of statistics of the encoder output $\mathbf{X}$ and its dependence with processing distortion $\mathbf{X}_t$.

2.2.1 Advantages of CAE-CID Framework

When compared to Costa's framework, the CAE-CID framework has the following advantages:

1. In Costa's scheme, both the encoder and decoder need to know the channel noise variance, while for the CAE-CID scheme only the encoder needs to know the channel noise variance. The channel dependent nature of the encoding, for the CAE-CID framework, is reflected on both inputs $\mathbf{X}$ and $\mathbf{X}_t$. Thus, channel state interference rejection at the decoder is achieved solely by the encoder's

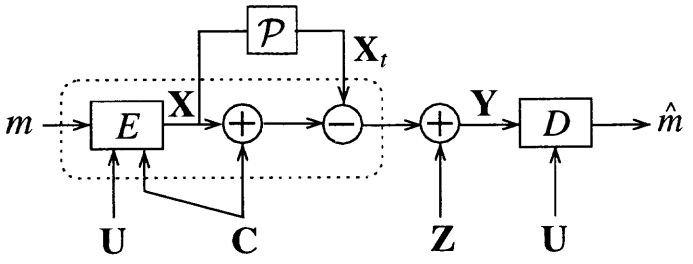


Figure 2.2 The channel model for the proposed CAE-CID framework corresponding to codebook design of $\mathbf{U} = \mathbf{X} + \mathbf{C}$.

ability to properly select σ_X and σ_{X_t} depending on the given σ_Z, Equation (2.15).

2. When the channel noise variance changes, in Costa's framework, successful decoding can no longer be sustained due to dependence of decoder on the channel noise level. However, in the CAE-CID framework, if the channel noise level changes, encoder-decoder can continue successful operation at a lower or higher rate by adjusting P at the encoder without updating the shared collection of $\mathbf{U}$ sequences as long as

$$\sigma_X^2 \geq 2\hat{\sigma}_Z \tag{2.19}$$

where $\hat{\sigma}_Z^2$ is the new channel noise power (derivation details are given in Appendix A).

3. CAE-CID framework provides a better theoretical basis for practical embedder-detector designs, as the post-processing, employed in practical methods, can be represented by the processing distortion term $\mathbf{X}_t$ in the formulations.

2.3 On the Duality of Communications and Data Hiding Frameworks

The theory of data hiding has been developed mainly through employing analytical tools of *communication with side information* and *spread spectrum communications.*

Table 2.1 Duality Between Communications and Data Hiding Frameworks

Communications Framework	**Data Hiding Framework**
Side information	Host signal
Encoder-Decoder	Embedder-Detector
Channel noise	All forms of modification on the stego signal (Attack)
Power constraints	Perceptual distortion limits
Bandwidth	Embedding signal size
Signal to Noise ratio	Embedding distortion to attack distortion ratio

This is achieved by reinterpreting and adapting basic concepts such as channel, side information, and power constraints within the context of data hiding.

In data hiding, channel is the medium between the hider and extractor, and it includes all forms of disturbances that affect the *stego* signal, which is an intelligent combination of the host signal and the message to be conveyed. Side information available at the encoder in a communication channel model, is associated with the host signal at the embedder in the equivalent data hiding model. Similarly, encoder-decoder pair (E, D) is functionally equivalent to embedder-detector pair $(\mathcal{E}, \mathcal{D})$. Power constraints in a channel communication scenario are analogous to the perceptual distortion limits that are determined based on the features of the host signal. The bandwidth is somewhat dual to embedding signal size as they are both resources of the communication, and signal to noise ratio (SNR) measure corresponds to embedding distortion to attack distortion ratio (WNR) measure. Table 2.3 shows the duality between the communications and data hiding frameworks.

Based on the communication frameworks given in Sections 2.1 and 2.2, encoding and decoding of a message index relies on proper selection of the codeword. Correspondingly, in the dual data hiding problem, the performance of an embedding and detection technique depends on the underlying codeword generation scheme. Hence, main goal of a data hiding method is to design practical codebook and codeword generation schemes that can deliver perfect host signal interference rejection at all noise levels.

A codebook is a collection of mappings from the set of messages to be conveyed. Each mapping, or codeword, is generated from the host signal by an intelligent process based on the imposed distortion constraints and the expected noise level. However, in the formulations of data hiding, a codeword is defined in two different ways. From the communications point of view, the side information is a state of the channel and the codeword is the signal transmitted through the channel. Then, due to the analogy with the communications framework, a codeword can be defined as the distortion introduced to the host signal due to the embedding operation. However, within the context of data hiding, side information is the host signal, and it is also transmitted through the channel. Correspondingly, one can define the stego signal to be the codeword, as it is the channel input. In order to better exploit the duality between the communications and data hiding frameworks, the former definition for codeword is adopted.

A typical data hiding system can be modeled as

$$\begin{array}{lll} \text{Embedding:} & \mathcal{W}:\ m \longrightarrow \mathbf{W}, & \\ & \mathbf{S} = \mathcal{E}(\mathbf{C}, \mathbf{W}), & \\ \text{Attack:} & \mathbf{Y} = \mathbf{S} + \mathbf{Z}, & \\ \text{Detection:} \quad \hat{m} = \mathcal{D}(\mathbf{Y}) & \text{or} & \hat{\mathbf{W}} = \mathcal{D}(\mathbf{Y}), \\ & & \mathcal{W}^{-1}:\ \hat{\mathbf{W}} \longrightarrow \hat{m}. \end{array} \tag{2.20}$$

where detector is assumed to have no access to the host signal during the extraction process. In the above model, m is the message to be hidden, $\mathbf{C}$ is the host signal, $\mathbf{W}$ is the watermark signal, $\mathbf{S}$ is the stego signal, $\mathbf{Z}$ is the intrusion of the attacker, $\mathbf{Y}$ is the distorted stego signal, $\hat{\mathbf{W}}$ is an estimate of $\mathbf{W}$, and $\hat{m}$ is the detected message. At the embedder, message index m is mapped to a sequence of information samples $\mathbf{W}$ by the mapping $\mathcal{W}$ which transforms message m into a better representation for embedding. Then, the resulting watermark signal $\mathbf{W}$ is embedded into the host signal $\mathbf{C}$. At the detector, sent message is detected from the received signal $\mathbf{Y}$ or from an extracted estimate $\hat{\mathbf{W}}$ of $\mathbf{W}$ by the inverse mapping $\mathcal{W}^{-1}$. In the model, the embedder, $\mathcal{E}$, and the detector, $\mathcal{D}$, may be linear or nonlinear functions that operate on scalar or vector variables, and are not necessarily inverses of each other. Not evident in the model is the distortion constraints imposed on hider and attacker for keeping the host signal intact. Ideally speaking, the measure used to quantify the hider's and attacker's distortion is expected to be in compliance with the perceptual properties of the host signal.

Due to the duality between the communications and data hiding frameworks, the underlying encoder-decoder design principles of Sections 2.1 and 2.2 can be applied to embedder-detector design of data hiding methods. The corresponding encoding-decoding schemes assume the presence of a very large number of $\mathbf{U}$ sequences both at the encoder and decoder, and achieving channel capacity relies on adapting the codeword to the channel state at a given channel noise level. The encoding operation is simply a brute search in the bin pointed by the message index, in order to find the $\mathbf{U}$ sequence that yields the codeword in the direction of the host signal $\mathbf{C}$. Accordingly, each codeword is orthogonal to $\mathbf{C}$ and satisfies the power constraint P. (These constraints take the form of $\mathbf{X}^T\mathbf{C} \approx 0$ and $\frac{1}{N}||\mathbf{X}||^2 = P$ in Costa's framework and $\mathbf{X}_n^T\mathbf{C} \approx 0$ and $\frac{1}{N}||\mathbf{X}_n||^2 = P$ in CAE-CID framework.) At the decoder, on the other hand, the same $\mathbf{U}$ sequence is searched in all bins based on joint typicality

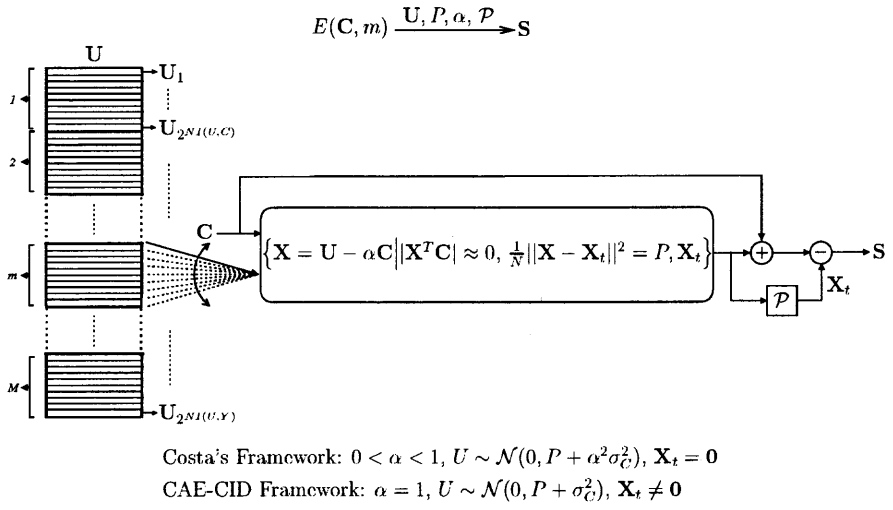


Figure 2.3 Encoding of message index m.

with the received $\mathbf{Y}$. Figures 2.3 and 2.4 depict the optimal encoding and decoding for message index m. In Costa's framework $0 < \alpha < 1$ and processing distortion is zero, whereas in CAE-CID framework $\alpha = 1$ and the processing distortion is non-zero. Hence, the main difference between the two frameworks is in *how the channel dependent nature is reflected in encoding and decoding operations.*

Despite their optimality, such encoding-decoding schemes cannot be applied to the design of practical embedding-detection techniques due to complexity issues. However, their structure has been an inspiration for the design of many embedder-detector pairs [18, 8, 20, 22, 21, 33]. Common to all these data hiding techniques is the use of quantization to simplify codebook generation and codeword selection. Also, they impose the power and orthogonality constraints in a less strict sense.

In quantization based methods, the optimal encoding-decoding procedure is effectively simplified by generating $\mathbf{U}$ sequences as sequences of reconstruction points where each reconstruction point is associated with a quantizer from a set of quantizers. The number of quantizers in the set corresponds to number of messages or message letters. Each quantizer of the set is uniquely described by a set of reconstruction points that are non-overlapping with other sets of reconstruction

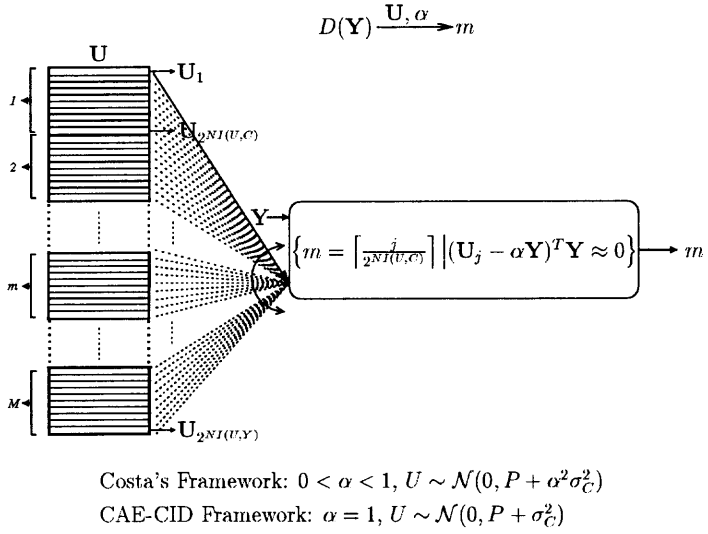


Figure 2.4 Decoding of sent message index m.

points. Therefore, each finite state of $\mathbf{U}$ is a sequence with values restricted to reconstruction values of the designated quantizers. The terms $\mathbf{X}$ and $\mathbf{X}_t$ are the embedding distortion due to quantization and the processing distortion, respectively. The codeword corresponding to a message is the distortion signal introduced to the host signal as a result of embedding operation, $\mathbf{S} - \mathbf{C}$. Consequently, it is denoted by $\mathbf{X}_n = \mathbf{X} - \mathbf{X}_t$ in the CAE-CID framework and by $\mathbf{X}$ in Costa's framework. The embedding operation, based on the CAE-CID framework, is the quantization of $\mathbf{C}$ vector with the quantizer(s) pointed by the watermark signal $\mathbf{W}$ to be embedded, and then processing the resulting quantized signal by a choice of (post-processing) function. Hence, input $\mathbf{X}$ in the CAE-CID framework is the distortion introduced to $\mathbf{C}$ due to quantization of embedding, and the processing distortion $\mathbf{X}_t$ is the result of processing $\mathcal{P}$, $\mathbf{X}_t = \mathcal{P}(\mathbf{X})$. The detection of the sent message, on the other hand, is by determining the nearest reconstruction point(s) to the received signal $\mathbf{Y}$, and generating the message by mapping the corresponding quantizer(s) to the message letters they are associated with. The crux of practical methods is that each codeword is directly generated from the given host signal and the watermark signal through quantization rather than maintaining a collection of shared $\mathbf{U}$ sequences.

Chou, *et al.*, in [22] applied the solution of a problem in distributed source coding to data hiding through the use of optimal quantizers. They proposed the use of robust optimization for codeword selection from Costa's huge codebook. In their work, the orthogonality of $\mathbf{C}$ and $\mathbf{X}$ is obtained by choosing $\mathbf{U}$ as a rate-distortion optimized and quantized version of a scaled version of $\mathbf{C}$. Although this approach approximates the optimal encoding and decoding scheme of Costa's framework, even the simplest implementations involve considerable complexity. Such complexity concerns draw attention to practical approaches with simpler implementations. Chen, *et al.*, Ramkumar, *et al.*, Eggers, *et al.*, and Perez-Gonzalez, *et al.* in [20, 8, 21, 33], respectively, proposed methods that handle codebook generation by uniform scalar quantization.

2.4 Codebook Generation for Data Hiding Methods

Practical data hiding approaches can be categorized into three main types within the frameworks studied in Sections 2.1 and 2.2 based on the design of embedder-detector pair, namely type-I, type-II, and type-III [3, 42]. Type-I methods refer to additive schemes where the stego signal is generated by adding the watermark signal to the host signal [12, 13, 14, 15, 16]. This type of methods suffer severely from host signal interference due to the non-optimal design that assumes the host signal $\mathbf{C}$ as a noise and tries to cancel it. Type-I methods have preferable performance only if channel noise is very strong or the host signal is available at the extractor.

Type-II methods are characterized by the use of quantization procedures and by the $(\mathcal{E}, D)$ pair which are exact inverses [25, 26, 27, 28, 29, 30, 23, 20]. The major drawback of this type of methods is that they perform well only if the attack is not severe. However, they are very suitable for oblivious data hiding applications with low noise levels.

Type-I and type-II methods correspond to designs of $\mathbf{U} = \mathbf{X}$, $\alpha = 0$, and $\mathbf{U} = \mathbf{X} + \mathbf{C}$, $\alpha = 1$, respectively, within Costa's framework. In the CAE-CID framework, however, corresponding designs for type-I and type-II methods take the form of $\mathbf{U} = \mathbf{X} + \mathbf{C}$ with the statistics of $\sigma_X^2 = \sigma_C^2 + \sigma_Z^2$ when $\rho = 1$, and $\sigma_X = \sqrt{P}$ when $\mathbf{X}_t = \mathbf{0}$, respectively. These two choices of designs for both frameworks correspond to two extreme cases in hiding rate vs. robustness curves. Namely, type-I methods are preferred for the case of "severe attacks" while type-II methods are superior for the case of "low attacks."

An optimal design is the one that designer has control over the operating characteristics of the method. In effect, this imposes some sort of dependency on the channel noise instead of the fixed severe noise (type-I) or low noise (type-II) assumptions. The type of methods that rely on this principle are called type-III which is a generalization of type-I and type-II. Codebook design of type-III methods follows $\mathbf{U} = \mathbf{X} + \mathbf{C}$ when $\rho = 1$ and $\mathbf{X}_t \neq \mathbf{0}$ within the CAE-CID framework, and $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$ where $0 < \alpha < 1$ within Costa's framework. Therefore, information hider has the freedom to adapt the codeword to the host signal at the presumed noise level. These methods are ideal for oblivious data hiding.

Type-III methods are developed from type-II methods by enhancing the functionality of type-II embedder with added processing, (*i.e.* thresholding, distortion compensation, Gaussian mapping) [20, 8, 21, 33]. In type-III methods, the post-processing is designed in a way that hiding rate is maximized for a presumed attack level [43]. However, codeword generation for most type-III methods does not explicitly follow Costa's framework due to the processing that takes place after quantization of the host signal. Therefore, type-III methods are better evaluated within the CAE-CID framework.

Table 2.4 summarizes the three types of methods. Based on the codebook designs, it is observed that type-I embedding does not exploit any information

on host signal or channel noise level. While type-II embedding exploits only host signal information. Type-III embedding, on the other hand, utilizes both forms of information.

Table 2.2 Three Types of Embedding-Detection Schemes

	Characterization	*Codebook Design*
Type-I	Additive schemes	$\mathbf{U} = \mathbf{X}$
Type-II	Quantization based schemes	$\mathbf{U} = \mathbf{X} + \mathbf{C}$
Type-III	Channel adaptive schemes	$\mathbf{U} = \mathbf{X} + \mathbf{C}$ with processing

Figures 2.5, 2.6, and 2.7, respectively, display the codeword generation of type-I, type-II and type-III methods for a set of watermark signals, denoted by $\mathbf{W}_1, \ldots, \mathbf{W}_M$, for the given host signal $\mathbf{C}$. In type-II and type-III methods, each message or watermark sample is assigned a particular quantizer $Q_\Delta(.)$. The base quantizer $Q_\Delta(.)$ may be a high dimensional vector quantizer or a Cartesian product of scalar quantizers with Δ as the distance between the reconstruction points. For type-II embedding, $\mathbf{C}$ is quantized with respect to the watermark signal, $Q_\Delta(\mathbf{C}, \mathbf{W})$. Consequently, the codeword $\mathbf{X}$ is the quantization error introduced to the host signal $\mathbf{C}$, $\mathbf{X} = Q_\Delta(\mathbf{C}, \mathbf{W}) - \mathbf{C}$. On the other hand in type-III methods, the quantization error (type-II codeword), undergoes the particular processing $\mathcal{P}$, which generates the codeword $\mathbf{X}_n = \mathbf{X} - \mathcal{P}(\mathbf{X})$. The post-processing function $\mathcal{P}$, may have the following forms

1. the distortion compensation [8, 21],
2. the thresholding [20], or
3. the Gaussian mapping [33].

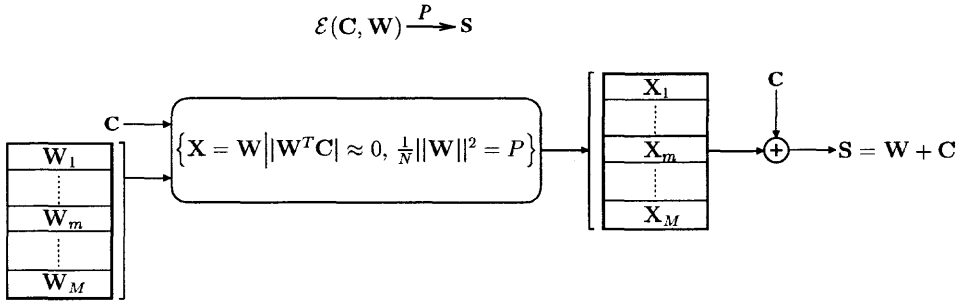


Figure 2.5 Encoding of message index m in type-I methods.

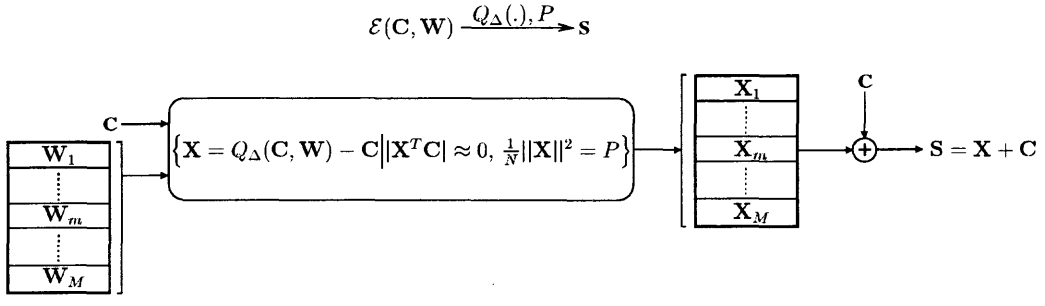


Figure 2.6 Encoding of message index m in type-II methods.

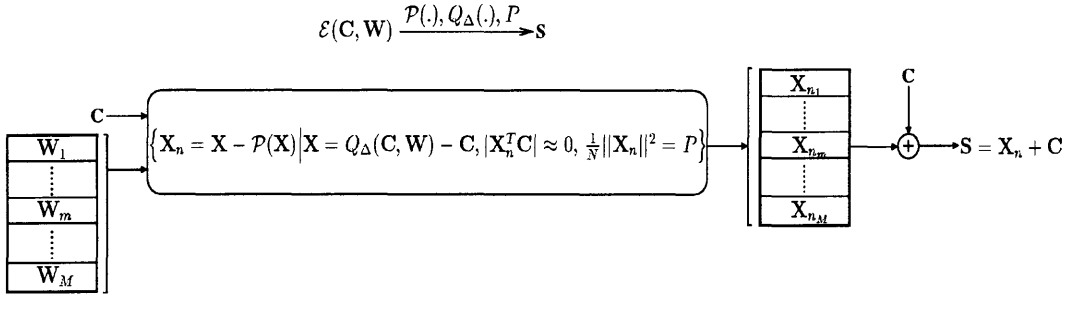


Figure 2.7 Encoding of message index m in type-III methods.

The performance of the three types of methods can also be judged by the structure of the corresponding detectors. Considering the very simple scenario where a two-level watermark sample is embedded to a signal coefficient and sent through a noisy channel, the three types of detectors take the following forms. The detector for the type-I scheme decides on the sent sample by comparing the received signal to a threshold. Whereas in type-II and type-III methods, detection of the embedded watermark sample is by some form of minimum distance decoding in order to determine the nearest reconstruction point to the received stego sample. Figure 2.8 displays the partitioning of the signal space between the two disjoint decision regions, $R_\times$ and $R_\circ$. In the figure, $\times$ and $\circ$ symbols denote the reconstruction points associated with the quantizers corresponding to two watermark samples. Obviously, the partitioning of the decision regions in type-I detector is far from being optimal when the channel noise level is low. This is because with a limited embedding distortion most (host) signal coefficients are not suitable for embedding (*i.e.*, in order to embed the information symbol denoted by $\circ$ to a host signal coefficient that is at the far left of the threshold, an arbitrarily large embedding distortion needs to be introduced to translate it to the region $R_\circ$). On the contrary, the layout of the decision regions of the type-II detector insure reliable detection from all stego coefficients, however, only up to channel distortions of power P. Type-III detector, on the other hand, gives control over the size of the decision regions, and as a result successful detection can be sustained up to noise level σ_Z^2 while embedding distortion is still limited to P as in type-II embedding. As the channel noise level σ_Z^2 increases the type-III detector will depart from the type-II detector and take the form of type-I detector.

Figures 2.9 and 2.10 display the hiding rate vs. robustness performances achievable by type-I, type-II and type-III methodologies computed using Equation (2.3) for $\alpha = 0$, $\alpha = 1$ and $\alpha = \frac{P}{P+\sigma_Z^2}$, respectively, or equivalently, solving Equations

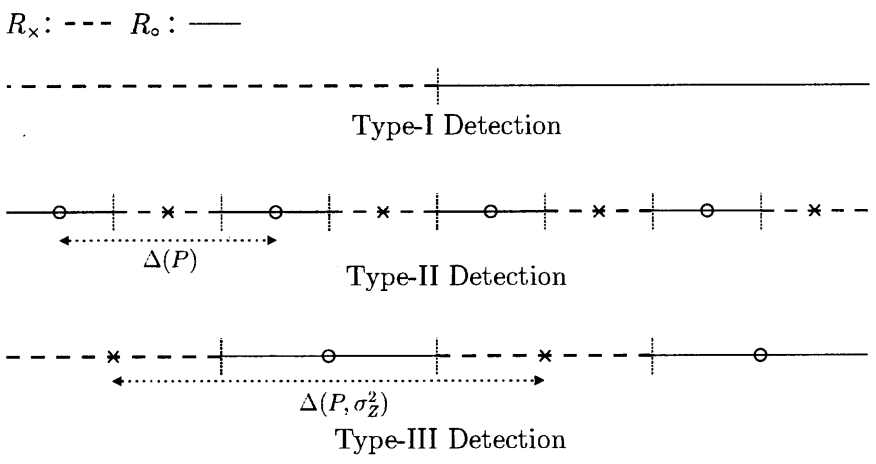


Figure 2.8 The partition of the signal space between decision regions $R_\times$ and $R_\circ$ corresponding to scalar embedding and detection of a binary signal.

(2.8) and (2.12) for $\sigma_X = \frac{P+\sigma_Z^2}{2\sqrt{P}}$ when $\rho = 1$, $\sigma_X = \sqrt{P}$ when $\mathbf{X}_t = \mathbf{0}$, and $\sigma_X = \frac{P+\sigma_Z^2}{\sqrt{P}}$ when $\rho = 1$.

The hiding rate is measured in the number of bits that can be hidden into a host signal coefficient, and the robustness measure is defined in terms of the ratio between the embedding distortion power and the channel noise power,

$$WNR = 10 \log_{10} \frac{P}{\sigma_Z^2} \text{ in dB.} \tag{2.21}$$

However, for type-I methods, WNR by itself can not be the indicator of the robustness as the host signal is considered to be a part of the noise. Therefore, another measure that can be considered is the ratio of the host signal power to embedding distortion power,

$$DWR = 10 \log_{10} \frac{\sigma_C^2}{P} \text{ in dB.} \tag{2.22}$$

In type-II methods, due to the ability to reject the host signal interference (depending on the WNR), the dependency of the performance to DWR level is weak. Type-I methods achieve the capacity at very low WNRs, and at high WNRs, there is almost a constant gap with the capacity. On the other hand, type-II methods

achieve the capacity at higher WNRs, and the hiding rate drops exponentially with the decreasing WNR. Furthermore, at low WNR range hiding is not possible. Since type-III is a superset of type-I and type-II methods, its optimal version can achieve the capacity at all WNRs.

A detailed analysis of type-I embedding-detection and capacity results can be found in [3, 44, 45, 46].

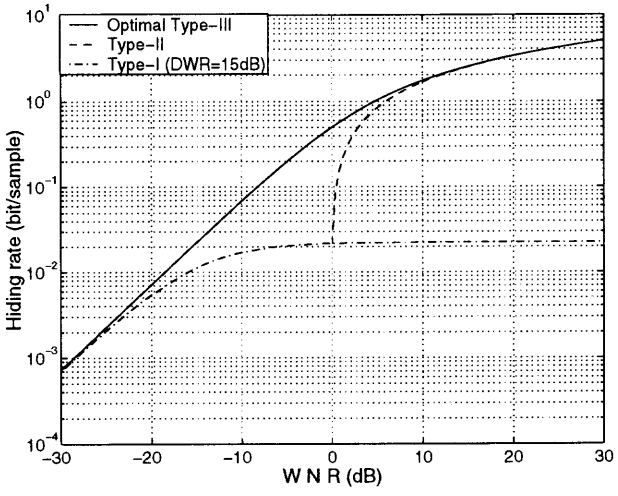


Figure 2.9 Hiding rate vs. robustness performance of type-I, type-II and type-III methods with $P = 10$ and DWR= 15dB.

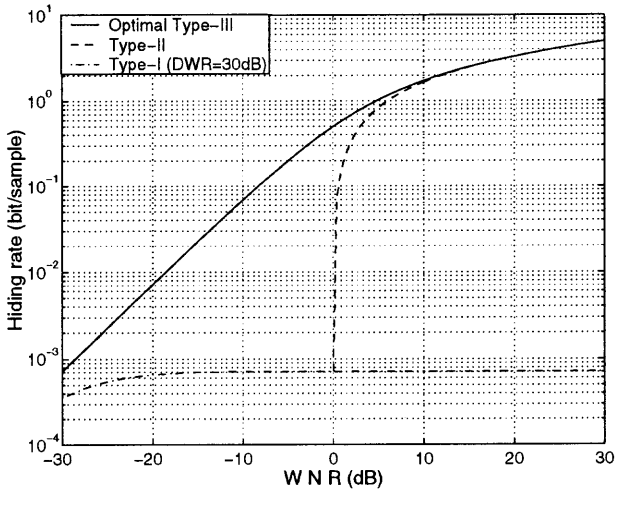


Figure 2.10 Hiding rate vs. robustness performance of type-I, type-II and type-III methods with $P = 10$ and DWR= 30dB.

CHAPTER 3

PERFORMANCE EVALUATION AND COMPARISON OF QUANTIZATION BASED EMBEDDING-DETECTION TECHNIQUES

Quantization based data hiding methods that rely on type-II and type-III [1] embedding-detection principles are studied together and compared based on three key characteristics as follows [47]:

1. the type of the distortion reduction technique (post-processing) employed in embedding;
2. the form of demodulation used (detection function);
3. the optimization criterion utilized in determining the embedding-detection parameters.

In the following sections, various type-II and type-III methods are examined and evaluated considering these three issues. The performance results for these methods, based on the above criteria, are provided in Section 3.3.

3.1 Type-II Embedding and Detection

The codebook generation for type-II methods is characterized by the design of $\mathbf{U} = \mathbf{X} + \mathbf{C}$ which corresponds to choice of $\alpha = 1$ within Costa's framework or $\mathbf{X}_t = 0$ ($\mathbf{X}_n = \mathbf{X}$) within the CAE-CID framework. framework. The generalized channel model for type-II hiding methods is displayed in Figure 3.1. In the model, $\mathbf{W}$ is the watermark signal corresponding to the message index m to be conveyed, $\mathbf{C}$ is the host signal, $\mathbf{X}$ is the codeword, $\mathbf{S}$ is the stego signal, $\mathbf{Z}$ is the additive noise (attack), and $\mathbf{Y}$ is the distorted stego signal at the detector defined as $\mathbf{Y} = \mathbf{S} + \mathbf{Z}$.

[1] Type-II can be considered as a special case of type-III where no post-processing is employed.

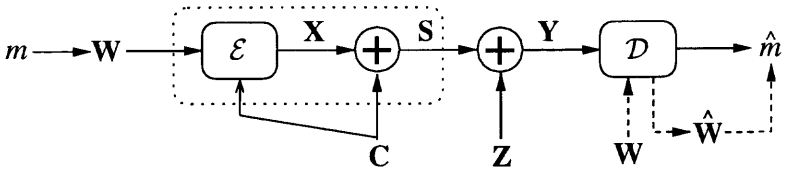


Figure 3.1 Block diagram of type-II embedding and detection stages.

The embedder, $\mathcal{E}$, imposes the power constraint as $\frac{1}{N}||\mathbf{X}||^2 = P$. At the detector, $\mathcal{D}$, the sent message $\hat{m}$ is detected from $\mathbf{Y}$ or from an extracted estimate $\hat{\mathbf{W}}$ of $\mathbf{W}$. Except the codebook design, type-II methods are also characterized by their $\mathcal{E}, D$ designs, which are exact inverses expressed as

$$\mathbf{S} = \mathcal{E}(\mathbf{C}, \mathbf{W}), \quad \mathbf{W} = \mathcal{D}(\mathbf{S}). \tag{3.1}$$

Chen, *et al.*, in [23], introduced QIM method which outlined the codeword generation for type-II methods. QIM achieves the upper bound on the hiding rate for low-level attacks (or high WNRs). In QIM method, embedding a message into a host signal refers to quantization of the host signal by a quantizer picked from an ensemble of quantizers, where each quantizer is associated with a message letter or message index. Thus, the stego signal $\mathbf{S}$ is a quantized form of $\mathbf{C}$, and the corresponding quantization error is the codeword $\mathbf{X}$. The number of quantizers in the ensemble determines the information embedding rate. The embedding distortion is measured using squared error distance measure, *viz.*, $\frac{1}{N}||\mathbf{X}||^2 = P$, and it varies with the size and shape of the quantization cells. The orthogonality constraint, $\mathbf{X}^T\mathbf{C} = 0$, however, is relaxed by assuming that $\mathbf{C}$ is uniformly distributed over all quantization cells and the number of quantization levels is not small such that $\mathbf{X}$ and $\mathbf{C}$ are approximately uncorrelated. This assumption also removes the dependence of embedding and detection operations on the host signal's statistics. In practice, this can be satisfied by the *small distortions scenario* where embedding and attack distortion powers are much less than the host signal power.

On the other hand, detection of a hidden message is achieved by the minimum distance decoder which computes the Euclidean distances of the received signal to surrounding reconstruction points. The message index associated with the nearest reconstruction point of the corresponding quantizer is regarded as the sent message. In QIM, embedding and detection are high-dimensional operations.

A practical implementation of QIM based on dithered quantizers, *viz.* dither modulation (DM), is presented and detailed in [23] and [24]. Dithered quantizers intend to decorrelate the quantization error of a quantizer from its input, [48]. In subtractive dithering, an *iid* dither vector (independent of the input) is added to the input prior to quantization, and then subtracted from the quantized output. Hence, the goal (decorrelation of the quantization error) is achieved. Within the context of data hiding, the dither signal is merely a mapping from the message index, the watermark signal. Therefore, the dither signal is not genuinely random and the orthogonality between the error and the input signals is not guaranteed. In DM, each quantizer in the ensemble is generated from a base quantizer by shifting the quantization cells and reconstruction points. The stego signal is generated by quantizing the host signal with the corresponding dithered quantizer as

$$\mathbf{S} = Q_{\Delta}(\mathbf{C} + \mathbf{W}_m) - \mathbf{W}_m \tag{3.2}$$

where $Q_{\Delta}(.)$ is the high dimensional base quantizer with reconstruction points Δ apart, and $\mathbf{W}_m$ is the watermark signal corresponding to message indexed by m, $1 \leq m \leq M$, where each component W_{m_i}, $1 \leq i \leq N$, of $\mathbf{W}_m$ is a representation from a set $\Omega \in \Re$. Consequently, the codeword $\mathbf{X}$ is defined as

$$\mathbf{X} = (Q_{\Delta}(\mathbf{C} + \mathbf{W}_m) - \mathbf{W}_m) - \mathbf{C}. \tag{3.3}$$

The power constraint on the embedding distortion $\mathbf{X}$ is controlled by adjusting the quantization step size Δ.

For the sake of practicality, $Q_\Delta(.)$ can be considered to be a product quantizer generated by a Cartesian product of N uniform scalar quantizers, $q_\Delta(.)$, each with step size Δ such that

$$q_\Delta(C) = i\Delta, \quad \text{for} \quad i\Delta - \frac{\Delta}{2} \leq C < i\Delta + \frac{\Delta}{2}. \tag{3.4}$$

Therefore, embedding can be viewed as N successive scalar quantization, of the coefficients of $\mathbf{C} = (C_1, \ldots, C_N)$, dithered with the watermark signal vector $\mathbf{W}_m = (W_{m_1}, \ldots, W_{m_N})$. Each distinct component of the watermark (dither) signal is associated with a quantizer that is generated by properly shifting the reconstruction points of $q_\Delta(.)$. The amount of shifting is determined by the number of possible values a watermark sample can take (the number of quantizers). For maximum separation of the reconstruction points of embedding quantizers, the watermark sample values are equally spaced along an interval of length that is equal to quantization step size Δ, *i.e.*, $[-\Delta/2, \Delta/2)$. It should be noted that, since the watermark signal is the subtractive dither signal, the sample values represented by the form $W_m + i\Delta$ for $i \in \mathcal{Z}$, where $\mathcal{Z}$ is the set of all integers, lead to the same dithered quantizer. (In other words, shifts differing by an integer multiple of Δ correspond to the same quantizer.) Considering a d-ary watermark sample, the set Ω that contains the d possible sample values is defined as

$$\Omega = \left\{ \delta + i\Delta, \delta + \frac{\Delta}{d} + i\Delta, \delta + 2\frac{\Delta}{d} + i\Delta, \ldots, \delta + (d-1)\frac{\Delta}{d} + i\Delta \right\} \tag{3.5}$$

where δ is a uniform random variable in $[-\frac{\Delta}{2}, \frac{\Delta}{2})$ and $i \in \mathcal{Z}$. As a result, reconstruction points and quantization cells of each quantizer in the ensemble are shifted by $\frac{\Delta}{d}$ with respect to each other. The reconstruction points of the embedding quantizers are also known to the detector for the extraction of the sent message. At the detector, the hidden message is extracted by the minimum distance decoder as

$$\hat{m} = \arg\min_m \left\| \mathbf{Y} - (q_\Delta(\mathbf{Y} + \mathbf{W}_m) - \mathbf{W}_m) \right\|. \tag{3.6}$$

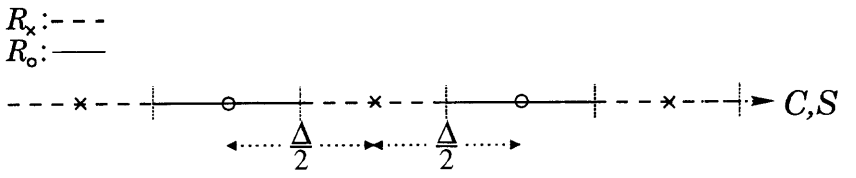


Figure 3.2 Reconstruction points of dithered quantizers corresponding to a binary watermark (dither) signal.

Figure 3.2 displays the reconstruction points of the dithered quantizers associated with the two watermark samples. The reconstruction points of the two quantizers are $\frac{\Delta}{2}$ apart. The decision regions denoted by $R_\times$ and $R_\circ$ determine the sustainable amount of noise for successful extraction of the message. The stego signal $\mathbf{S}$ is generated by quantizing each host signal coefficient C with the quantizer pointed by the binary watermark sample W of $\mathbf{W}$ to be embedded. (Accordingly, embedding of the watermark sample associated with the symbol $\times$ or $\circ$ refers to translation of the host signal coefficient C in the direction of nearest $\times$ or $\circ$, respectively.) Similarly, detection of a sent message is achieved by determining the nearest reconstruction points, denoted by $\times$ and $\circ$ symbols, to the coefficients of the received signal $\mathbf{Y}$.

The main disadvantage of type-II methods is that they perform well only if the attack is not severe (less than distortion P). In other words, its performance is equivalent to that of optimal design *only for the low attack case*, Section 2.1. For all other attack levels there's a performance gap with the upper bound, which increases with the attack level. This is due to the non-optimal codebook design based on $\alpha = 1$ or equivalently $\mathbf{X}_t = \mathbf{0}$, which undermines the dependency of codebook generation to the channel noise level. The poor performance of type-II methods with increasing attack levels is improved by the modifications proposed by the class of methods called as type-III.

3.2 Type-III Embedding and Detection Methods

The data hiding rate (payload) vs. robustness performance of type-II methods is substantially improved by enhancing the functionality of the embedder with further processing capabilities (*i.e.*, thresholding, distortion compensation, Gaussian mapping), see [20, 8, 21, 33]. In type-III methods, embedding quantization is followed by a processing stage (post-processing) that generates the stego signal. The improvement in the performance of type-III methods, compared to type-II, at the same noise level can be explained by the fact that codebook design depends on channel noise level or by the deviation from the non-optimal design of $\mathbf{X}_t = \mathbf{0}$ through the added processing. Alternately, in terms of Costa's framework, the improvement can be attributed to the effective value of α used in codebook generation which is less than one rather than being equal to one, as the latter is optimal for the no attack case. Data hiding methods with post-processing abilities enable the embedder to increase the distance between the reconstruction points of quantizers at a fixed embedding distortion. Therefore, they have improved detection capabilities for any finite WNR level (type-II is optimal only for the case of infinite WNR). On the other hand, since the detector is blind to the additional processing at the embedder, its structure is not altered.

The channel model for type-III hiding methods, based on the model for type-II methods given in Figure 3.1, is displayed in Figure 3.3. In the model, $\mathbf{X}$ is the type-II codeword (embedding distortion introduced due to the quantization), $\mathbf{X}_t$ is the processing distortion, and the channel output is $\mathbf{Y} = \mathbf{C} + \mathbf{X} - \mathbf{X}_t + \mathbf{Z}$. The processing distortion $\mathbf{X}_t$ is derived from $\mathbf{X}$ by the post-processing depending on the expected noise level. The type-III codeword that yields the stego signal, $\mathbf{S} = \mathbf{C} + \mathbf{X}_n$, is defined as $\mathbf{X}_n = \mathbf{X} - \mathbf{X}_t$. Correspondingly, embedder imposes the power constraint as $\frac{1}{N}||\mathbf{X}_n||^2 = P$.

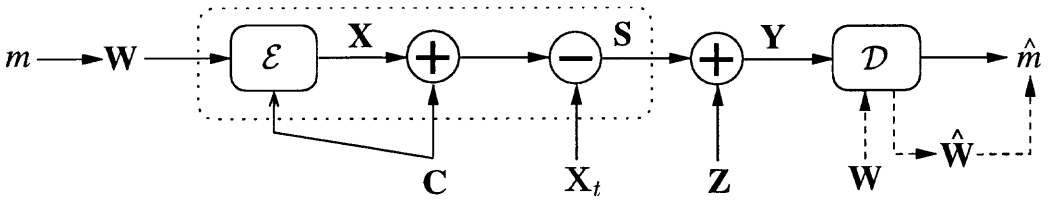


Figure 3.3 Block diagram of type-III embedding and detection stages.

In type-III methods, since the detector is not aware of the processing at the embedder, the processing distortion $\mathbf{X}_t$ can effectively be considered to be a part of the channel noise at the detector. Therefore, type-II codeword $\mathbf{X}$, which would yield an errorless extraction of the watermark signal $\mathbf{W}$, is distorted by two sources of noise, *viz.*, the attack $\mathbf{Z}$ and the processing distortion $\mathbf{X}_t$. (In other words, the signal $\mathbf{C}+\mathbf{X}$ refers to a signal quantized by the quantizer(s) associated with the watermark signal $\mathbf{W}$, and $\mathbf{W}$ can be perfectly recovered from this signal.) Therefore, the effective noise at the detector that distorts the embedded watermark signal is represented as $\mathbf{Z}_{eff} = \mathbf{Z} - \mathbf{X}_t$. In type-III methods, the invertibility condition on the $\mathcal{E}, D$ pair is sacrificed as a result of the processing that follows quantization of the host signal, $\mathcal{D}(\mathcal{E}(C, W)) \neq W$.

Performance of type-III hiding methods vary based on three factors: the type of post-processing that is incorporated with type-II embedding, the choice of demodulation function used in message extraction, and the criterion used for optimizing the embedding and detection parameters. Therefore, the performance of any type-III data hiding method can be evaluated further by considering these three issues.

3.2.1 Post-Processing Types

There are three types of post-processing that are employed in type-III embedder-detector designs. These are:

- Distortion compensation

- Thresholding
- Gaussian mapping

In [8], Chen, *et al.* identified the capacity achieving variant of QIM as distortion compensated QIM (DC-QIM). In DC-QIM, the quantization index modulated signal is perturbated by subtracting the $1-\alpha^*$ scaled version of the embedding distortion $\mathbf{X}$. Therefore, $\mathbf{X}_t = (1-\alpha^*)\mathbf{X}$, $\rho = 1$, and $\mathbf{X}_n = \alpha^*\mathbf{X}$. Ramkumar, *et al.* [20] proposed thresholding type of post-processing where the magnitude of distortions, that can be introduced to host signal samples, are limited to $\pm\frac{\beta}{2}$. Hence, the type-III codeword $\mathbf{X}_n$ is generated by limiting the values of $\mathbf{X}$, $\mathbf{X}_n = \min(|\mathbf{X}|, \frac{\beta}{2})\text{sign}(\mathbf{X})$. The processing distortion $\mathbf{X}_t$, in this case, is the thresholding noise, $\mathbf{X}_t = \max(0, |\mathbf{X}| - \frac{\beta}{2})\text{sign}(\mathbf{X})$. Perez-Gonzalez *et al.* [33], considering uniform scalar quantization, proposed to generate the processing distortion $\mathbf{X}_t$ from $\mathbf{X}$ by transforming each *iid* component X into a zero-mean Gaussian distributed random variable with a variance of σ_v^2, $\mathbf{X}_t = -\sigma_v \text{Q}^{-1}\left(\frac{\mathbf{X}+\frac{\Delta}{2}}{\Delta}\right)$ where $\text{Q}^{-1}(.)$ is the inverse Gaussian Q-function.

In type-III methods, the parameters α, β, and σ_v, depending on the type of post-processing, are selected such a way that the power constraint $\frac{1}{N}||\mathbf{X}_n||^2 = P$ is satisfied and the performance at the presumed noise (attack) level is maximized. Corresponding expressions for the processing distortion $\mathbf{X}_t$ and the codeword $\mathbf{X}_n$ for the three types of post-processing are as given in Table 3.1.

Table 3.1 Expressions for $\mathbf{X}_t$ and $\mathbf{X}_n$

Processing, $\mathcal{P}$	*Processing distortion*, $\mathbf{X}_t$	*Codeword*, $\mathbf{X}_n$
Thresholding	$\max(0, \lvert\mathbf{X}\rvert - \frac{\beta}{2})\text{sign}(\mathbf{X})$	$\min(\lvert\mathbf{X}\rvert, \frac{\beta}{2})\text{sign}(\mathbf{X})$
Distortion Compensation	$(1-\alpha)\mathbf{X}$	$\alpha\mathbf{X}$
Gaussian mapping	$-\sigma_v \text{Q}^{-1}\left(\frac{\mathbf{X}+\frac{\Delta}{2}}{\Delta}\right)$	$\mathbf{X} - \mathbf{X}_t$

Vectoral embedding and detection. The optimal processing, within the CAE-CID framework, requires that the processing distortion $\mathbf{X}_t$ be a linear function of the processing distortion $\mathbf{X}$. Accordingly, the power σ_X^2 of the embedding distortion $\mathbf{X}$ corresponding to distortion compensation type of processing can be computed in the limit, using $\frac{1}{N}\|\mathbf{X}_n\|^2 = P$, as

$$\sigma_X^2 = \frac{1}{N}\|\mathbf{X}\|^2 = \frac{1}{N}\left\|\frac{\mathbf{X}_n}{\alpha^*}\right\|^2 = \frac{(P+\sigma_Z^2)^2}{P} \tag{3.7}$$

where $\alpha^* = \frac{P}{P+\sigma_Z^2}$. It should be noted that, the variance of the *iid* components of the channel input $\mathbf{X}$ (the power of the input $\mathbf{X}$) in Equation (2.14) is the same as the power of the optimal embedding distortion $\mathbf{X}$ found in Equation (3.7), $\sigma_X = \sigma_X^*$. Therefore, distortion compensation is the optimal processing when the embedding distortion is Gaussian distributed. This can be satisfied by the use of high-dimensional quantization for embedding which yields Gaussian distributed quantization error. However, a capacity achieving embedding-detection scheme based on thresholding or Gaussian mapping types of post-processing is not possible since the relation between $\mathbf{X}$ and $\mathbf{X_t}$ is not linear.

Scalar embedding and detection. In the practical cases, where scalar quantization rather than high-dimensional vector quantization is employed at the embedder, $\mathbf{X}$ is an *iid* vector with a non-Gaussian distribution. Therefore, the optimal post-processing is not necessarily the distortion compensation. For the scalar quantization case, the embedding operation of all embedding-detection techniques can be represented by a form of dithered quantization. Thus, each component X of the embedding distortion $\mathbf{X}$, defined as $\mathbf{X} = q_\Delta(\mathbf{C}, \mathbf{W}_m) - \mathbf{W}_m - \mathbf{C}$, is uniformly distributed. However, the processing distortion $\mathbf{X}_t$ and its dependency on $\mathbf{X}$ are different for the three types of post-processing.

Eggers, *et al.*, in [21] optimized the value of α for scalar quantization, rather than assuming $\alpha^* = \frac{P}{P+\sigma_Z^2}$, and provided the approximation

$$\alpha = \sqrt{\frac{P}{P + 2.71\sigma_Z^2}}. \tag{3.8}$$

Expressions for the optimal values of Δ and the threshold β based on the expected attack level were reported in [20]. Although [33] does not provide the optimal σ_v values for Gaussian mapping, the optimization procedure is straightforward.

3.2.2 Forms of Demodulation

Detection of the sent message is achieved either by sample-wise hard decisions or soft decisions based on the availability of the set of watermark signals at the extractor side. The presence of watermark signals leads to an improved detection of the sent message since they can be utilized in detection operation [24, 20].

There are two forms of demodulation employed in detection of the sent message. In [24, 21, 33], demodulation of the sent message, from the received signal $\mathbf{Y}$, is realized by minimum distance decoding, and in [20], demodulation takes the form of maximum correlation rule.

Minimum distance detector. With the use of minimum distance detector, detection is simply the quantization of the received signal $\mathbf{Y}$ by all quantizers in the ensemble. The message letter or message index associated with the quantizer that yields the minimum Euclidean distance to received $\mathbf{Y}$ is deemed to be the sent message. The general form of minimum distance decoding based on dithered quantization can be rewritten, in terms of $\mathbf{Y}_m = \mathbf{Y} + \mathbf{W}_m$, as

$$\hat{m} = \mathcal{D}(\mathbf{Y}) = \arg\min_m ||\mathbf{Y}_m - Q_\Delta(\mathbf{Y}_m)||, \quad 1 \leq m \leq M. \tag{3.9}$$

It should be noted that Equation (3.9) is a minimization of the quantization error over all quantizers. For the case of scalar quantization, $Q_\Delta(.)$ takes the form of dithered quantizer $q_\Delta(.)$, Equation (3.6).

Figure 3.4 displays the detectors for the binary signaling case where the embedding operation is based on scalar quantization. In the figure, the symbols $\times$ and $\circ$ denote the reconstruction points of the quantizers associated with the watermark sample values of $-\frac{\Delta}{4}$ and $\frac{\Delta}{4}$. (However, it should be noted that, within the scope of DM, any two sample values with $\frac{\Delta}{2}$ difference are valid choices, see Equation (3.5).)

When the extractor has no access to the watermark signals but only knows the reconstruction points, each sample of the embedded watermark signal is detected from each coefficient Y of the received signal $\mathbf{Y}$ by individual hard decisions as

$$\hat{W}_i = \arg \min_{W_i \in \Omega} ||Y_i + W_i - q_\Delta(Y_i + W_i)|| \quad \text{for,} \quad i = 1, \ldots, N \tag{3.10}$$

where Ω is the set of signal representations for watermark samples. Equation (3.10) is based on determining the minimum Euclidean distance of the received signal coefficients to reconstruction points which can equivalently be achieved by mapping each coefficient Y over the square wave function displayed in Figure 3.4-a. Then, the extracted binary watermark samples, $\hat{W}_1, \ldots, \hat{W}_N$, are combined into the sequence $\hat{\mathbf{W}}$ to generate the embedded watermark signal.

On the other hand, when the watermark signals are present at the detector, detection of each sample is by soft decisions. Accordingly, each coefficient Y_m of the signal $\mathbf{Y}_m$, that is obtained from the received signal $\mathbf{Y}$, is mapped over the sawtooth function displayed in Figure 3.4-b. The norm of the resulting signal values is the distance between $\mathbf{Y}$ and $\mathbf{W}_m$. Hence, the watermark signal that has the minimum distance to $\mathbf{Y}$ is regarded as the embedded signal.

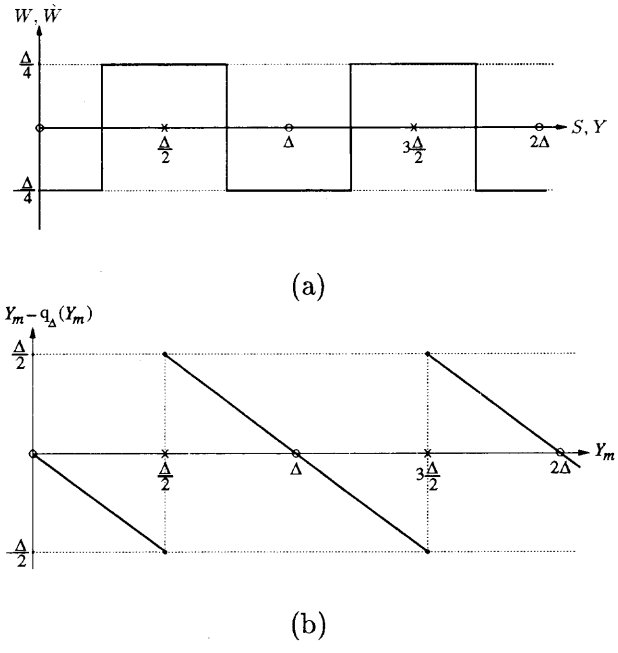


Figure 3.4 Demodulation for DM based on (a) hard decisions and (b) soft decisions.

Maximum correlation detector. When the demodulation scheme is based on maximum correlation detector, watermark signals are assumed to be present at the detector. In this form of demodulation, at first, an estimate $\hat{\mathbf{W}}$ of the embedded watermark signal is extracted from the received signal by soft decisions. Then, the sent message is detected by matching the estimate of the embedded watermark signal to one of the watermark signals using a correlation based similarity measure as

$$\begin{aligned} \hat{\mathbf{W}} &= \mathcal{D}(\mathbf{Y}), \\ \hat{m} &= \arg\max_{m} \frac{\mathbf{W}_m \hat{\mathbf{W}}}{||\mathbf{W}_m||\ ||\hat{\mathbf{W}}||}, \ 1 \leq m \leq M. \end{aligned} \tag{3.11}$$

Since the hard decisions are caused by the discontinuities in the extraction function, Figure 3.4-a, Reference [20] proposed a continuous periodic triangular extraction function. Figure 3.5 displays the corresponding function used for extracting embedded binary watermark samples that are confined to values $-\frac{\Delta}{4}$ and $\frac{\Delta}{4}$ for maximum separation, $\Omega = \{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$. An estimate of the embedded watermark signal

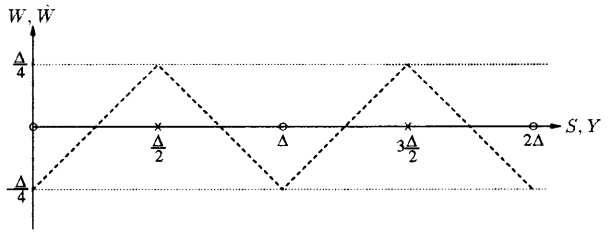


Figure 3.5 Periodic extraction function corresponding to soft decisions.

is obtained by mapping each coefficient of $\mathbf{Y}$ over the periodic triangular function, rather than making a hard decision by the Euclidean distance decoder. As a result, each extracted sample $\hat{W}$ is a real valued signal in the range of $[-\frac{\Delta}{4}, \frac{\Delta}{4}]$. Message detection is achieved by combining the sample estimates into $\hat{\mathbf{W}} = (\hat{W}_1, \ldots, \hat{W}_N)$ and then matching $\hat{\mathbf{W}}$ to one of $\mathbf{W}_1, \ldots, \mathbf{W}_M$.

3.2.3 Optimization Criteria for Embedding and Detection Parameters

The embedding and detection operations are controlled by a pair of parameters. The values for these parameters are optimized for the given channel noise and permitted distortion levels, σ_Z^2 and P.

One of the parameters which is common to all techniques is Δ which designates the distance between the reconstruction points of the embedding quantizers. The choice of Δ determines the embedding distortion due to quantization, and it is known to both embedder and detector. The other parameter controls the amount of processing distortion introduced to quantized signal (type-II embedded signal) by the post-processing and, it limits the distortion due to embedding operation to the permitted amount. This parameter is known only to embedder and parameterized as β, α, or σ_V depending on the type of post-processing. The values for the two interdependent parameters can be optimized based on various performance criteria as discussed in the following sections.

Optimization of parameters for vectoral embedding and detection. In [8], researchers optimized the embedding-detection parameters by maximizing the ratio of the embedding distortion to the sum of processing and channel distortions, $(\frac{\sigma_X^2}{\sigma_{X_t}^2+\sigma_Z^2})$, at the extractor as

$$(\Delta, \sigma_{X_t}^2) = \arg\max_{\Delta,\sigma_{X_t}^2} \left\{ \frac{\sigma_X^2}{\sigma_{X_t}^2+\sigma_Z^2} \;\middle|\; \sigma_Z^2, \frac{1}{N}\|\mathbf{X}_n\|^2 = P, \mathbf{X_t} \right\}. \tag{3.12}$$

With the use of high dimensional quantization for embedding and detection, the marginal pdf of embedding distortion $\mathbf{X}$ approximates Gaussian distribution and consequently distortion compensation becomes the optimal processing. Hence, for the given channel noise level, Δ and α are selected such a way that Equation (3.12) is satisfied where $\mathbf{X}_t = (1-\alpha)\mathbf{X}$ and $\mathbf{X}_n = \alpha\mathbf{X}$, *i.e.* $\sigma_{X_t}^2 = (1-\alpha)^2\sigma_X^2$ and $\sigma_X^2 = \frac{P}{\alpha^2}$. This leads to $\alpha = \frac{P}{P+\sigma_Z^2}$ which is in accord with the results of Section 2.1 due to the duality between the two channel models.

Optimization of parameters for scalar embedding and detection. Researchers in [20, 21, 33] modeled the effective noise that distorts the embedded watermark signal in terms of the statistics of the channel noise $\mathbf{Z}$ and the processing distortion $\mathbf{X}_t$, $\mathbf{Z}_{eff} = \mathbf{Z} - \mathbf{X}_t$. The optimum values for embedding-detection parameters are then selected such a way that the distortion in the extracted watermark signal is minimized.

When the host signal is uniformly distributed over all quantization intervals, the embedding distortion X introduced to each host signal coefficient C is uniformly distributed in $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$. For thresholding type of post-processing the parameters are the step size Δ and the threshold β. The corresponding pdf and statistics of processing distortion X_t and the codeword X_n are expressed as

$$f_{X_t}(x_t) = \frac{\beta}{\Delta}\delta(x_t) + \frac{1}{\Delta}\text{rect}(\Delta-\beta), \tag{3.13}$$

$$m_{X_t} = 0, \tag{3.14}$$

$$\sigma^2_{X_t} = \frac{(\Delta-\beta)^3}{12\Delta}, \tag{3.15}$$

$$f_{X_n}(x) = \frac{1}{\Delta}\text{rect}(\beta) + \frac{\Delta-\beta}{2\Delta}\left(\delta(x_n-\frac{\beta}{2})+\delta(x_n+\frac{\beta}{2})\right), \tag{3.16}$$

$$m_{X_n} = 0, \tag{3.17}$$

$$\sigma^2_{X_n} = \frac{\beta^2}{12\Delta}(3\Delta-2\beta). \tag{3.18}$$

where $\text{rect}(x)$ is the rectangular function in x with a value of one in the interval $(-\frac{1}{2},\frac{1}{2})$ and zero elsewhere. Similarly for distortion compensation type of post-processing, corresponding pdfs and statistics are found in terms of Δ and α as

$$f_{X_t}(x_t) = \frac{1}{(1-\alpha)\Delta}\text{rect}((1-\alpha)\Delta), \tag{3.19}$$

$$m_{X_t} = 0, \tag{3.20}$$

$$\sigma^2_{X_t} = \frac{(1-\alpha)^2\Delta^2}{12}, \tag{3.21}$$

$$f_{X_n}(x) = \frac{1}{\alpha\Delta}\text{rect}(\alpha\Delta), \tag{3.22}$$

$$m_{X_n} = 0, \tag{3.23}$$

$$\sigma^2_{X_n} = \frac{\alpha^2\Delta^2}{12}. \tag{3.24}$$

When the post-processing takes the form of Gaussian mapping, X_t is a zero mean Gaussian random variable with variance σ_v^2 and the parameters are Δ and σ_v. However, as the dependency between X and X_t is through a Gaussian transformation, the pdf of X_n is not a straightforward one but its statistics can be calculated as

$$E[X_n^k] = \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \left(x+\sigma_v \text{Q}^{-1}\left(\frac{x+\frac{\Delta}{2}}{\Delta}\right)\right)^k \frac{1}{\Delta}dx. \tag{3.25}$$

Figures 3.6 and 3.7 display $f_X(x)$, $f_{X_t}(x_t)$ and $f_{X_n}(x_n)$ for thresholding and distortion compensation types of post-processing, respectively.

Given the host signal is *iid*, $\mathbf{X}$ and $\mathbf{X}_t$ are *iid* random vectors with the marginal distributions given above, since embedding operation is memoryless. It should also

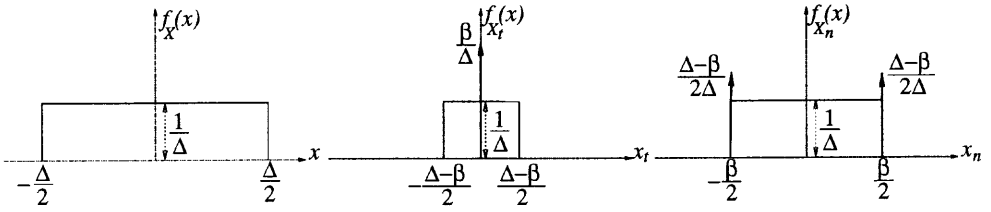


Figure 3.6 Probability density functions *(left)* $f_X(x)$, *(center)* $f_{X_t}(x_t)$, *(right)* $f_{X_n}(x_n)$ corresponding to thresholding type of processing for $0 < \beta < \Delta$.

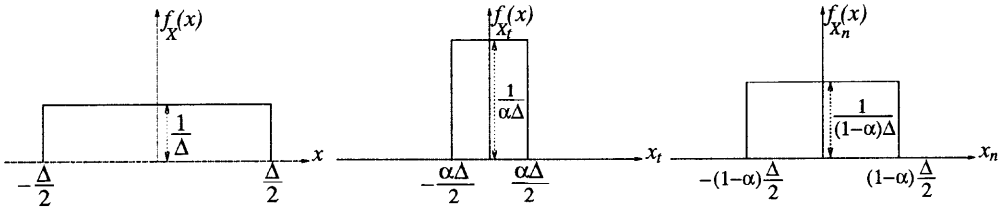


Figure 3.7 Probability density functions *(left)* $f_X(x)$, *(center)* $f_{X_t}(x_t)$, *(right)* $f_{X_n}(x_n)$ corresponding to distortion compensation type of processing for $\alpha < 1$.

be noted that, for large N, the distortion P introduced to host signal $\mathbf{C}$, due to embedding operation, is equal to $\sigma^2_{X_n}$, *i.e.* $\frac{1}{N}||\mathbf{X}_n^2|| = P$.

Assuming Z and X_t are independent, the resulting pdf of Z_{eff}, $f_{Z_{eff}}(z_{eff})$ can be computed by the convolution of the individual pdfs $f_Z(z)$ and $f_{X_t}(x_t)$ as

$$f_{Z_{eff}}(z_{eff}) = \int_{-\infty}^{\infty} f_Z(z_{eff} - x) f_{X_t}(x) dx. \tag{3.26}$$

Thus, for $Z \sim \mathcal{N}(0, \sigma_Z^2)$, $f_{Z_{eff}}(z_{eff})$ corresponding to thresholding type of processing is derived as

$$\begin{aligned} f_{Z_{eff}}(z_{eff}) &= \frac{\beta}{\Delta\sqrt{2\pi\sigma_Z^2}} \exp^{-\frac{z_{eff}^2}{2\sigma_Z^2}} \\ &+ \frac{1}{2\Delta}\left(\text{erf}\left(\frac{z_{eff} + \frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_Z}\right) - \text{erf}\left(\frac{z_{eff} - \frac{\Delta-\beta}{2}}{\sqrt{2}\sigma_Z}\right)\right), \end{aligned} \tag{3.27}$$

where erf(.) is the Gaussian error function, $\text{erf}(z) = \frac{2}{\pi}\int_0^z \exp^{-x^2} dx$. Similarly, for distortion compensation and Gaussian mapping cases $f_{Z_{eff}}(z_{eff})$ is expressed as

$$f_{Z_{eff}}(z_{eff}) = \frac{1}{2(1-\alpha)\Delta}\left(\text{erf}\left(\frac{z_{eff} + \frac{(1-\alpha)\Delta}{2}}{\sqrt{2}\sigma_Z}\right) - \text{erf}\left(\frac{z_{eff} - \frac{(1-\alpha)\Delta}{2}}{\sqrt{2}\sigma_Z}\right)\right) \tag{3.28}$$

and

$$f_{Z_{eff}}(z_{eff}) = \frac{1}{\sqrt{2\pi(\sigma_Z^2 + \sigma_v^2)}} \exp^{-\frac{z_{eff}^2}{2(\sigma_Z^2+\sigma_v^2)}}, \tag{3.29}$$

respectively.

The embedding-detection parameters are optimized by proper selection of the step size Δ and the amount of processing distortion $\sigma_{X_t}^2$. Such a selection can be based on one of the three criterion for the given statistics of $\mathbf{Z}_{eff}$.

Maximizing correlation. With this criterion, the selection of parameters is based on maximizing the normalized correlation between the embedded and extracted watermark signals [20]. Since $\mathbf{Z}_{eff}$ is the noise that distorts the type-II codeword $\mathbf{X}$ corresponding to watermark signal $\mathbf{W}$, the signal $\hat{\mathbf{W}}$ extracted from $\mathbf{Y}$ can be expressed in terms of $\mathbf{Z}_{eff}$ and $\mathbf{W}$ using the extraction function shown in Figure 3.5. (Note that if $\mathbf{Z}_{eff} = 0$, then $\mathbf{W}$=$\hat{\mathbf{W}}$.) Hence, a binary distributed watermark signal sample W with values in $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$ embedded in a host signal coefficient is extracted as

$$\hat{W} = \begin{cases} (\frac{(2i+1)\Delta}{4} - Z_{eff})(-1)^i, \; i\frac{\Delta}{2} < Z_{eff} \leq \frac{(i+1)\Delta}{2}, i \in \mathcal{Z} \text{ if } W = \frac{\Delta}{4}, \\ (-\frac{(2i+1)\Delta}{4} + Z_{eff})(-1)^i, \; i\frac{\Delta}{2} < Z_{eff} \leq \frac{(i+1)\Delta}{2}, i \in \mathcal{Z} \text{ if } W = -\frac{\Delta}{4}. \end{cases} \tag{3.30}$$

Due to memoryless embedding-detection and attack schemes, the vectors $\mathbf{W}$ and $\hat{\mathbf{W}}$ are *iid* with sample values W and $\hat{W}$. Hence the normalized correlation ρ between $\mathbf{W}$ and $\hat{\mathbf{W}}$ can be analytically computed for large N as

$$\begin{aligned} \rho &= E\left[\frac{\mathbf{W}^T\hat{\mathbf{W}}}{||\mathbf{W}||\,||\hat{\mathbf{W}}||}\right], \\ &= \frac{E[W\hat{W}]}{\sqrt{E[W^2]E[\hat{W}^2]}}, \\ &= \frac{R(1)}{\sqrt{R(2)}}, \end{aligned} \tag{3.31}$$

where $E[W\hat{W}]$ is the first joint moment of the random variables W and $\hat{W}$ and

$$R(p) = 2\sum_{i=0}^{i=\infty} \int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}} \left(\left(\frac{(2i+1)\Delta}{4} - z_{eff} \right) (-1)^i \right)^p f_{Z_{eff}}(z_{eff}) dz_{eff}. \tag{3.32}$$

Therefore, the optimal parameter values for the utilized post-processing technique is computed by maximizing Equation (3.31) over Δ and $\sigma^2_{X_t}$ using the pdfs given in Equations (3.27-3.29) for the given channel noise level and permitted distortion as

$$(\Delta, \sigma^2_{X_t}) = \arg \max_{\Delta, \sigma^2_{X_t}} \left\{ \rho \;\middle|\; \sigma^2_Z, \mathbf{X}_t \in \mathcal{X}_t, \sigma^2_{X_n} = P \right\} \tag{3.33}$$

where

$$\mathcal{X}_t = \left\{ \max\left(\mathbf{0}, |\mathbf{X}| - \frac{\beta}{2}\right) \text{sign}(\mathbf{X}), (1-\alpha)\mathbf{X}, -\sigma_v \mathrm{Q}^{-1}\left(\frac{\mathbf{X} + \frac{\Delta}{2}}{\Delta} \right) \right\}, \tag{3.34}$$

and $\mathbf{X} = q_\Delta(\mathbf{C} + \mathbf{W}) - \mathbf{W} - \mathbf{C}$.

Minimizing probability of error. The embedding-detection parameters are selected to minimize the probability of error in detecting an embedded watermark sample [33]. Since Z_{eff} indicates the deviation of the received signal coefficient Y from the reconstruction points, the probability of detection error, P_e, can be calculated by integrating $f_{Z_{eff}}(z_{eff})$ over all decision regions but excluding the one associated with the sent sample as

$$\mathrm{P}_e = \mathrm{P}\{Y \notin \mathcal{R}_W \mid W\} \tag{3.35}$$

where $\mathcal{R}_W$ denotes the decision region associated with the sample W. For the binary signaling case depicted in Figure 3.8, the symbols $\times$ and $\circ$ denote the reconstruction points of two quantizers associated with sample values $\frac{\Delta}{4}$ and $-\frac{\Delta}{4}$, respectively. The decision regions $R_\times$ and $R_\circ$ are used to map the received signal coefficient Y to $\frac{\Delta}{4}$ or $-\frac{\Delta}{4}$ by hard decisions. Assuming $\frac{\Delta}{4}$ and $-\frac{\Delta}{4}$ are equally likely to be embedded,

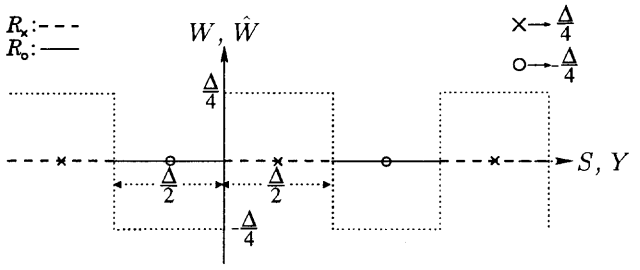


Figure 3.8 Embedding and detection of a binary watermark sample.

corresponding P_e is calculated as

$$\begin{aligned} P_e &= P\left\{\left\|Y + \frac{\Delta}{4} - q_\Delta\left(Y + \frac{\Delta}{4}\right)\right\| > \left\|Y - \frac{\Delta}{4} - q_\Delta\left(Y - \frac{\Delta}{4}\right)\right\| \middle| w = \frac{\Delta}{4}\right\}, \\ &= P\{Y \in \mathcal{R}_o \mid w = \frac{\Delta}{4}\}, \\ &= \int_{\mathcal{R}_o} f_{Z_{eff}}(z_{eff} - \frac{\Delta}{4}) dz_{eff} \end{aligned} \tag{3.36}$$

Then, the parameters can be selected to minimize P_e for the given P, σ_Z^2, and the type of post-processing as

$$(\Delta, \sigma_{X_t}^2) = \arg\min_{\Delta, \sigma_{X_t}^2} \left\{ P_e \;\middle|\; \sigma_Z^2, X_t \in \mathcal{X}_t, \sigma_{X_n}^2 = P \right\} \tag{3.37}$$

where $\mathcal{X}_t$ is given in Equation (3.34).

Maximizing mutual information. The parameters are selected to maximize the mutual information between the embedded watermark sample W and the received signal coefficient Y [21]. The mutual information between W and Y is expressed as

$$I(W,Y) = H(Y) - H(Y|W) \tag{3.38}$$

where $H(.)$ is the differential entropy of a random variable in bits that is defined as $H(X) = -\int_{-\infty}^{\infty} f_X(x) \log_2[f_X(x)] dx$. As the erroneous detection of W from Y is due to the noise Z_{eff}, $H(Y|W)$ in Equation (3.38) can be computed in terms of the effective noise pdf conditioned on W, $f_{z_{eff}|w}(z_{eff}|w)$. The pdf $f_{Z_{eff}|W}(z_{eff}|w)$

can be calculated over any quantization interval Δ, since the signal constellation is periodic with Δ (reconstruction points corresponding to quantizer associated with W are Δ apart). However, one should take into account that when Z_{eff} is heavy tailed (the range of $f_{Z_{eff}}(z_{eff})$ is larger than Δ), its pdf will be wrapped around Δ due to the periodicity. Consequently, $H(Y)$ is computed from $H(Y|W)$ by averaging it over W. (Assuming all samples $W \in \Omega$ are equally likely, $H(Y)$ is obtained as $\frac{1}{|\Omega|}\sum_{W\in\Omega} H(Y|W)$.) With this criterion, optimization of parameter values is by maximizing Equation (3.38) for the given constraints over Δ and $\sigma^2_{X_t}$ as

$$(\Delta, \sigma^2_{X_t}) = \arg\max_{\Delta,\sigma^2_{X_t}} \left\{ I(Y,W) \;\middle|\; \sigma^2_Z, X_t \in \mathcal{X}_t, \sigma^2_{X_n} = P \right\}. \tag{3.39}$$

The use of Equation (3.38) also enables computation of the maximum hiding rate in bits per host signal coefficient achievable with a particular embedding-detection technique. Therefore, it is a useful performance evaluation tool.

3.3 Performance Comparisons

Figure 3.9 displays the achievable data hiding rates of various embedding-detection techniques for the binary signaling case, obtained using Equations (2.3) and (3.38), compared to hiding rates of type-I (additive scheme) and optimal type-III (capacity). The embedding-detection parameters for type-II and type-III methods are selected so that the hiding rate is maximized, Equation (3.39). The additive scheme (type-I) and DM (type-II) have preferable performances, respectively, at very low and very high WNRs. For DM, the gap with the upper bound at higher WNRs is due to binary signaling. Thus the performance can be improved for multi-level signal representations. The poor performance of both methods in mid-WNR range is due to non-optimal codebook designs, as discussed in Section 2.4. In the former, the codebook design does not utilize the host signal, and in the latter, the design disregards the channel noise level.

The type-III versions of DM, implemented by incorporating the embedding of DM with thresholding, distortion compensation, and Gaussian mapping types of post-processing, have better performances than DM due to the deviation from the optimistic "low-noise" assumption in the codebook design. These methods have significantly improved performances in the mid-WNR range, however, in order to achieve higher rates, embedding through scalar quantization has to be substituted by high-dimensional vector quantization.

Type-III methods employing thresholding and distortion compensation types of post-processing perform closely in the whole WNR range. On the other hand, Gaussian mapping processing has a comparable performance only for WNRs higher than -7.8 dB. Below that range the rate drops rapidly. At WNRs lower than -8.7 dB thresholding performs marginally better, while from -8.7 dB to -7 dB, distortion compensation performs best. Above -7 dB, both distortion compensation and Gaussian mapping are the preferred post-processing types. Figures 3.10-3.13 show the hiding rates for the corresponding methods with multi-level signaling. With the decreasing noise level and higher signal representation levels, all methods yield similar data hiding rates as the need for post-processing reduces. Ultimately when there's no noise, the DM is the optimal embedding-detection technique.

The normalized correlation, ρ, and probability of error, P_e, performances for the considered methods are respectively given in Figures 3.14 and 3.15. The corresponding embedding-detection parameters for the hiding methods are selected as described in Section 3.2, Equations (3.33) and (3.37). The correlation between an embedded binary watermark signal $\mathbf{W}$ and extracted watermark signal $\hat{\mathbf{W}}$ is calculated by using Equation (3.31), and the probability of the error in detecting an embedded binary watermark sample is computed by using Equation (3.36).

The relative performances of the three types of post-processing obtained for the three criteria, Figures 3.9, 3.14 and 3.15, are in accord with each other.

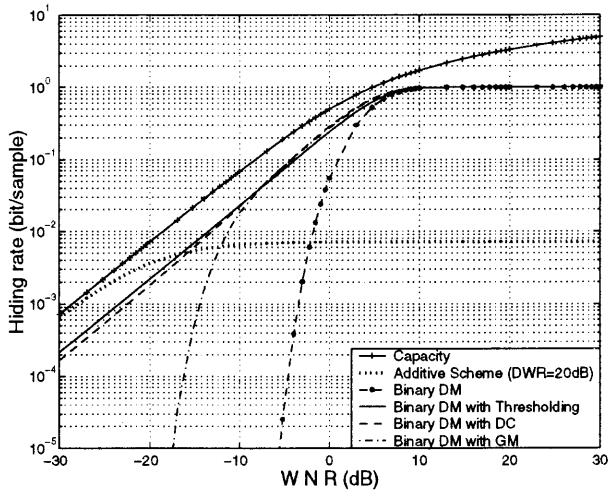


Figure 3.9 Comparison of the hiding rates corresponding to various hiding methods considering binary signaling obtained for $P = 10$.

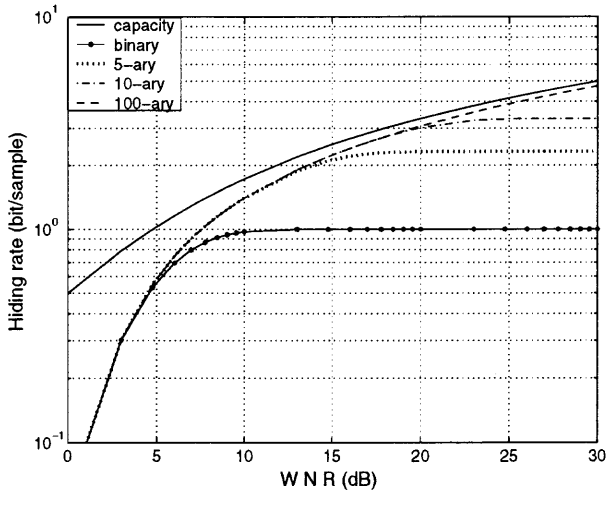


Figure 3.10 Data hiding rates for DM with binary, 5-ary, 10-ary, and 100-ary signaling.

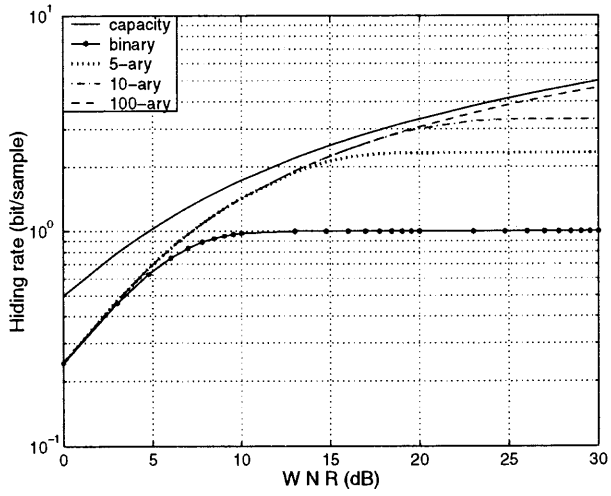


Figure 3.11 Data hiding rates for DM followed by thresholding type of post-processing with binary, 5-ary, 10-ary, and 100-ary signaling.

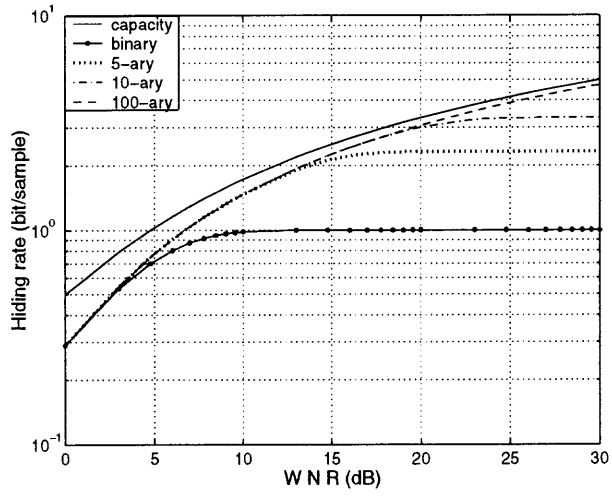


Figure 3.12 Data hiding rates for DM followed by distortion compensation type of post-processing with binary, 5-ary, 10-ary, and 100-ary signaling.

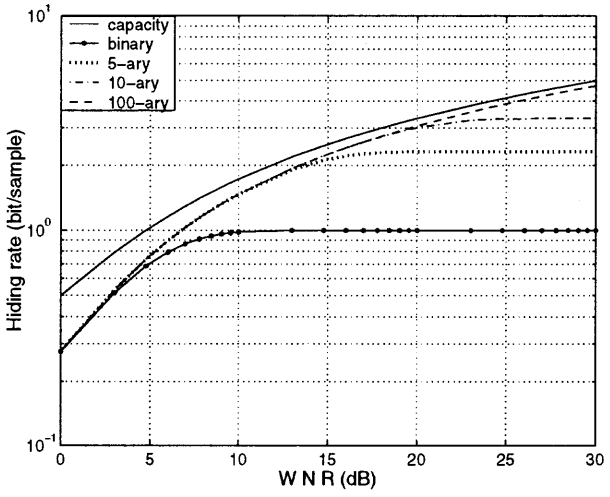


Figure 3.13 Data hiding rates for DM followed by Gaussian mapping type of post-processing with binary, 5-ary, 10-ary, and 100-ary signaling.

Thus, thresholding type of post-processing performs better when WNR is below approximately -9 dB, and at higher WNRs distortion compensation has better performance. Above -7 dB, Gaussian mapping and distortion compensation have comparable performances, and DM performs well only at higher WNR range, as expected. Figures 3.16 and 3.17 display the actual simulation results obtained by embedding and detecting binary watermark signals. In Figure 3.16, the normalized correlation ρ between the embedded vector $\mathbf{W}$ and its extracted version $\hat{\mathbf{W}}$ is measured, and in Figure 3.17, the error probability in detecting an embedded watermark sample W is measured. Both simulation results are in accord with theoretical values computed in Figures 3.14 and 3.15.

One intuitive way to evaluate the performance characteristics of type-I, type-II, and type-III methods at varying noise levels is by considering the size of decision cells at the detector, as discussed in Section 2.4. For type-II methods in the absence of noise, the extracted watermark signals correspond to reconstruction points of the embedding quantizers. Thus, decision cells can collapse to points and the data hider can afford to use higher level signaling without any performance penalty. However, with the increasing noise level, the successful extraction of the embedded watermark

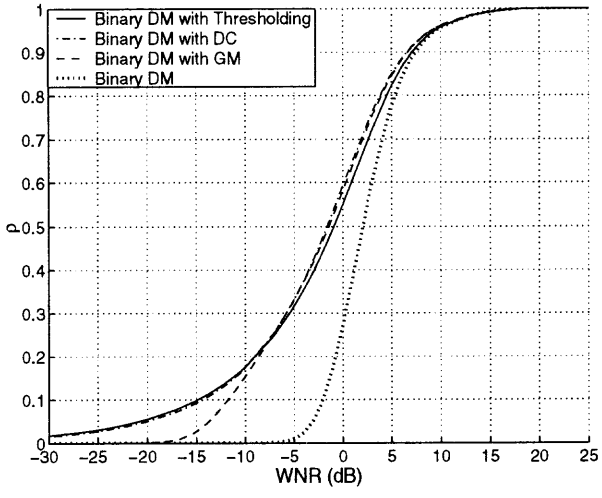


Figure 3.14 The normalized correlation between $\mathbf{W}$ and $\hat{\mathbf{W}}$ for the considered hiding methods when $P = 10$.

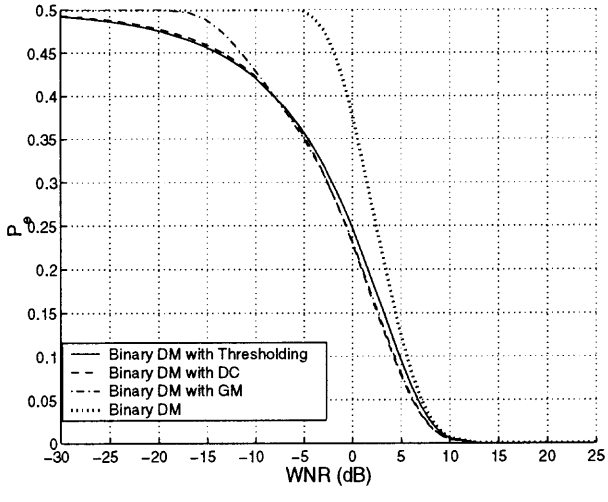


Figure 3.15 The probability of error in detecting W for the considered hiding methods when $P = 10$.

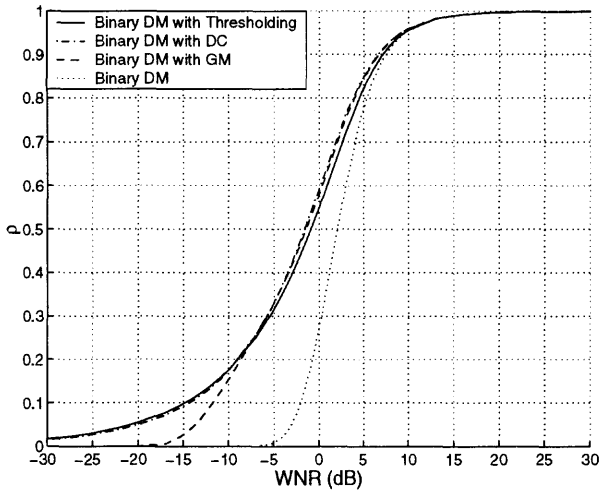


Figure 3.16 The actual measured normalized correlation between embedded $\mathbf{W}$ and extracted $\hat{\mathbf{W}}$ for the considered hiding methods when $P = 10$.

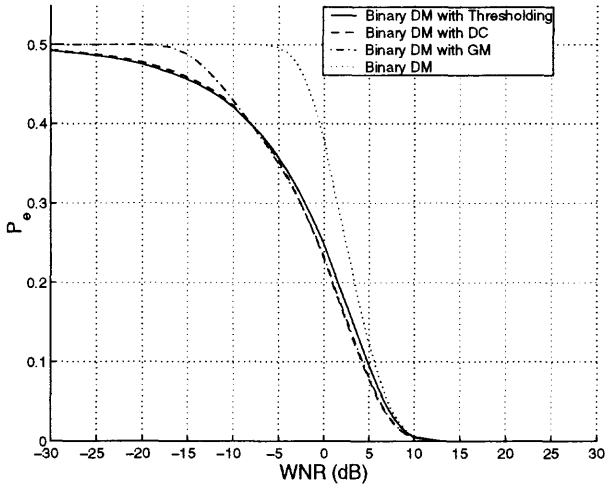


Figure 3.17 The actual measured error probability in detecting W for the considered hiding methods when $P = 10$.

signal requires decision cells to be enlarged accordingly. In type-III methods Δ is increased in accordance with the channel noise level σ_Z^2 and the corresponding increase in embedding distortion due to increased Δ is compensated by the post-processing. Hence, the data hider has the freedom to change the size of the decision cell depending on the noise level. Ultimately when the noise level is very high, the optimal strategy becomes making the decision regions arbitrarily large as in type-I methods where even for very high noise levels the detector is able to extract some of the embedded watermark signals.

3.4 Perceptual Constraints

As the resource of the communication between the hider and the attacker is the total imperceptible distortion that can be introduced to a given host signal, achieving the optimal rate vs. robustness performance requires a higher level understanding of the host signal in the perceptual sense. Data hiding methods, most generally, approach the problem by incorporating simplified perceptual models or the findings of perceptual compression with the embedding process.

Most elaborate formulations of the data hiding (as discussed in this chapter) rely on a fixed distortion measure, *e.g.* mean squared error distortion, for analytical tractability. Hence, the corresponding analyses and results oversimplify this aspect of the problem. Evaluated from imperceptibility perspective, type-I methods can exploit the host signal information better than type-II and type-III methods.

Within the additive schemes, embedding is by adding a scaled version of the watermark signal to the host signal or to a transformed form of it. The proper weighting for each watermark signal sample can be locally determined according to just noticeable difference (JND) thresholds and masking principles, thereby complying with perceptual constraints. In quantization based techniques, however, the distortion introduced to each host signal coefficient can only be controlled in an indirect manner

by adjusting either the quantization step size or the amount of processing distortion. Since optimization procedure for the embedding and detection parameters assume power limited distortion, which disregards the perceptual properties of the host signal, the corresponding embedding operation is a non-optimal one in terms of perceptual criteria. In this respect, scalar quantization based embedding-detection schemes provide a better control, since each coefficient is embedded individually and Δ or the post-processing parameter can be selected to comply with perceptual constraints. Whereas in schemes that employ high dimensional quantization, the introduced distortion due to embedding is minimized over the quantized vector which would not necessarily limit the distortion introduced to each coefficient.

In order to achieve imperceptibility, type-II and type-III methods select the power constraint conservatively. This leads to an under-utilization of the communication resource. Compared to type-II methods, the post-processing involved in type-III methods give hider another degree of freedom in controlling the distortion introduced to each host signal sample. Hence, the embedding parameter that designates the amount of processing distortion introduced to the quantized host signal (*i.e.* β in thresholding, α in distortion compensation, σ_V in Gaussian mapping) can be fine-tuned in accordance with the perceptual features of the host signal. Thresholding and distortion compensation types of post-processing can be readily adapted to applications with more strict imperceptibility requirements through adjusting β and α. Whereas with Gaussian mapping, modulating the processing distortion is a more complex task due to the non-linear transformation. However, the optimal approach is to revise the optimization procedures given in Section 3.2 (Equations (3.33), (3.37), and (3.39)) by taking into account the perceptual properties of the host signal as constraints (rather than limiting the distortion power to P) during the optimization of embedding-detection parameter values.

CHAPTER 4

PERFORMANCE AND COMPLEXITY TRADEOFFS

In order to further improve the performance of embedding-detection techniques, performance and complexity tradeoffs are to be made depending on the host signal size N. The two extreme cases are when the embedding signal size N is very large and very small.

For the case where the host signal size is large, *spread transforming* can be employed. Inspired by the spread-spectrum communications, the authors in Reference [18] used spread transforming in order to increase the WNR at the extractor by sacrificing in signal size N. However, they did not consider the problem of choosing the optimal "spreading factor." The concept of optimal spreading factor was addressed by Ramkumar *et al.* in [20] and by Eggers *et al.* in [49], independently.

On the contrary, when the signal size is small, *multiple codebook hiding method* can be used. The authors in References [50, 51, 52] introduced multiple codebook hiding to enable the embedding of watermark signal at lower embedding distortion levels. The use of multiple codebooks provides embedder with the choice of the codeword that better adapts to the host signal at the expense of increased complexity.

4.1 Spread Transforming

The underlying idea of spread transforming is to embed the watermark signal into a projection of the host signal and generate the stego signal by spreading the corresponding lower dimensional embedding distortion over the high dimensional host signal. In spread transforming, a pseudo-random vector $\mathbf{u}$ of size L with unit norm, $1 = \mathbf{u}^T\mathbf{u}$, is designated as the spreading vector and made known to both

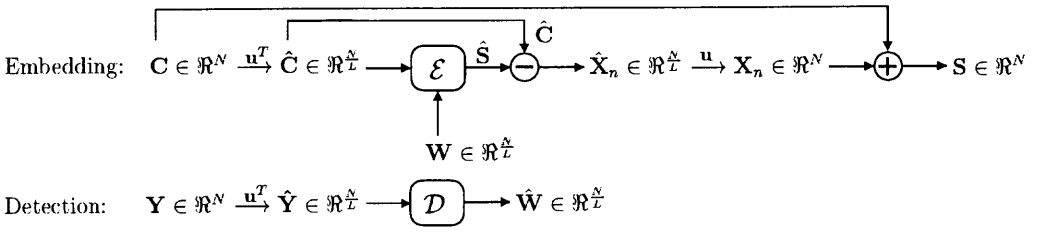


Figure 4.1 Embedding and detection with spread transforming.

embedder and detector. The embedding and detection operations are performed as follows.

At the embedder, the host signal vector $\mathbf{C} \in \Re^N$, is split into blocks of length L such that $\mathbf{C}^T = [\mathbf{C}_1^T, \ldots, \mathbf{C}_{\frac{N}{L}}^T]$ where $\mathbf{C}_i \in \Re^L$. Each block of data is projected onto $\mathbf{u}$ to generate the projected host signal $\hat{\mathbf{C}}^T = [\hat{C}_1, \ldots, \hat{C}_{\frac{N}{L}}]$ where $\hat{\mathbf{C}} \in \Re^{\frac{N}{L}}$ and

$$\hat{C}_i = \mathbf{C}_i^T \mathbf{u}, \quad i = 1, \ldots, \frac{N}{L}. \tag{4.1}$$

Then, the watermark signal $\mathbf{W} \in \Re^{\frac{N}{L}}$, corresponding to a message index, is embedded into $\hat{\mathbf{C}}$, $\hat{\mathbf{S}} = \mathcal{E}(\hat{\mathbf{C}}, \mathbf{W})$. The stego signal $\mathbf{S}^T = [\mathbf{S}_1^T, \ldots, \mathbf{S}_{\frac{N}{L}}^T]$ is generated from $\hat{\mathbf{S}}^T = [\hat{S}_1, \ldots, \hat{S}_{\frac{N}{L}}]$ as

$$\mathbf{S}_i = \mathbf{C}_i + (\hat{S}_i - \hat{C}_i)\mathbf{u}, \quad i = 1, \ldots, \frac{N}{L}. \tag{4.2}$$

Similarly, at the detector, the received signal is partitioned into blocks, $\mathbf{Y}^T = [\mathbf{Y}_1^T, \ldots, \mathbf{Y}_{\frac{N}{L}}^T]$, and each block of data is projected onto $\mathbf{u}$, $\hat{\mathbf{Y}} = [\hat{Y}_1, \ldots, \hat{Y}_{\frac{N}{L}}]$ where $\hat{Y}_i = \mathbf{Y}_i^T \mathbf{u}$. This is followed by the detection of the hidden signal, $\mathcal{D}(\hat{\mathbf{Y}})$. Figure 4.1 depicts the embedding and detection operations with spread transforming.

With spreading, the bandwidth is reduced by a factor of L, from N to $\frac{N}{L}$, as $\frac{N}{L}$ coefficients are information embedded. However, the embedding distortion is spread over all of the N coefficients. On the other hand, the distortion introduced to the host signal is $\frac{N}{L}P = ||\hat{\mathbf{S}} - \hat{\mathbf{C}}||^2$ which would be NP without the spreading. Therefore, the hider can afford to increase the embedding power by a factor of L. At the embedder, this reflects as an increase in the distance between the reconstruction points of the

embedding quantizers (when scalar quantization is considered, spreading with a factor of L leads to an increase in Δ by a factor of $\sqrt{L}$, *i.e.* $LP = \frac{(\sqrt{L}\Delta)^2}{12}$ where $P = \frac{\Delta^2}{12}$ is the embedding distortion per coefficient). Therefore, the system operates at a higher WNR level. An alternate interpretation of the gain due to spreading is that the stego signal can only be distorted by the component of the noise that is in the direction of the vector $\mathbf{u}$, which improves the robustness against noise.

Spread transforming method can be generalized to include non-integer spreading factors by adopting a transform domain embedding-detection approach where each basis vector of the transform basis is treated as a spreading vector. Let $\mathbb{U} \in \Re^{L \times L}$ be a unitary transformation matrix, $\mathbb{I} = \mathbb{U}^T\mathbb{U}$ where $\mathbb{I}$ is an $L \times L$ identity matrix, and the host signal vector $\mathbf{C} \in \Re^N$ be mapped to the matrix $\mathbb{C} \in \Re^{L \times \frac{N}{L}}$ by arranging its coefficients into L rows and $\frac{N}{L}$ columns, $\mathbb{C} = [\mathbf{C}_1; \ldots; \mathbf{C}_L]$ where $\mathbf{C}_i \in \Re^{1 \times \frac{N}{L}}$. Let $\hat{\mathbb{C}}$ represent the unitary transformation of $\mathbb{C}$ as

$$\hat{\mathbb{C}} = \mathbb{U}\mathbb{C}, \tag{4.3}$$

where $\hat{\mathbb{C}} = [\hat{\mathbf{C}}_1; \ldots; \hat{\mathbf{C}}_L]$ and $\hat{\mathbf{C}}_i \in \Re^{1 \times \frac{N}{L}}$. In other words, the coefficients of the host signal vector are broken down into L channels, each consisting of $\frac{N}{L}$ coefficients. The watermark signal $\mathbf{W} \in \Re^{\frac{N}{L}}$ is embedded into the coefficients of designated channel(s), *i.e.* $\hat{\mathbf{C}}_1, \ldots, \hat{\mathbf{C}}_L$.

For the general case, let's assume $\mathbf{W}$ is embedded into ith channel coefficients. This yields the embedded signal $\hat{\mathbf{S}}_i = \mathcal{E}(\hat{\mathbf{C}}_i, \mathbf{W})$ at the ith channel while the transform coefficients in the rest of the channels are not changed. Then the transformed and embedded signal $\hat{\mathbb{S}} = [\hat{\mathbf{C}}_1; \ldots; \hat{\mathbf{S}}_i; \ldots; \hat{\mathbf{C}}_L]$ is inverse transformed as

$$\mathbb{S} = \mathbb{U}^T\hat{\mathbb{S}}, \tag{4.4}$$

and mapped to the stego signal vector $\mathbf{S}$. At the detector, the embedded signal is extracted from the stego channel(s) obtained by segmenting and transforming the

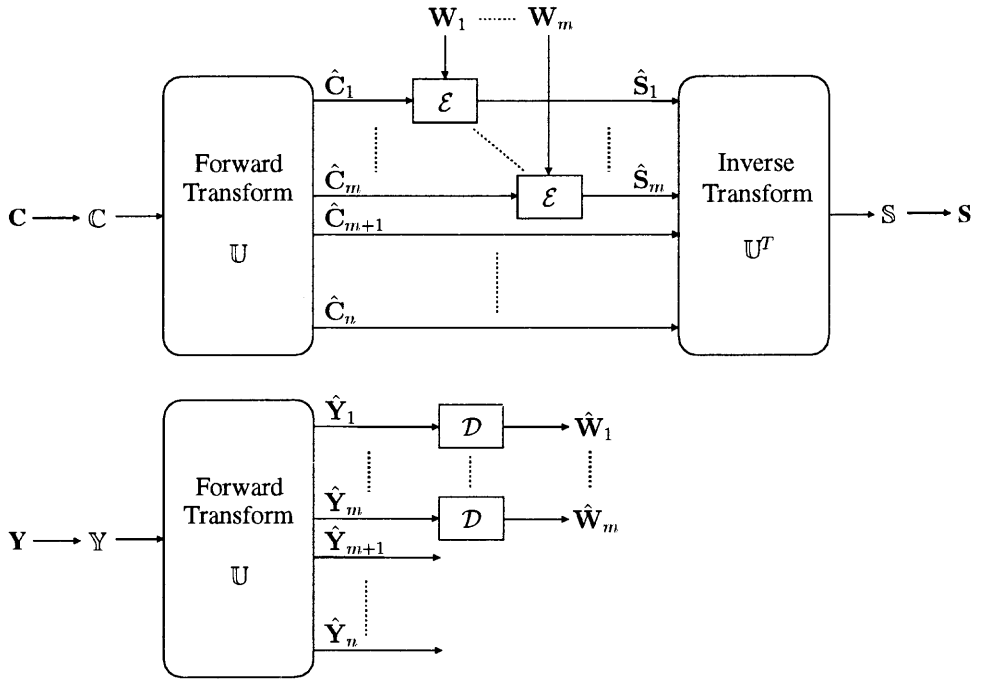


Figure 4.2 Embedding and detection of $\mathbf{W}^T = [\mathbf{W}_1^T, \ldots, \mathbf{W}_m^T]$ into $\mathbf{C}$ with the spreading gain $L = \frac{n}{m}$.

received signal $\mathbf{Y}$. Although only particular transform coefficients are used for data hiding, the resulting embedding distortion, in the transform domain, is spread over all samples in the signal domain. This enables hider to exploit the bandwidth vs. WNR tradeoff at the detector by selecting the spreading factor by choosing $\mathbb{U}$. Spreading factor, in this case, is the ratio of the total number of channels to the number of channels used for data hiding. In order to obtain a spreading factor of $L = \frac{n}{m}$, where L may also be a rational number, m channels of an $n \times n$ unitary transform of the host signal ($\mathbb{C} \in \Re^{n \times \frac{N}{n}}$) are information embedded. Figure 4.2 illustrates this scenario where the first m channels of $\hat{\mathbb{C}}$ are used for hiding the watermark signal $\mathbf{W}^T = [\mathbf{W}_1^T, \ldots, \mathbf{W}_m^T]$ where $\mathbf{W}_i \in \Re^{\frac{N}{n}}$.

The affect of spread transforming on the data hiding rate of a method can be determined in terms of N and WNR. The capacity of any communication scheme, in general, can be expressed in terms of its bandwidth and signal-to-noise ratio. Therefore, the data hiding capacity can be formulated as $C = Nf(WNR)$. Due to

the tradeoff between N and WNR, the capacity with spread transforming takes the form of $C_S = \frac{N}{L} f(L \times WNR)$. Thus, the optimal spreading factor L for a given method can be found from the measured C through maximizing C_S. It should be noted that if the capacity formulation of a scheme is such that the linear increase in the WNR can compensate for the linear reduction in N, then spread transforming offers an improvement in performance. As the performance drop in type-II and type-III methods are exponential in WNR, spreading becomes an efficient tool by enabling them to operate at higher WNR levels where they perform reasonably well. However, for type-I schemes and the upper bound (optimal type-III scheme), where all variables are assumed to be Gaussian, the fall in the hiding rate is logarithmic, $\frac{1}{2}\log_2(\frac{WNR}{WNR \times DWR+1})$ and $\frac{1}{2}\log_2(1+WNR)$, respectively. Consequently, the optimal spreading factor L that maximizes $\frac{1}{2L}\log_2(\frac{L \times WNR}{L \times WNR \times DWR+1})$ or $\frac{1}{2L}\log_2(1+L \times WNR)$ is computed as one.

The hiding rate vs. robustness curves of DM and type-III methods with spread transforming, computed using the results of Figure 3.9, are displayed in Figure 4.3. When compared to the hiding capacity, the hiding rates corresponding to DM and the type-III implementations of DM with Gaussian mapping type of processing improved remarkably in the low WNR range. With spread transforming, distortion compensation and Gaussian mapping types of processing deliver slightly better performances than thresholding type of processing. This is not surprising since the improvement with spreading depends on the performance of the scheme at higher WNRs where distortion compensation and Gaussian mapping types of post-processing were seen to perform better than thresholding type of proccessing, Section 3.3. Measured spreading factors for the methods are shown in Figure 4.4. However, one should be careful since very large spreading factors enable large embedding distortions, *i.e.* increased Δ values, and this may violate the assumption that host signal is uniformly distributed over all quantization cells. Therefore, large

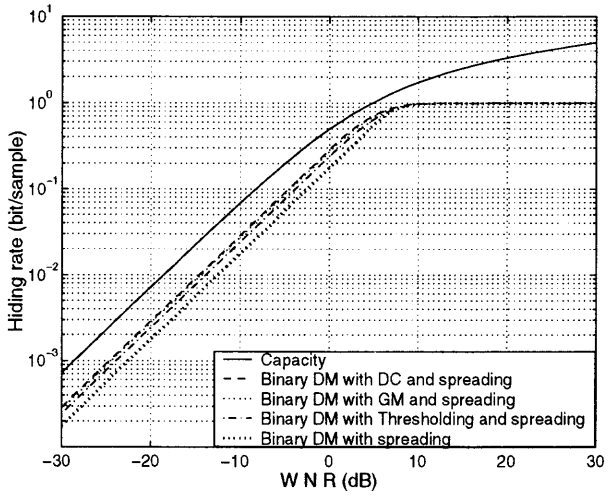


Figure 4.3 The improvement in the hiding rate of type-II and type-III methods when $P = 10$.

spreading factors may not be practically feasible as the embedding operation becomes dependent on the statistics of the host signal.

4.2 Multiple Codebook Data Hiding

When the embedding signal size N is small, multiple codebook data hiding can be used to embed the watermark signal at lower embedding distortion levels. The distortion P introduced to host signal $\mathbf{C}$ due to embedding operation is computed over all stego signal coefficients as $P = \frac{1}{N}||\mathbf{X}_n||^2$. Assuming that the pdf of the host signal is smooth enough, such that it can be considered as uniformly distributed over all quantization intervals, the distortion introduced to each host signal sample C has the statistics of X_n, Equations (3.16), (3.22), and (3.25). In other words, the distortion P is a random variable and its distribution approximates $\mathcal{N}(\sigma^2_{X_n}, \frac{\sigma^2_P}{N})$ where

$$\frac{\sigma^2_P}{N} = \frac{1}{N}\int_{-\infty}^{\infty} x_n^4 f_{X_n}(x_n)\mathrm{d}x_n - (\sigma^2_{X_n})^2. \tag{4.5}$$

Accordingly, when N is large, the distortion P introduced to the host signal becomes $\sigma^2_{X_n}$. However, when N is small, P varies around the mean $\sigma^2_{X_n}$ depending on the distribution of X_n and the signal size N. The variation in the embedding distortion

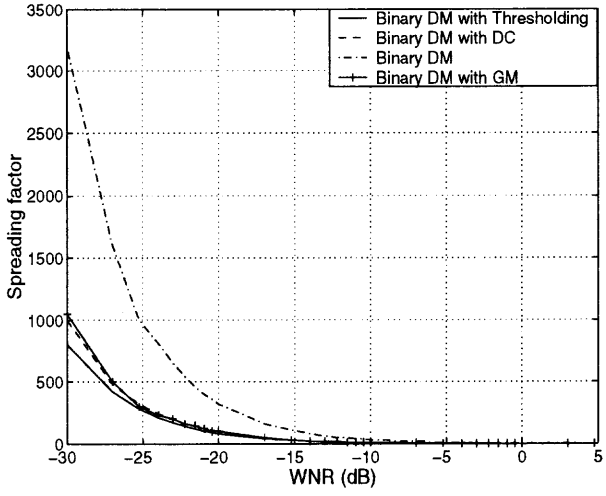


Figure 4.4 Corresponding spreading factors.

becomes more significant with the decreasing value of N. Therefore, embedding in a host signal with limited signal size requires a more careful selection of embedding and detection parameters. In general embedding-detection parameters are optimized to maximize the performance at the given noise level σ_Z^2 and the permitted distortion $\sigma_{X_n}^2$ as described in Section 3.2.3. Therefore, implicitly, a very large embedding signal size N is assumed. Embedding and detection with the parameters obtained through an optimization procedure that disregards this aspect of the problem may cause the data hiding method to operate on a lower hiding rate vs. robustness curve due to the variation in the embedding distortion with respect to N.

Figures 4.5 and 4.6 display the hiding rates corresponding to binary DM with thresholding and distortion compensation types of post-processing for various N values when the embedding distortion deviates from the mean $\sigma_{X_n}^2$ by five times the standard deviation, $P = \sigma_{X_n}^2 - 5\sqrt{\frac{\sigma_P^2}{N}}$. As displayed in figures, with decreasing N, the hiding rate drops in both cases. However, since X_n corresponding to distortion compensation type of post-processing has higher variance around the mean, the reduction in rate is more drastic. These results indicate that, given two host signals with similar statistics, if the same watermark signal is

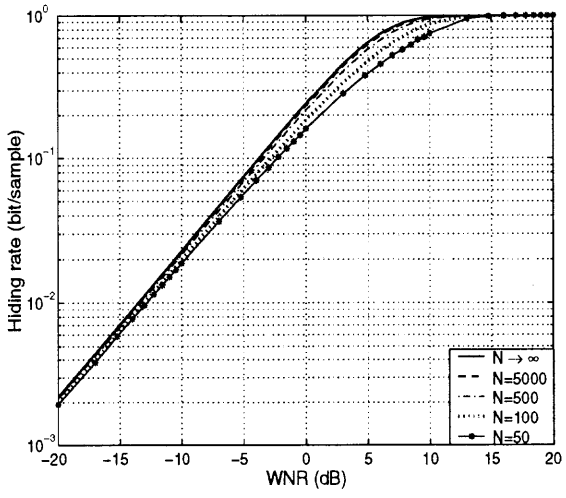


Figure 4.5 Hiding rates corresponding to Binary DM with thresholding for various N when $P = \sigma_{X_n}^2 - 5\sqrt{\frac{\sigma_P^2}{N}}$.

embedded in both signals using the same parameters, the resulting distortion due to embedding may differ significantly for the two signals depending on size N. Therefore, more sophisticated optimization techniques are needed for determining the embedding-detection parameters for limited N. An obvious approach is to fine-tune the parameters obtained with the assumption of large N, so that the resulting distortion is neither above nor below the permitted distortion level. The question now is, can better be done? Can the fact that the embedding distortion has a large variance be utilized to improve the performance of data hiding? It will be soon seen that this is indeed possible. Multiple codebook hiding method exploits this phenomenon by choosing a transformation of $\mathbf{C}$ which yields the minimum embedding distortion when $\mathbf{W}$ is embedded. The ability to embed a watermark signal at a lower embedding distortion, rather than at the permitted distortion level, is translated into more robust embedding of the watermark signal.

The essence of the method is depicted in Figures 4.7 and 4.8 where the embedding signal size is two. In both cases, one of the binary symbols is embedded into a signal vector $\mathbf{c}$ composed of two signal samples, $\mathbf{c} = (c_1, c_2)$, using either a

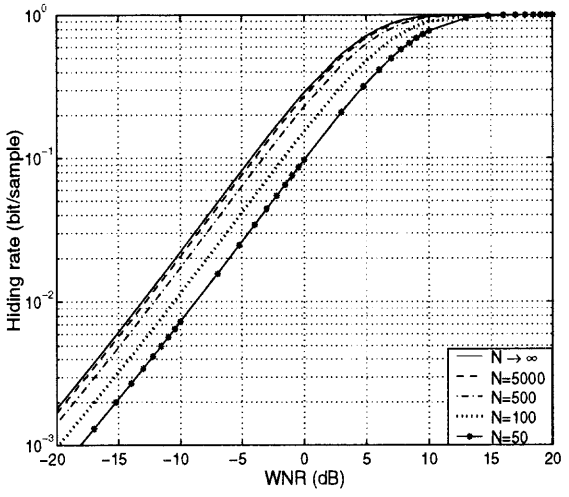


Figure 4.6 Hiding rates corresponding to Binary DM with distortion compensation for various N when $P = \sigma_{X_n}^2 - 5\sqrt{\frac{\sigma_P^2}{N}}$.

two-dimensional lattice or two unidimensional lattices. The lattice points or the reconstruction points associated with each binary sample is marked by $\times$ and $\circ$ symbols. Embedding operation is the translation of the vector $\mathbf{c}$ to the nearest centroid associated with the symbol to be embedded. The decision regions in Figures 4.7 and 4.8 determine the sustainable amount of noise that does not impair the detection performance.

In the considered cases, the binary symbol corresponding to $\times$ is embedded into $\mathbf{c}$ and into two of its transformed (rotated) versions $\mathbf{c}_2$ and $\mathbf{c}_3$. The embedding distortions between the signal pairs $(\mathbf{c}, \hat{\mathbf{c}})$, $(\mathbf{c}_2, \tilde{\mathbf{c}}_2)$, and $(\mathbf{c}_3, \tilde{\mathbf{c}}_3)$ are measured, in terms of Euclidean distance, as d_1, d_2, and d_3, respectively, as displayed in Figure 4.7. Similarly, in Figure 4.8 the resulting embedding distortions are measured as $\sqrt{d_{11}^2 + d_{12}^2}$, $\sqrt{d_{21}^2 + d_{22}^2}$, and $\sqrt{d_{31}^2 + d_{32}^2}$. When $\tilde{\mathbf{c}}_2$ and $\tilde{\mathbf{c}}_3$ are inverse transformed, one can observe that the distortions introduced to $\mathbf{c}$ due to three embedding operations are not the same. For both of the cases depicted in Figures 4.7 and 4.8, $\hat{\mathbf{c}}_2$ (inverse transformed $\tilde{\mathbf{c}}_2$) yields the smallest embedding distortion, d_2. It is important to note that the amount of embedding distortion, due to embedding into

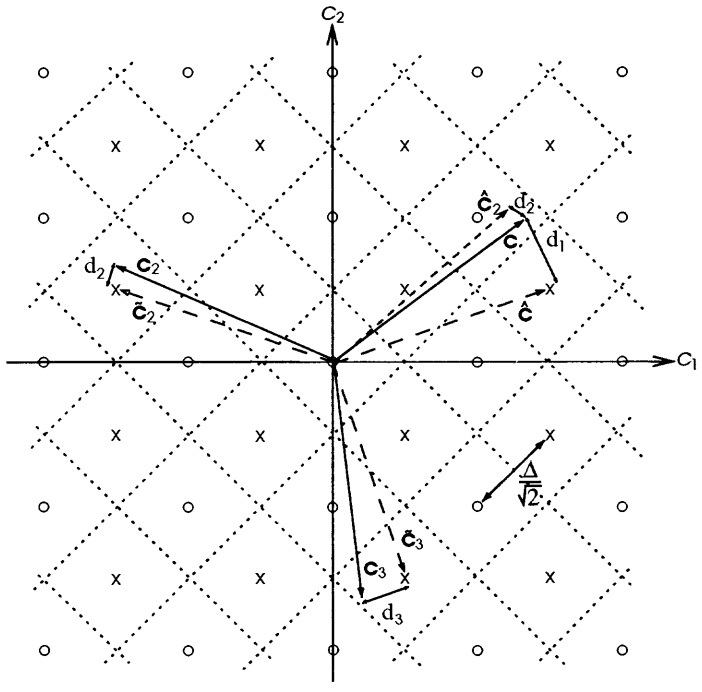


Figure 4.7 Depiction of embedding a binary symbol into the host signal $\mathbf{c} = (c_1, c_2)$ and into its two transformations using a 2-D lattice.

transformations of $\mathbf{c}$, $\mathbf{c}_2$, and $\mathbf{c}_3$, remains the same in magnitude after the inverse transformation since the transformation is assumed to be unitary or energy preserving. One can now easily see that, with the added complexity of transformations, a binary symbol can be embedded into $\mathbf{c}$ at a smaller embedding distortion level. Multiple codebook hiding incorporates these savings in embedding distortion with type-III hiding methodology.

Type-III methods, as described earlier, are derived from type-II methods by increasing the distance between the reconstruction points, and introducing a processing distortion that is also a function of the expected noise level. In type-III methods, the resulting increase in the embedding distortion, due to the increased separation of the reconstruction points, is reduced to the permitted amount by the post-processing while performance is maximized at the expected noise level [43]. In other words, the distortion due to embedding operation is limited to permitted amount P by proper selection of the separation between the reconstruction points

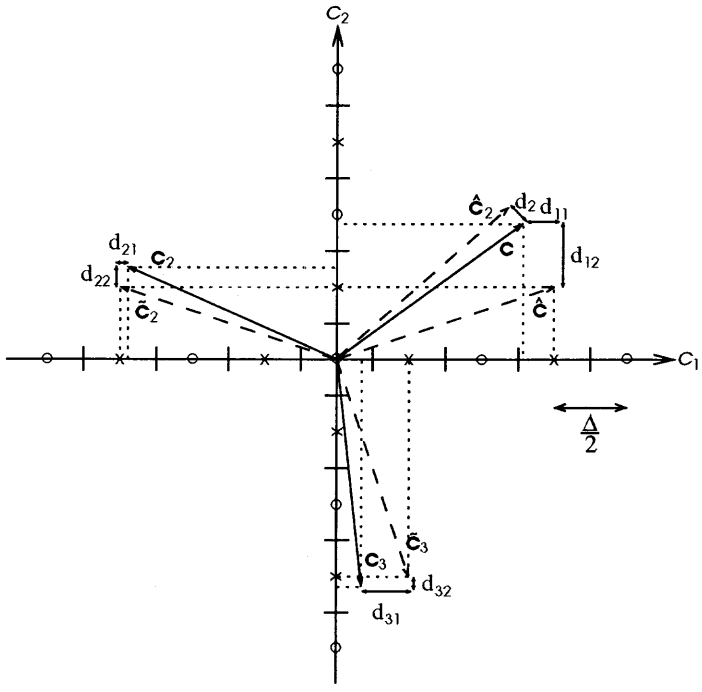


Figure 4.8 Depiction of embedding two binary symbols into the host signal vector $\mathbf{c} = (c_1, c_2)$ and into its two transformations using uniform scalar quantizers.

(Δ) and the amount of processing distortion ($\sigma^2_{X_t}$). The Δ and $\sigma^2_{X_t}$ values that yield the distortion P are not unique, and in order to maintain a fixed distortion level of P, an increase or decrease in either of Δ or $\sigma^2_{X_t}$ values should be followed by the other in the same manner. *Since the employment of transformations enables embedding at lower distortion levels, the difference between the permitted and actual embedding distortions can be utilized by the type-III embedder to either reduce the $\sigma^2_{X_t}$ value at the given Δ or to further increase the Δ value at the fixed $\sigma^2_{X_t}$.* Both actions lead to an improvement in the detection performance.

Employing multiple codebooks resembles the optimal binning technique in the manner that the size of each bin is increased from one to the number of codebooks. Therefore, for a message to be transmitted, the embedder generates a set of codewords and chooses the best among them. Correspondingly, the detector has to search over all codebooks for a successful extraction of the message. Modifying the multiple codebook hiding method by assigning one of the codebooks for embedding and

detection while discarding the others reduces it to a type-III method. Due to this freedom in selecting one of the many codebooks being utilized, the method is referred to as multiple codebook hiding.

In multiple codebook hiding, each codeword is generated from a unitary transformation of the host signal. From this point of view, the design of the ideal codebook requires the derivation of the optimal transform basis for embedding and detection (both at the embedder and detector). This is an impractical task considering the dependency on the host signal. (In Figure 4.7 and Figure 4.8, where $N = 2$, this refers to the transformation that translates $\mathbf{c}$ to a point that coincides with one of the $\times$ points.) Therefore rather than computing the optimal transformation basis, a set of transformation bases is selected with the intention that, for a given host signal some of the bases will yield codewords similar to that of the optimal transformation. Thus, the use of multiple codebooks provide the embedder with a freedom in choosing the best among a number of sub-optimal codewords. However, when $N \longrightarrow \infty$, for any $\mathbf{C}$, the embedding distortion converges to the expected value, $P \longrightarrow \sigma^2_{X_n}$, and multiple codebook hiding does not provide any advantage over single codebook hiding. (In other words, with the increasing N all transformations of $\mathbf{C}$ become equally preferable for embedding as they all yield the same distortion.) On the other hand, detector should be able to differentiate the correct transformation from among all transformations of the received signal, in order to successfully detect the embedded message. Apparently, such a detection of the message is more prone to errors. Ultimately, the question to be answered is whether at a fixed N and permitted embedding distortion, the improvement in the detection performance due to the ability to increase the Δ (or to reduce the $\sigma^2_{X_t}$), can compensate for the additional detection errors due to the uncertainty in the transform basis used for embedding. It is shown that for AWGN channel, Gaussian distributed host signal and squared error distortion measure, the increase in probability of error due to use of

multiple codebooks is compensated by a reduction (in probability of error) due to the embedder's ability to adapt the codeword to the host signal. Type-III schemes like binary DM with thresholding and distortion compensation types of post-processing employing soft decision rule based detectors are incorporated with multiple codebook hiding technique. However, the concept is applicable to all type-III hiding methods.

4.2.1 Channel Model for Multiple Codebook Data Hiding

In the multiple codebook data hiding scenario, information hider and extractor share two sets of information. One is the set of sequences $\mathbf{W}_1, \ldots, \mathbf{W}_M \in \Re^N$ that are associated with M distinct messages the other is the set of L, $N \times N$, unitary transform bases, *i.e.*

$$\mathbb{I} = \mathbb{T}_i^T \mathbb{T}_i, \; i = 1, \ldots, L \tag{4.6}$$

where $\mathbb{I}$ is the $N \times N$ identity matrix and ^{T} denotes the matrix transpose operation. The overall data hiding system is outlined in Equations (4.7) through (4.12) in an additive model as

$$\mathcal{W} \; : \quad m \longrightarrow \mathbf{W}_m, \tag{4.7}$$

$$\hat{\mathbf{S}}_k = \mathcal{E}(\mathbb{T}_k \mathbf{C}, \mathbf{W}_m), \; 1 \leq k \leq L, \tag{4.8}$$

$$\mathbf{S}_k = \mathbb{T}_k^T \hat{\mathbf{S}}_k, \tag{4.9}$$

$$\mathbf{Y} = \mathbf{S}_k + \mathbf{Z} = \mathbf{C} + \mathbf{X}_{n_k} + \mathbf{Z}, \tag{4.10}$$

$$\hat{\mathbf{W}}_m^i = \mathcal{D}(\mathbb{T}_i \mathbf{Y}), \; i = 1, \ldots, L, \tag{4.11}$$

$$\mathcal{W}^{-1} \; : \quad \hat{\mathbf{W}}_m^i \longrightarrow \hat{m}. \tag{4.12}$$

In the model, $\mathbf{C}$ is the *iid* Gaussian distributed host signal with the marginal $C \sim \mathcal{N}(0, \sigma_C^2)$, $\mathbf{X}_n = \mathbf{X}_{n_k}$ is the distortion introduced by the type-III embedder (type-III codeword, Section 3.2) and $\mathbf{Z}$ is the AWGN vector where $Z \sim \mathcal{N}(0, \sigma_Z^2)$. One selection criterion for $\mathbb{T}_i$, $i = 1, \ldots, L$, is to require the transformations of a random

signal vector $\mathbf{r}$ be maximally separated from each other in $\Re^N$ with respect to a predesignated distance measure. For squared error distortion measure, selection of $\mathbb{T}_1, \ldots, \mathbb{T}_L$ is based on the maximization of the following criterion

$$E[||\mathbb{T}_k\mathbf{r} - \mathbb{T}_i\mathbf{r}||^2] \ , 1 \leq i, k \leq L \text{ and } i \neq k \tag{4.13}$$

where the expectation is performed over all $\mathbf{r} \in \Re^N$. Among the L unitary transformations $\mathbf{C}_i = \mathbb{T}_i\mathbf{C}$, $i = 1, \ldots, L$, embedder picks the one that is expected to yield highest detection statistics at the permitted embedding distortion. Assuming k is the index of the selected transform basis, the sequence $\mathbf{W}_m$, corresponding to the message indexed by m, $1 \leq m \leq M$, is embedded into $\mathbb{T}_k$ transformation of the host signal, $\mathbf{C}_k$. Then, the stego signal in the transform domain, $\hat{\mathbf{S}}_k$, is inverse transformed to signal domain, $\mathbf{S}_k$. Uninformed of the particular transform $\mathbb{T}_k$ used for embedding, detector generates L transformations of the received signal $\mathbf{Y}$ and detects the hidden message $\hat{m}$ in a blind manner. With the use of multiple codebooks, the choice of T_k determines the codeword $\mathbf{X}_{n_k}$ among codewords $\{\mathbf{X}_{n_1}, \ldots, \mathbf{X}_{n_L}\}$. Therefore, embedding operation can be viewed as a vectorial operation where embedder chooses one of the L codewords based on the given host signal $\mathbf{C}$ and the message m to be conveyed.

Figure 4.9 displays codeword generation for multiple codebook hiding. Compared to Figure 2.7, the main difference is that for a message index m, L number of codewords are generated by embedding $\mathbf{W}_m$ into $\mathbb{T}_1, \ldots, \mathbb{T}_L$ transformations of $\mathbf{C}$. Consequently, the embedder chooses the best one among the codewords $\mathbf{X}_{1,m}, \ldots, \mathbf{X}_{L,m}$.

Table 4.2.1 lists all the notations used in the analysis in addition to the previous notation, *i.e.* the vectors are denoted by boldfaced characters, the random variables and their realizations are respectively symbolized by the capital letters and the corresponding lower case letters. For the general case all signals are assumed to

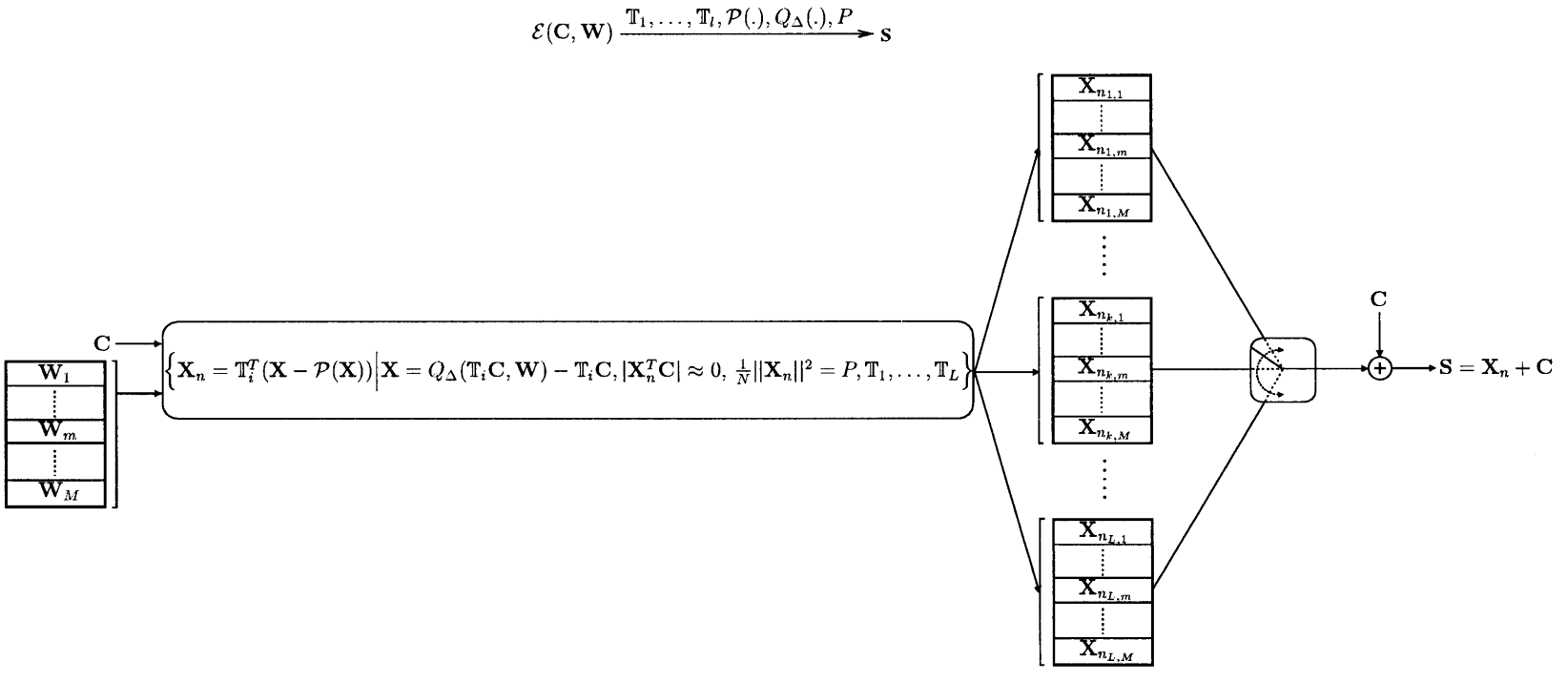


Figure 4.9 Encoding of message index m using multiple codebooks.

be random vectors of size N, however, in some of the derivations individual random variables are used for the sake of simplicity. In such cases vector extensions are straightforward due to *iid* assumption.

Table 4.1 Notation Used in the Chapter

$\mathbf{C}$	Host signal vector
$\mathbf{X}$	Codeword
$\mathbf{S}$	Information hidden signal vector
$\mathbf{Z}$	Channel noise vector
$\mathbf{Y}$	Distorted $\mathbf{S}$
$\mathbf{W}_m$	Watermark signal vector corresponding to message m to be conveyed
$\hat{\mathbf{W}}_m$	Extracted signal vector when $\mathbf{W}_m$ is embedded
$\rho_{m,j}$	The normalized correlation between $\hat{\mathbf{W}}_m$ and $\mathbf{W}_j$
$d_{m,j}$	The mean squared distance between $\hat{\mathbf{W}}_m$ and $\mathbf{W}_j$
$\hat{\mathbf{W}}_m^i$	Extracted signal vector from $\mathbb{T}_i$ transformation of $\mathbf{Y}$ when $\mathbf{W}_m$ is embedded
$\tilde{\mathbf{W}}_m^i$	Extracted signal vector from $\mathbb{T}_i$ transformation of $\mathbf{S}$ when $\mathbf{W}_m$ is embedded
$\rho_{m,j}^i$	The normalized correlation between $\hat{\mathbf{W}}_m^i$ and $\mathbf{W}_j$
$\tilde{\rho}_{m,j}^i$	The normalized correlation between $\tilde{\mathbf{W}}_m^i$ and $\mathbf{W}_j$
$d_{m,j}^i$	The mean squared distance between $\hat{\mathbf{W}}_m^i$ and $\mathbf{W}_j$
$\tilde{d}_{m,j}^i$	The mean squared distance between $\tilde{\mathbf{W}}_m^i$ and $\mathbf{W}_j$

The most crucial step of multiple codebook hiding is the selection of the transformation basis $\mathbb{T}_k$, $1 \leq k \leq L$, which yields the codeword that adapts to $\mathbf{C}$ best at the permitted embedding distortion. For this, the watermark signal $\mathbf{W}_m$ is embedded into L transformations of the host signal, $\mathbf{C}_i = \mathbb{T}_i\mathbf{C}$, $i = 1, \ldots, L$, consecutively. Noting that in a type-III method embedding and detection functions are not inverses of each other, the signal $\mathbf{W}_m$ embedded into $\mathbf{C}_i$ will differ from the

corresponding extraction $\tilde{\mathbf{W}}_m^i$ due to the processing distortion $\mathbf{X}_t$, $\mathcal{D}(\mathcal{E}(\mathbf{C}, \mathbf{W}_m)) \neq \mathbf{W}_m$. Therefore, embedder can decide on the transformation basis by measuring the similarity (or the dissimilarity) between $\mathbf{W}_m$ embedded into all transformations of $\mathbf{C}$ and the corresponding extractions $\tilde{\mathbf{W}}_m^i$ through computing and comparing normalized correlations, $\tilde{\rho}_{m,m}^i$, or mean squared distances, $\tilde{d}_{m,m}^i$. If the decision on the transform basis is made using correlation, *maximum correlation* criterion, the value of i index that yields the highest correlation $\tilde{\rho}_{m,m}^i$, is chosen as the index of the transformation basis $\mathbb{T}_k$, $k = \arg\max_i \left(\tilde{\rho}_{m,m}^i\right)$ for $\tilde{\rho}_{m,m}^i = \frac{\mathbf{W}_m^T \tilde{\mathbf{W}}_m^i}{\|\mathbf{W}_m\| \, \|\tilde{\mathbf{W}}_m^i\|}$. Alternately, if squared error distance is used as the decision metric, or *minimum distance* criterion, the embedder picks the transform basis $\mathbb{T}_k$ that yields the smallest mean squared distance between $\mathbf{W}_m$ and $\tilde{\mathbf{W}}_m^i$, $k = \arg\min_i\{\tilde{d}_i\}$, $i = 1, \ldots, L$ where $\tilde{d}_i = \frac{1}{N} \| \mathbf{W}_m - \tilde{\mathbf{W}}_m^i \|^2$.

Such a selection of the transformation basis can be justified as follows. In order to embed a signal into a host signal, embedder has to determine the optimal embedding parameters depending on the employed post-processing (*i.e.* (Δ, β) for thresholding, (Δ, α) for distortion compensation). These parameters are computed in advance for the permitted embedding distortion (P_E) and the given channel noise (σ_Z^2) levels assuming N is very large and host signal is uniformly distributed in each quantization interval. It should be noted that the embedding parameters computed using the optimization criteria described in Section 3.2 are valid when N is relatively large. However, due to limitation on the size N, the embedding distortion P introduced to $\mathbf{C}$ by using the optimal embedding parameter values differs from P_E. Therefore, embedder has to fine-tune those parameters for the given host signal in order to comply with P_E. Since Δ is also revealed to the extractor, it should remain same for all embedding operations while processing distortion due to the choice of β or α may vary for each embedding. As discussed earlier, β and α designates the amount of processing distortion applied on the type-II codeword due to the post-processing. Ultimately, when $\beta = \Delta$ or $\alpha = 1$ no post-processing is performed and therefore

embedded and extracted watermark signals are the same. On the other hand, when embedding of $\mathbf{W}_m$ with $\beta < \Delta$ and $\alpha < 1$ is considered, the extracted signal, $\tilde{\mathbf{W}}_m$, will be distorted at various levels depending on the amount of processing distortion. Thus, correlation (respectively distance) between the embedded and extracted signals reduces (respectively increases) with decreasing β or α.

The sent message is detected from the received signal $\mathbf{Y}$ without knowing which of the L transformation bases is used for embedding. Hence, the extractor tries all transformations of $\mathbf{Y}$ and extracts signals $\hat{\mathbf{W}}_m^i = \mathcal{D}(\mathbb{T}_i\mathbf{Y})$. Then, the set of extracted signals $\{\hat{\mathbf{W}}_m^1, \ldots, \hat{\mathbf{W}}_m^L\}$, of which only $\hat{\mathbf{W}}_m^k$ is a valid extraction, is compared with the set of watermark signals $\{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$ by computing the normalized correlations $\rho_{m,j}^i$ or mean squared distances $d_{m,j}^i$, where $i = 1, \ldots, L$ and $j = 1 \ldots, M$, depending on the decision metric used at the embedder. Among all (i, j) index pairs, the j index of the pair that maximizes $\rho_{m,j}^i$ or minimizes $d_{m,j}^i$ is the index of the detected message $\hat{m}$, $\hat{m} = \arg_j \max_{i,j} \left(\rho_{m,j}^i\right)$ or $\hat{m} = \arg_j \min_{i,j} \left(d_{m,j}^i\right)$.

Figure 4.10 displays an L codebook embedding and detection scheme. In the block diagram, $\mathbf{W}$ is the watermark signal corresponding to message index m. The decision block b_E, at the embedder side, decides on the transform basis $\mathbb{T}_i$, $1 \leq i \leq L$, to be used for embedding using one of the decision metrics. Then, it transmits the stego signal corresponding to $\mathbf{W}$ and $\mathbf{C}$. At the detector side, b_D detects the message with index $\hat{m}$ by computing the correlations or distances between the extracted signals and the set of watermark signals. A detection error occurs whenever m and $\hat{m}$ are not the same.

With multiple codebook hiding, as mentioned earlier, the embedder is able to better adapt the codeword to the host signal. However, this improvement at the embedder is accompanied by an increase in the probability of detection error. This error is due to two sources of noise: the channel noise and the interference from the other transformations. When extraction is made from the correct transformation of

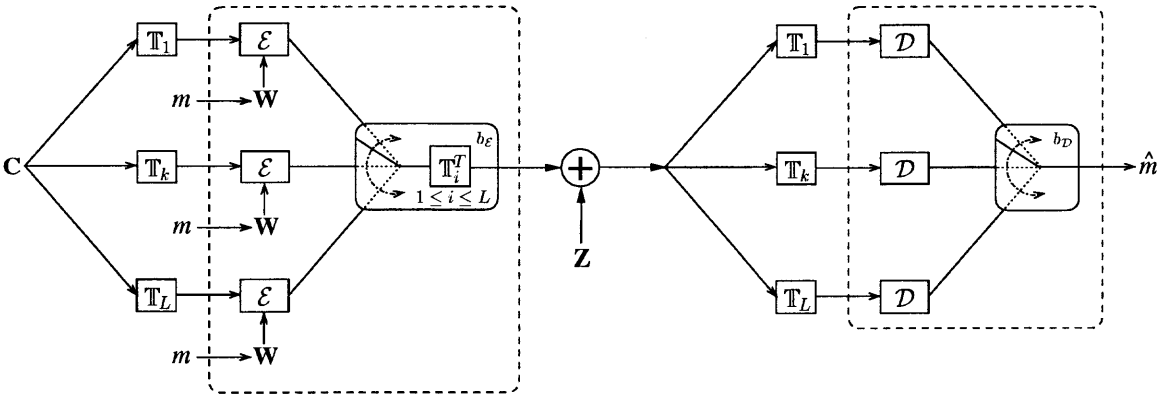


Figure 4.10 Multiple codebook embedding and detection.

the received signal, the sent message may be falsely detected because of the channel's distortion of the stego signal. This is the same as the detection error in single codebook hiding. However, for the multiple codebook case, the error may also be due to the interference from the other $L-1$ transformations. This occurs when detection of a message is from a transformation of the received signal *other* than the transformation used for embedding. This error is independent of the channel noise and can be minimized by the proper selection of the transformation bases.

In Sections 4.2.2-4.2.5, single and multiple codebook hiding methods using *maximum correlation* and *minimum distance* criteria are studied and analyzed with respect to their probability of error performances.

4.2.2 Single Codebook Hiding Based on Maximum Correlation Criterion

Let $\mathbf{W}_m^T = [W_{m_1}, \ldots, W_{m_N}]$ be a length N *iid* zero mean binary random vector corresponding to message m and $\hat{\mathbf{W}}_m^T = [\hat{W}_{m_1}, \ldots, \hat{W}_{m_N}]$ be the extracted real valued signal at the detector. Since the embedding and detection processes are memoryless and both host signal and channel noise are white, $\hat{\mathbf{W}}_m$ is an *iid* zero mean random vector. For the single codebook case, the embedder employs an $M \times N$ sized codebook composed of M length-N codewords. A detection error is due to $\hat{\mathbf{W}}_m$ having the highest correlation with any of $\{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$ other than $\mathbf{W}_m$. Then, an event E_j

that the detector will pick $\hat{m}$ as the detected message instead of m is denoted as

$$\mathrm{E}_j = \{\mathrm{p}(\rho_{m,j} \geq \rho_{m,m})\}, j = 1, \ldots, M \text{ and } j \neq m. \tag{4.14}$$

The event E that detector makes a detection error is expressed as,

$$\mathrm{E} = \bigcup_{j=1, j\neq m}^{M} \mathrm{E}_j. \tag{4.15}$$

Hence, the probability of error for single codebook hiding, P_e^{one}, is expressed as

$$\mathrm{P}_e^{one} = Pr\{\mathrm{E}\} \leq \sum_{j=1, j\neq m}^{M} Pr\{\mathrm{E}_j\}. \tag{4.16}$$

Using Equation (4.14), the upper bound on P_e^{one} can be rearranged as

$$\mathrm{P}_\mathrm{e}^{\mathrm{one}} \leq \sum_{j=1, j\neq m}^{M} \mathrm{p}(\rho_{m,j} \geq \rho_{m,m}). \tag{4.17}$$

In Equation (4.17), $\rho_{m,j}$ and $\rho_{m,m}$ are random variables that are respectively equivalent to random variables ρ_{ind} and ρ_{dep} in their statistics. The relationship of $\rho_{m,j}$, $1 \leq m, j \leq M$, to ρ_{ind} and ρ_{dep} is explained in the following subsections. Based on those results, the pdf of r.v. $\rho_{m,j}$ can be generalized as

$$\rho_{m,j} \sim \begin{cases} \mathcal{N}(0, \frac{1}{N}), & \text{if } m \neq j \\ \rho_{dep}, & \text{if } m = j. \end{cases} \tag{4.18}$$

Assuming m is the index of the transmitted message for all the cases, the first subscript, m, of $\rho_{m,j}$ can be dropped for the sake of simplicity. Thus, Equation (4.17) can be rewritten using Equation (4.18) as

$$\mathrm{P}_e^{one} \leq \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\rho_j}(\rho_j \geq \rho_m) f_{\rho_m}(\rho_m) \mathrm{d}\rho_j \mathrm{d}\rho_m, \tag{4.19}$$

$$\leq \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \left(\int_{\rho_m}^{\infty} f_{\rho_j}(\rho_j) \mathrm{d}\rho_j \right) f_{\rho_m}(\rho_m) \mathrm{d}\rho_m. \tag{4.20}$$

The inner integral in Equation (4.20) can be expressed in terms of Gaussian Q-function, *i.e.* $\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} \mathrm{d}t$. Since statistics of ρ_j are independent of the index j when $j \neq m$, the sum operator in Equation (4.20) can be replaced with the factor $M-1$ and the inequality in P_e^{one} simplifies to

$$\mathrm{P}_e^{one} \leq (M-1) \int_{-\infty}^{\infty} \mathrm{Q}(\rho_m \sqrt{N}) f_{\rho_m}(\rho_m) \mathrm{d}\rho_m. \tag{4.21}$$

Distribution of ρ_{ind}. If $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ have a zero covariance matrix, $\hat{\mathbf{W}}_m$ carries no information about $\mathbf{W}_m$ due to the channel noise, the normalized correlation ρ_{ind} between $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ is defined as

$$\rho_{ind} = \frac{\mathbf{W}_m^T \hat{\mathbf{W}}_m}{||\mathbf{W}_m||\ ||\hat{\mathbf{W}}_m||} = \sum_{l=1}^{l=N} \frac{W_{m_l} \hat{W}_{m_l}}{||\mathbf{W}_m||\ ||\hat{\mathbf{W}}_m||}. \tag{4.22}$$

The r.v. W_{m_l}, $1 \leq l \leq N$, has the variance $\frac{||\mathbf{W}_m||^2}{N}$ due to the *iid* assumption, where $||\mathbf{W}_m||^2$ is the power of $\mathbf{W}_m$. Similarly, the variance of $\hat{W}_{m_l}$ is $\frac{||\hat{\mathbf{W}}_m||^2}{N}$ irrespective of its pdf. Hence, the normalized random variables $\frac{W_{m_l}}{||\mathbf{W}_m||}$ and $\frac{\hat{W}_{m_l}}{||\hat{\mathbf{W}}_m||}$ are both zero mean with variance $\frac{1}{N}$. The normalized correlation ρ_{ind} is a r.v. with the mean $m_{\rho_{ind}}$ and the variance $\sigma^2_{\rho_{ind}}$ calculated as

$$\begin{aligned} m_{\rho_{ind}} &= \sum_{l=1}^{l=N} E[\frac{W_{m_l}}{||\mathbf{W}_m||}] E[\frac{\hat{W}_{m_l}}{||\hat{\mathbf{W}}_m||}], \\ &= 0, \end{aligned} \tag{4.23}$$

$$\begin{aligned} \sigma^2_{\rho_{ind}} &= \sum_{l=1}^{l=N} Var[\frac{W_{m_l}}{||\mathbf{W}_m||}] Var[\frac{\hat{W}_{m_l}}{||\hat{\mathbf{W}}_m||}], \\ &= N \frac{1}{N^2} = \frac{1}{N}. \end{aligned} \tag{4.24}$$

The r.v. ρ_{ind} has approximately Gaussian distribution, due to central limit theorem, $\rho_{ind} \sim \mathcal{N}(0, \frac{1}{N})$.

Similarly, if $\mathbf{W}_m$ and $\mathbf{W}_j$ are independent *iid* random vectors, then $\hat{\mathbf{W}}_m$ is also independent with $\mathbf{W}_j$. Consequently, the normalized correlation $\rho_{m,j} \sim \rho_{ind}$.

Distribution of ρ_{dep}. When $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ are dependent, a similar analysis can be performed. However, in this case, the samples W_{m_l} and $\hat{W}_{m_l}$, $1 \leq l \leq N$ are somewhat correlated. The normalized correlation ρ_{dep}, defined between $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$, is the normalized inner product of the two *iid* correlated random vectors, as given in Equation (4.22).

For relatively small N, the embedding distortion P introduced to $\mathbf{C}$ with the use of optimal embedding parameters (that are computed for large N) becomes a r.v. distributed around $P_E = \sigma^2_{X_n}$ with the variance $\frac{\sigma_P^2}{N}$ as discussed in Section 4.2. Based on the measured distortion P, embedder has to adjust the processing distortion $\mathbf{X}_t$ by changing β or α in order to ensure an embedding distortion of P_E. Consequently, the effective noise level, $\mathbf{Z}_{eff} = \mathbf{Z} - \mathbf{X}_t$, at the detector changes and the embedded signal $\mathbf{W}_m$ is distorted accordingly. The relation between the embedded binary watermark signal samples and the extracted samples is expressed in terms of Z_{eff} as in Equation (3.30). The pdf of Z_{eff} for thresholding and distortion compensation types of processing are given in Equations (3.27) and (3.28) as a function of embedding parameters. Ultimately, the correlation coefficient ρ_{dep} between the dependent $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ can be calculated in terms of embedding parameters, N, and statistics of Z_{eff} and W.

It should be noted that a change in the embedding parameter β or α will induce a similar change on the value of correlation coefficient as they designate the amount of processing distortion applied. When N is not large enough, the embedding distortion P deviates from $P_E = \sigma^2_{X_n}$. This is reflected as a deviation of embedding parameters from their optimal values so that adjusted β or α value yields $P = P_E$. Hence, the correlation of $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ is actually a r.v. conditioned on P, $\rho_{dep}|P$ with the mean m_{ρ^*} and the variance $\sigma^2_{\rho^*}$. The mean m_{ρ^*} is calculated as,

$$m_{\rho^*} = E\left[\frac{\mathbf{W}_m^T \hat{\mathbf{W}}_m}{||\mathbf{W}_m||\ ||\hat{\mathbf{W}}_m||}\right],$$

$$= \frac{E[W_m \hat{W}_m]}{\sqrt{E[W_m^2]E[\hat{W}_m^2]}},$$
$$= \frac{R(1)}{\sqrt{R(2)}}, \tag{4.25}$$

where $E[W_m^p \hat{W}_m^p]$ is the p-th joint moment of random variables W_m and $\hat{W}_m$ and

$$R(p) = 2\sum_{i=0}^{i=\infty} \int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}} \left(\left(\frac{(2i+1)\Delta}{4} - z_{eff} \right) (-1)^i \right)^p f_{Z_{eff}}(z_{eff}) \mathrm{d}z_{eff}. \tag{4.26}$$

Similarly, the variance $\sigma_{\rho^*}^2$ of the r.v. $\rho_{dep}|P$ is expressed as

$$\sigma_{\rho^*}^2 = Var\left[\frac{\mathbf{W}_m^T \hat{\mathbf{W}}_m}{||\mathbf{W}_m||\ ||\hat{\mathbf{W}}_m||} \right]. \tag{4.27}$$

The details of the derivations for the equations (4.25) and (4.27) are given in Appendix B.

The covariance matrix of the *iid* signal vector $\mathbf{W}_m$ and the extracted signal vector $\hat{\mathbf{W}}_m$ is diagonal (*i.e.* $E[W_{m_l} \hat{W}_{m_s}] = 0$, if $l \neq s$, $1 \leq l, s \leq N$). Therefore, the distribution of r.v. $\rho_{dep}|P$ approximates Gaussian distribution, $\rho_{dep}|P \sim \mathcal{N}(m_{\rho^*}, \sigma_{\rho^*}^2)$ with mean and variance as given in Equations (4.25) and (4.27), respectively. The pdf of the r.v. ρ_{dep} is therefore

$$f_{\rho_{dep}}(\rho_{dep}) = \int_{-\infty}^{\infty} f_{\rho_{dep}|P}(\rho_{dep}|P) f_P(P) \mathrm{d}P, \tag{4.28}$$

where $P \sim \mathcal{N}(\sigma_{X_n}^2, \frac{\sigma_P^2}{N})$.

4.2.3 Multiple Codebook Hiding Using Maximum Correlation Criterion

In multiple codebook hiding method, the transmitted codeword, corresponding to a message, is expected to yield highest detection statistics at the presumed noise level σ_Z^2. The embedder achieves this by searching for the transformation basis that yields less processing distortion than the others. This is done by choosing the maximum of the correlations $\tilde{\rho}_{m,m}^i$, $i = 1, \ldots, L$, that are measured between

$\mathbf{W}_m$ embedded into L transformations of $\mathbf{C}$ and the corresponding extractions $\tilde{\mathbf{W}}_m^i$. However, due to channel noise $\mathbf{Z}$, the dependency between the embedded watermark signal and the extracted signal at the detector reduces. Therefore the correlation $\tilde{\rho}_{m,m}^i$, between $\mathbf{W}_m$ and its extracted version from $\mathbf{Y}$, would be less than $\tilde{\rho}_{m,m}^i$ measured at the embedder. The correlation values $\tilde{\rho}_{m,m}^i$ and $\rho_{m,m}^i$ can be calculated from Equation (3.30) for $\mathbf{Z}_{eff} = -\mathbf{X}_t$ and $\mathbf{Z}_{eff} = \mathbf{Z} - \mathbf{X}_t$, respectively. Ultimately, the transformation basis that yields the highest correlation at the embedder will also yield the highest correlation at the detector, $\arg_i \max\left(\tilde{\rho}_{m,m}^i\right) = \arg_i \max\left(\rho_{m,m}^i\right)$.

Let the maximum of $\rho_{m,m}^i$ be denoted by ρ_{max} with the pdf given as

$$\rho_{max} \sim \max\left(\rho_{m,m}^1, \ldots, \rho_{m,m}^L\right) \tag{4.29}$$

where $\rho_{m,m}^i$ are independent random variables with $\rho_{m,m}^i \sim \rho_{dep}$, Section 4.2.2. With multiple codebook hiding, then, detection errors are due to any of the normalized correlation values $\rho_{m,j}^i$, $j \neq m$, being greater than the correlation value ρ_{max}. Compared to the single codebook case, probability of error for multiple codebook hiding, P_e^{mul}, is expected to increase with the number of codebooks as there are L times more number of normalized correlation values that can exceed ρ_{max}. On the other hand, since ρ_{max} is expected to have higher mean than $\rho_{m,m}$, the probability of error for each comparison of the normalized correlations is reduced.

Assuming $\mathbb{T}_k$ is the transformation basis used for embedding in all cases, an event E_j^i that the detector will pick $\hat{m}$ instead of m is denoted as

$$\mathrm{E}_j^i = \{\mathrm{p}(\rho_{m,j}^i \geq \rho_{max})\}, i = 1 \ldots, L,\ j = 1, \ldots, M \text{ and } j \neq m. \tag{4.30}$$

The event E^{mul} that the detector makes an error is

$$\mathrm{E}^{mul} = \bigcup_{i=1}^{L} \bigcup_{j=1, j\neq m}^{M} \mathrm{E}_j^i. \tag{4.31}$$

Hence, the probability of detecting a wrong message for multiple codebook hiding, P_e^{mul}, is obtained as

$$\mathrm{P}_e^{mul} = Pr\{\mathrm{E}^{mul}\} \leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} Pr\{\mathrm{E}_j^i\} \tag{4.32}$$

The union bound on the probability of error can be rewritten using Equation (4.30) as

$$\mathrm{P}_e^{mul} \leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} \Pr(\rho_{m,j}^i \geq \rho_{max}). \tag{4.33}$$

Comparing Equation (4.17) with Equation (4.33), one sees that the advantage of multiple codebook embedding over single codebook embedding is reflected in the statistics of $\rho_{m,m}$ and ρ_{max}.

The distribution of $\rho_{m,j}^i$, $1 \leq j \leq M$ and $1 \leq i \leq L$, can be generalized as

$$\rho_{m,j}^i \sim \begin{cases} \mathcal{N}(0, \frac{1}{N}), & \text{if } i \neq k, \\ \mathcal{N}(0, \frac{1}{N}), & \text{if } i = k \text{ and } j \neq m, \\ \rho_{dep}, & \text{if } i = k \text{ and } j = m, \end{cases} \tag{4.34}$$

The probability of error for multiple codebook hiding, Equation (4.33), can be further rewritten using the above results as

$$\mathrm{P}_{\mathrm{e}}^{\mathrm{mul}} \leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\rho_j^i}(\rho_j^i \geq \rho_{max}) f_{\rho_{max}}(\rho_{max}) \mathrm{d}\rho_j^i \mathrm{d}\rho_{max}, \tag{4.35}$$

$$\leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \left(\int_{\rho_{max}}^{\infty} f_{\rho_j^i}(\rho_j^i) \mathrm{d}\rho_j^i \right) f_{\rho_{max}}(\rho_{max}) \mathrm{d}\rho_{max}, \tag{4.36}$$

where the first subscript referring to the transmitted message m is dropped. Since the inner integral in Equation (4.36) is the Gaussian Q function and does not depend on the index j, Equation (4.36) can be simplified to

$$\mathrm{P}_e^{mul} \leq L(M-1) \int_{-\infty}^{\infty} \mathrm{Q}(\rho_{max}\sqrt{N}) f_{\rho_{max}}(\rho_{max}) \mathrm{d}\rho_{max}. \tag{4.37}$$

Note that for $L = 1$, P_e^{mul} given in Equation (4.37) reduces to P_e^{one} in Equation (4.21).

Distribution of $\rho_{m,j}^i$. The distribution of the random variables $\rho_{m,j}^i$ can be found based on the choice of i and j. When detector assumes $i = k$, the transformations used for embedding and detection are the same. Then, the extracted signal $\hat{\mathbf{W}}_m^k$ is expressed as

$$\hat{\mathbf{W}}_m^k = \mathcal{D}\left(\mathbb{T}_k\left(\mathbb{T}_k^T\mathcal{E}\left(\mathbb{T}_k\mathbf{S}, \mathbf{W}_m\right) + \mathbf{Z}\right)\right), \tag{4.38}$$

$$= \mathcal{D}\left(\mathcal{E}\left(\mathbb{T}_k\mathbf{C}, \mathbf{W}_m\right) + \mathbf{Z}'\right). \tag{4.39}$$

Since $\mathbf{Z}$ is assumed to be a white noise vector (*iid* Gaussian), a unitary transformation of it, $\mathbf{Z}' = \mathbb{T}_k\mathbf{Z}$, is also *iid* Gaussian with the same mean and variance. Therefore, the results of the analysis given in Sections 4.2.2 and 4.2.2 also apply to multiple codebook hiding. Consequently, the normalized correlation $\rho_{m,j}^k$, $1 \leq j \leq M$, is equivalent to random variables ρ_{dep} and ρ_{ind} in its statistics respectively for $j = m$ and $j \neq m$.

If there is a mismatch between the embedding and detection transformations, $i \neq k$, then $\hat{\mathbf{W}}_m^i$ is obtained as

$$\hat{\mathbf{W}}_m^i = \mathcal{D}\left(\mathbb{T}_i\left(\mathbb{T}_k^T\mathcal{E}\left(\mathbb{T}_k\mathbf{C}, \mathbf{W}_m\right) + \mathbf{Z}\right)\right), \tag{4.40}$$

$$= \mathcal{D}\left(\mathbb{T}_i\mathbb{T}_k^T\mathcal{E}\left(\mathbb{T}_k\mathbf{C}, \mathbf{W}_m\right) + \mathbf{Z}'\right). \tag{4.41}$$

where $\mathbf{Z}' = \mathbb{T}_i\mathbf{Z}$. In Equation (4.41), $\hat{\mathbf{W}}_m^i$ is related to $\mathbf{W}_m$ through the transformation $\mathbb{T}_i$ followed by a non-linear detection, Section 3.2. For properly selected transform bases, *i.e.* $E[||\mathbb{T}_i\mathbf{C} - \mathbb{T}_k\mathbf{C}||]$ is maximized. An extraction from $\mathbb{T}_i$ transformation of the received signal does not provide any meaningful information about $\mathbf{W}_m$ since embedding transformation was $\mathbb{T}_k$. Consequently, the binary distributed $\mathbf{W}_m$ with values in $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$ is extracted, $\hat{\mathbf{W}}_m^i$, as a uniformly distributed sample sequence in the range $[-\frac{\Delta}{4}, \frac{\Delta}{4}]$ which is independent from $\mathbf{W}_m$. Therefore, the normalized correlation $\rho_{m,j}^i$, $i \neq k$ and $\forall j$, has the same statistics as the r.v. ρ_{ind}, $\rho_{m,j}^i \sim \mathcal{N}(0, \frac{1}{N})$.

Distribution of ρ_{max}. The r.v. ρ_{max} is the maximum of L random variables, Equation (4.29), that are all distributed according to pdf of r.v. ρ_{dep}. The distribution of ρ_{max}, for any finite L, can be expressed in terms of the distribution function of ρ_{dep} as

$$F_{\rho_{max}}(\rho_{max}) = F^{L}_{\rho_{dep}}(\rho_{max}) \tag{4.42}$$

where $F_X(x) = \int_{-\infty}^{x} f_X(x)\mathrm{d}x$ and the superscript ^{L} refers to the L th order power of the distribution function $F_{\rho_{dep}}(\rho_{max})$. Correspondingly, the pdf of ρ_{max} is found as $f_{\rho_{max}}(\rho_{max}) = LF^{L-1}_{\rho_{dep}}(\rho_{max}) f_{\rho_{dep}}(\rho_{max})$.

4.2.4 Single Codebook Hiding Using Minimum Distance Criterion

Considering the minimum distance criterion for the single codebook hiding case, a detection error is the result of $\hat{\mathbf{W}}_m$ having the smallest distance with any of $\{\mathbf{W}_1, \ldots, \mathbf{W}_M\}$ other than $\mathbf{W}_m$. Hence, the upper bound on the probability of detection error, $\mathrm{P}_{\mathrm{e}}^{\mathrm{one}}$, can be expressed similar to Section 4.2.2, Equations (4.14)-(4.17), as

$$\mathrm{P}_e^{one} \leq \sum_{j=1, j\neq m}^{M} \mathrm{p}(d_{m,j} \leq d_{m,m}). \tag{4.43}$$

As will be shown in the following sections, the statistics of the random variables $d_{m,j}$ and $d_{m,m}$, in Equation (4.43), are respectively the same as those of d_{ind} and d_{dep}. Consequently, the pdf of r.v. $d_{m,j}$, $1 \leq m, j \leq M$, can be expressed as

$$d_{m,j} \sim \begin{cases} \mathcal{N}(\frac{\Delta^2}{12}, \frac{\Delta^4}{N180}), & \text{if } m \neq j \\ d_{dep}, & \text{if } m = j. \end{cases} \tag{4.44}$$

Assuming m is the index of the transmitted message for the generic case, Equation (4.43) can be rewritten using Equation (4.44) as

$$\mathrm{P}_{\mathrm{e}}^{\mathrm{one}} \leq \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{d_j}(d_j \leq d_m) f_{d_m}(d_m) \mathrm{d}d_j \mathrm{d}d_m, \tag{4.45}$$

$$\leq \sum_{j=1,j\neq m}^{M} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{d_m} f_{d_j}(d_j) \mathrm{d}d_j \right) f_{d_m}(d_m) \mathrm{d}d_m, \tag{4.46}$$

$$\leq (M-1) \int_{-\infty}^{\infty} F_{d_j}(d_m) f_{d_m}(d_m) \mathrm{d}d_m \tag{4.47}$$

where $F_{d_j}(d_j)$ is the probability distribution function of the r.v. d_j.

Distribution of d_{ind}. When $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ have a zero covariance matrix, the distance d_{ind} between the iid $\mathbf{W_m}$ and $\hat{\mathbf{W}}_m$ can be defined as

$$\begin{aligned} d_{ind} &= \frac{1}{N} \parallel \mathbf{W}_m - \hat{\mathbf{W}}_m \parallel^2, \\ &= \frac{1}{N} (\mathbf{W}_m - \hat{\mathbf{W}}_m)^T (\mathbf{W}_m - \hat{\mathbf{W}}_m), \\ &= \frac{1}{N} \sum_{l=1}^{l=N} (W_{m_l}^2 + \hat{W}_{m_l}^2 - 2W\hat{W}_{m_l}). \end{aligned} \tag{4.48}$$

Introducing the random variable $\lambda = W^2 + \hat{W}^2 - 2W\hat{W}$, such that $d_{ind} = \frac{1}{N} \sum_{l=1}^{l=N} \lambda_{m_l}$, the statistics of random variable d_{ind} can be computed in terms of the statistics of λ. The mean and variance of λ are respectively derived in Appendix B as

$$m_\lambda = \frac{\Delta^2}{12}, \tag{4.49}$$

$$\sigma_\lambda^2 = \frac{\Delta^4}{180}. \tag{4.50}$$

Therefore,

$$\begin{aligned} m_{d_{ind}} &= E[\frac{1}{N} \sum_{j=1}^{j=N} \lambda_{m_l}], \\ &= \frac{1}{N} N m_\lambda = \frac{\Delta^2}{12}, \end{aligned} \tag{4.51}$$

$$\begin{aligned} \sigma_{d_{ind}}^2 &= Var[\frac{1}{N} \sum_{j=1}^{j=N} \lambda], \\ &= \frac{1}{N^2} N \sigma_\lambda^2 = \frac{1}{N} \frac{\Delta^4}{180}. \end{aligned} \tag{4.52}$$

As both $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ are *iid*, the distribution of d_{ind} approximates Gaussian, $d_{ind} \sim \mathcal{N}(\frac{\Delta^2}{12}, \frac{\Delta^4}{N180})$. Similarly, the distance $d_{i,j}$ measured between the extracted signal $\hat{\mathbf{W}}_i$ and the watermark signal $\mathbf{W}_j$, is equivalent to d_{ind} in its statistics when $\mathbf{W}_i$ and $\mathbf{W}_j$ are mutually independent *iid* random vectors.

Distribution of d_{dep}. When $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$ have a diagonal covariance matrix, an analysis similar to the one given in Section 4.2.2 is performed. The distance d_{dep} is the mean squared difference of the *iid* correlated random vectors $\mathbf{W}_m$ and $\hat{\mathbf{W}}_m$, as defined in Equation (4.48). Given that optimal embedding parameters yield an embedding distortion of P, the distance between $\hat{\mathbf{W}}_m$ and $\mathbf{W}_m$ can be expressed as a r.v. conditioned on P. The mean m_{d^*} and the variance $\sigma^2_{d^*}$ of $d_{dep}|P$ can be calculated in terms of the statistics of λ_{m_l} as

$$\begin{aligned} m_{d^*} &= E\left[\frac{1}{N}\sum_{l=1}^{N}\lambda_{m_l}\right], \\ &= \left(\frac{\Delta}{4}\right)^2 - 2\frac{\Delta}{4}R(1) + R(2), \end{aligned} \tag{4.53}$$

$$\begin{aligned} \sigma^2_{d^*} &= Var\left[\frac{1}{N}\sum_{l=1}^{N}\lambda_{m_l}\right], \\ &= \left(\frac{\Delta}{4}\right)^4 - 4\left(\frac{\Delta}{4}\right)^3 R(1) + \frac{1}{N}\left(R(4) + 6\left(\frac{\Delta}{4}\right)^2 R(2) - \Delta R(3)\right) \\ &\quad \frac{N-1}{N}\left(2\left(\frac{\Delta}{4}\right)^2 R(2) + 2\left(\frac{\Delta}{4}\right)^2 R(1)^2 + 4\left(\frac{\Delta}{4}\right)^2 R(1)^2 - \Delta R(1)R(2)\right. \\ &\quad \left. + R(2)^2\right) - m^2_{d^*}, \end{aligned} \tag{4.54}$$

where $R(p)$ is as given in Equation (4.26). Derivation details for Equations (4.53) and (4.54) are given in Appendix B, Equations (B.12)-(B.13). The distribution of $d_{dep}|P$ also converges to a Gaussian distribution, $d_{dep}|P \sim \mathcal{N}(m_{d^*}, \sigma^2_{d^*})$. The pdf of r.v. d_{dep}

is calculated as

$$f_{d_{dep}}(\rho_{dep}) = \int_{-\infty}^{\infty} f_{d_{dep}|P}(d_{dep}|P) f_P(P) \mathrm{d}P, \tag{4.55}$$

where $P \sim \mathcal{N}(\sigma_{X_n}^2, \frac{\sigma_P^2}{N})$.

4.2.5 Multiple Codebook Hiding Using Minimum Distance Criterion

In this version of the method, embedder selects the transformation basis by choosing the minimum of the distances $\tilde{d}_{m,m}^i$, $i = 1, \ldots, L$ computed between $\mathbf{W}_m$ and $\tilde{\mathbf{W}}_m^i$ for each transformations of $\mathbf{C}$. At the detector, on the other hand, the distance between the embedded and extracted signals is measured as $d_{m,m}^i$, $1 \leq i \leq L$. The degradation in the measured distance from $\tilde{d}_{m,m}^i$ to $d_{m,m}^i$ is due to the channel noise $\mathbf{Z}$ as discussed in Section 4.2.3. However, the transformation basis that yields the minimum distance at the embedder will yield the minimum distance at the detector, $\arg_i \min(\tilde{d}_{m,m}^i) = \arg_i \min\left(d_{m,m}^i\right)$. Defining the minimum of $d_{m,m}^i$ as d_{min}, its pdf is given as

$$d_{min} \sim \min\left(d_{m,m}^1, \ldots, d_{m,m}^L\right) \tag{4.56}$$

where $d_{m,m}^i$ are independent random variables with $d_{m,m}^i \sim d_{dep}$, Section 4.2.4. Consequently, a detection error occurs if any of the distance values $d_{m,j}^i$, $1 \leq j \leq M$, $j \neq m$, and $1 \leq i \leq L$, is smaller than d_{min}. Compared to the single codebook case, similar to Section 4.2.3, probability of error is expected to increase with respect to the number of codebooks since there are L times more distance values that may be smaller than d_{min}. Whereas d_{min} has lower mean than $d_{m,m}$ which will reduce the probability of error. The union bound on the probability of error for multiple codebook hiding, P_e^{mul}, is found similar to Equations 4.30-4.33 as

$$\mathrm{P}_e^{mul} \leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} \Pr(d_{m,j}^i \leq d_{min}). \tag{4.57}$$

The advantage of multiple codebook hiding stems from the difference in the distributions of the random variables $d_{m,m}$ and d_{min} in Equations (4.43) and (4.57), respectively. The distribution of $d^i_{m,j}$, $1 \leq j \leq M$ and $1 \leq i \leq L$, can be generalized as

$$d^i_{m,j} \sim \begin{cases} \mathcal{N}(\frac{\Delta^2}{12}, \frac{\Delta^4}{N180}), & \text{if } i \neq k, \\ \mathcal{N}(\frac{\Delta^2}{12}, \frac{\Delta^4}{N180}), & \text{if } i = k \text{ and } j \neq m, \\ d_{dep}, & \text{if } i = k \text{ and } j = m. \end{cases} \tag{4.58}$$

The bound on the probability of error given in Equation (4.57) can be rewritten using the above results (by dropping the first subscript referring to the transmitted message m) as

$$\mathrm{P}_e^{mul} \leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{d^i_j}(d^i_j \leq d_{min}) f_{d_{min}}(d_{min}) \mathrm{d}d^i_j \mathrm{d}d_{min}, \tag{4.59}$$

$$\leq \sum_{i=1}^{L} \sum_{j=1, j\neq m}^{M} \int_{-\infty}^{\infty} \left(\int_{-\infty}^{d_{min}} f_{d^i_j}(d^i_j) \mathrm{d}d^i_j \right) f_{d_{min}}(d_{min}) \mathrm{d}d_{min}, \tag{4.60}$$

$$\leq L(M-1) \int_{-\infty}^{\infty} F_{d^i_j}(d_{min}) f_{\mathbf{d_{min}}}(d_{min}) \mathrm{d}d_{min}. \tag{4.61}$$

where $d^i_j \sim \mathcal{N}(m_{d_{ind}}, \sigma^2_{d_{ind}})$.

Distribution of $d^i_{m,j}$. The distribution of the random variables $d^i_{m,j}$ can be found based on the choice of i and j, as in Section 4.2.3. When detector assumes $i = k$, the transformations used for embedding and detection are the same. The detected watermark signal $\hat{\mathbf{W}}^k_{m,j}$ can be expressed as in Equation (4.39). Thus, the analysis given for single codebook case also applies to multiple codebook case. The distance between the $\mathbf{W}_m$ and $\hat{\mathbf{W}}^k_{m,j}$, $d^k_{m,j}$ for $1 \leq j \leq M$ and $j \neq m$, has the same statistics with the r.v. d_{ind}, $d^k_{m,j} \sim \mathcal{N}(\frac{\Delta^2}{12}, \frac{\Delta^4}{N180})$. In the same manner, $d^k_{m,m}$, $j = m$, has the same statistics with the r.v. d_{dep}, $d^k_{m,m} \sim d_{dep}$.

If there is a mismatch between the embedding and detection transformations such that $i \neq k$, then $\hat{\mathbf{W}}^k_{m,j}$ is obtained as in Equation (4.41). Due to the transformation $\mathbb{T}_i$, $i \neq k$, and the non-linear detection that follows it, $\hat{\mathbf{W}}^k_{m,j}$ becomes independent of $\mathbf{W}_m$. Therefore, the mean squared distance, $d^i_{m,j}$ for $i \neq k$, is equivalent to the r.v. d_{ind} in its statistics, $d^i_{m,j} \sim \mathcal{N}(\frac{\Delta^2}{12}, \frac{\Delta^4}{N180})$.

Distribution of d_{min}. Since, d_{min} is the minimum of L independent random variables, Equation (4.56), distributed according to $F_{d_{dep}}(d_{dep})$, the probability distribution function of d_{min} is found as

$$F_{d_{min}}(d_{min}) = 1 - \Big(1 - F_{d_{dep}}(d_{min})\Big)^L. \tag{4.62}$$

The pdf of r.v. d_{min} is therefore $f_{d_{min}}(d_{min}) = L\Big(1 - F_{d_{dep}}(d_{min})\Big)^{L-1} f_{d_{dep}}(d_{min})$.

4.2.6 Comparisons

The robustness measure used to compare multiple codebook hiding with single codebook hiding is defined in terms of the ratio between the embedding distortion power and the channel noise power, WNR= $\frac{P_E}{\sigma_Z^2}$. Figures 4.11-4.13 and 4.14-4.16 display the union bound on the probability of error for thresholding type of post-processing using both criteria. The curves are obtained by numerically solving Equations (4.37) and (4.61) at different WNRs and for various numbers of codebooks and codebook sizes $M \times N$. Corresponding results for distortion compensation type of post-processing are similarly displayed in Figures 4.17-4.19 and 4.20-4.22. In all cases, as the number of codebooks increases, the bound on the probability of error decreases exponentially. On the other hand, the probability of error for single codebook hiding also decreases with the increasing signal size N. Consequently, a lesser number of codebooks is required to further improve the performance. Results show that for WNR$\geq$ 1 and WNR$\geq$ 0.2 (equivalently in logarithmic scale WNR$\geq$ 0 dB and WNR$\geq$ -7 dB) the use of multiple codebooks is not necessary if $N \simeq 100$

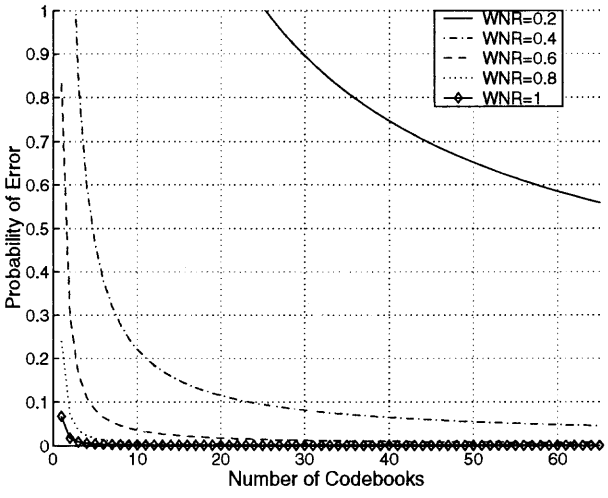


Figure 4.11 Probability of error performance for multiple codebook hiding based on maximum correlation criterion and thresholding type of processing for M=100 and N=50.

and $N \simeq 500$, respectively. Intuitively, this is due to the increasing confidence in the detection with the increasing N. With reference to the analyses in Sections 4.2.3 and 4.2.5, as $m_{\rho_{dep}}$ increases and $\sigma^2_{\rho_{dep}}$ decreases, the maximum of the ensemble of random variables $\tilde{\rho}^1_{m,m}, \ldots, \tilde{\rho}^L_{m,m}$ is less likely to differ from the rest. Respectively, as $m_{d_{dep}}$ decreases the minimum of $\tilde{d}^1_{m,m}, \ldots, \tilde{d}^L_{m,m}$ will not differ significantly from any of the other measured distances. Consequently, all codebooks become almost equally favorable.

In multiple codebook hiding method, since detector forces the extracted signal to match one of the watermark signals, one concern is the probability of false-positive (false-alarm). This is the probability of detecting a message when no message is embedded, and it can be derived based on the results of analysis given in Section 4.2.2 and Section 4.2.3. Under the assumption that host signal is distributed uniformly in each quantization interval ($\sigma_C^2 \gg \Delta$), the extracted signal $\hat{\mathbf{W}}_{null}$ is *iid* uniformly distributed in $[-\frac{\Delta}{4}, \frac{\Delta}{4}]$ and uncorrelated with any of the watermark signals. As a result, the normalized correlation $\rho_{null,j}$ or the squared error distance $d_{null,j}$ between

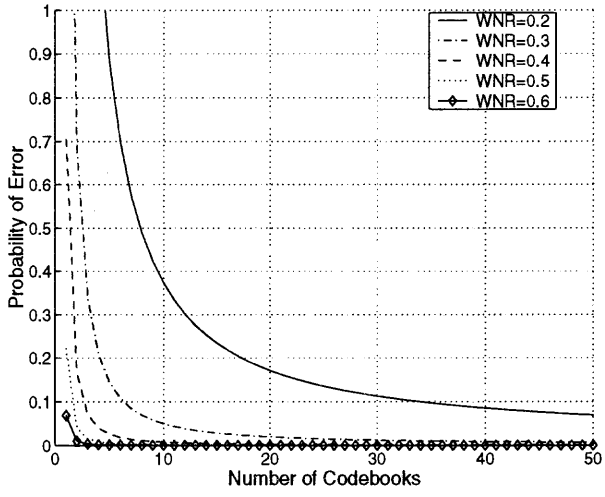


Figure 4.12 Probability of error performance for multiple codebook hiding based on maximum correlation criterion and thresholding type of processing for M=200 and N=100.

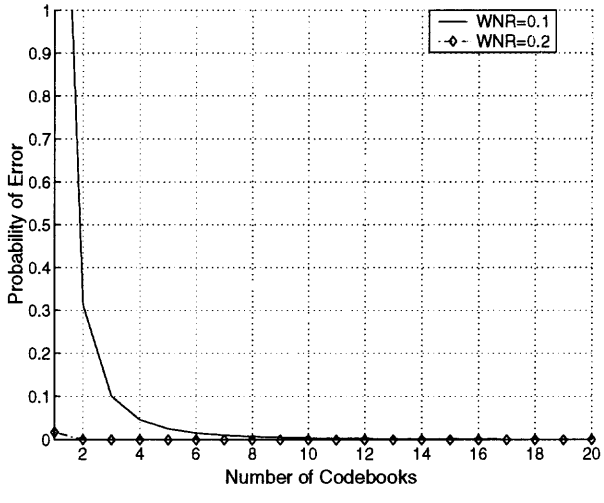


Figure 4.13 Probability of error performance for multiple codebook hiding based on maximum correlation criterion and thresholding type of processing for M=1000 and N=500.

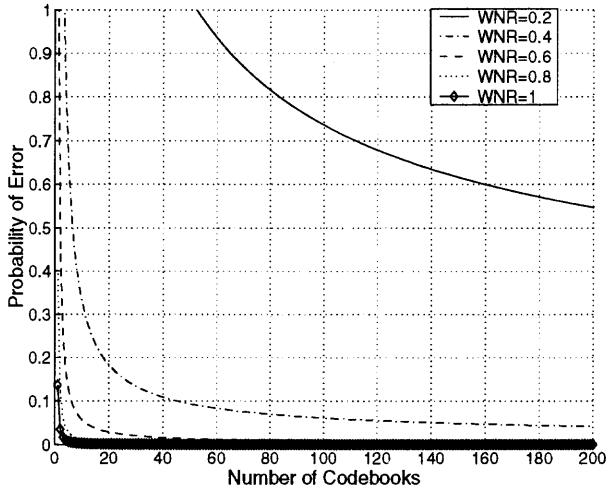


Figure 4.14 Probability of error performance for multiple codebook hiding based on minimum distance criterion and thresholding type of processing for M=100 and N=50.

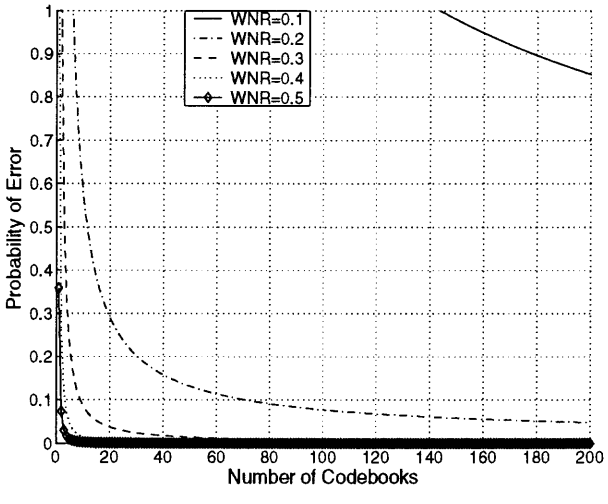


Figure 4.15 Probability of error performance for multiple codebook hiding based on minimum distance criterion and thresholding type of processing for M=200 and N=100.

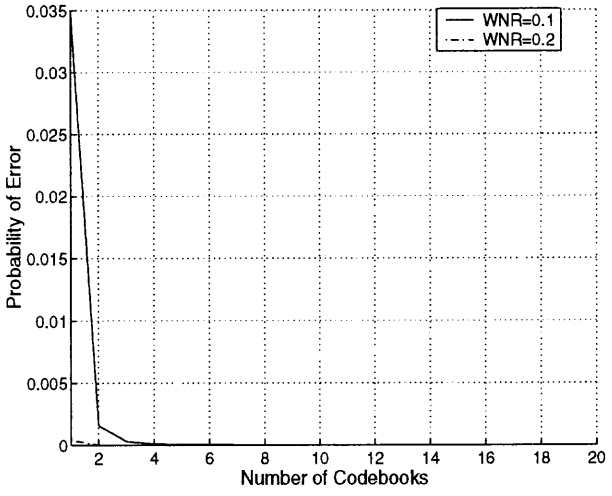


Figure 4.16 Probability of error performance for multiple codebook hiding based on minimum distance criterion and thresholding type of processing for M=1000 and N=500.

$\hat{\mathbf{W}}_{null}$ and $\mathbf{W}_j$, $1 \leq j \leq M$ is distributed as $\mathcal{N}(0, \frac{1}{N})$ *irrespective of the channel noise level.*

For single codebook hiding, a false-positive occurs when $\rho_{null,j}$ is greater or $d_{null,j}$ is smaller than a preset threshold. Using maximum correlation criterion, the threshold is set based on the statistics of ρ_{dep}, which is the normalized correlation between an embedded watermark signal and its extracted version, so that the embedded message can be distinguished from the rest at a constant false-alarm rate. Respectively using minimum distance criterion, the threshold is determined based on the statistics of d_{dep}.

With multiple codebook hiding, where extractions are made from unitary transformations of the received signal, the extracted signals $\hat{\mathbf{W}}^i_{null}$, $1 \leq i \leq L$, have the same statistics from $\hat{\mathbf{W}}_{null}$. Consequently, the correlation $\rho^i_{null,j}$ and the distance $d^i_{null,j}$, computed between $\hat{\mathbf{W}}^i_{null}$ and $\mathbf{W}_j$, have same statistics with $\rho_{null,j}$, and $d_{null,j}$, respectively. Correspondingly, the probability of false-positive is due to $\rho^i_{null,j}$ being greater or $d^i_{null,j}$ being smaller than the preset threshold. Considering a fixed threshold for message detection, the false-alarm rate within multiple codebook

hiding increases with a factor of L compared to single codebook hiding (as there are so many comparisons that may yield a false positive). However, noting that the use of multiple codebooks enables embedding a watermark signal with less processing distortion, the correlation and distance properties of the extracted signal are improved. Therefore, using maximum correlation criterion, one can afford to increase the threshold in accordance with the statistics of ρ_{max}. Alternately, using minimum distance criterion, the threshold can be decreased depending on the statistics of d_{min}.

The numerical solutions of Equation (4.37) indicates that the increase in the P_e^{mul} by the factor of L, compared to P_e^{one}, is compensated by embedder's ability to better adapt the codeword to the host signal as a result of which detection statistics are improved from those of ρ_{dep} to ρ_{max}. Similarly, the linear increase in false-alarm rate with the number of codebooks can be compensated by an exponential decrease trough proper selection of the threshold which relies on the statistics of ρ_{max} rather than of ρ_{dep}. A similar reasoning based on solution of Equation (4.61) is valid for minimum distance criterion due to the improvement in distance properties from d_{dep} to d_{min}.

A complete comparison of multiple codebook hiding and single codebook hiding schemes would involve calculating the actual probability of errors (not the union bound), which would be extremely difficult. However, the analytical results indicate that, Equations (4.37) and (4.61), the upper bound on the probability of error decreases exponentially for multiple codebook hiding scheme. Therefore, schemes employing multiple codebooks, rather than a single codebook, will perform better when N is limited.

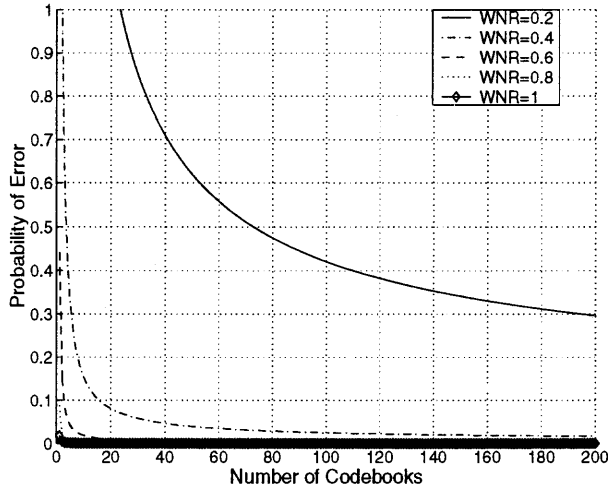


Figure 4.17 Probability of error performance for multiple codebook hiding based on maximum correlation criterion and distortion compensation type of processing for M=100 and N=50.

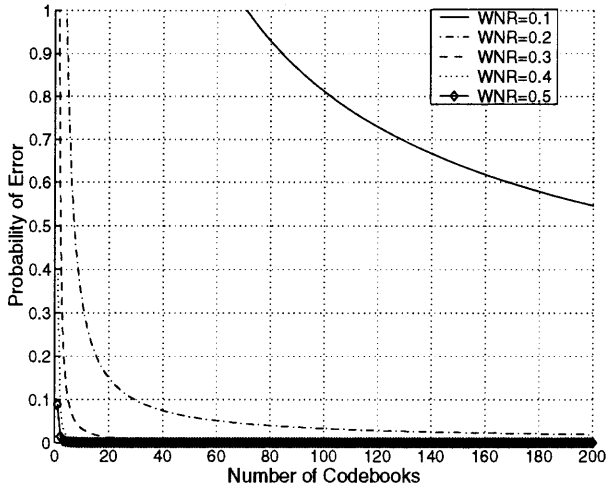


Figure 4.18 Probability of error performance for multiple codebook hiding based on maximum correlation criterion and distortion compensation type of processing for M=200 and N=100.

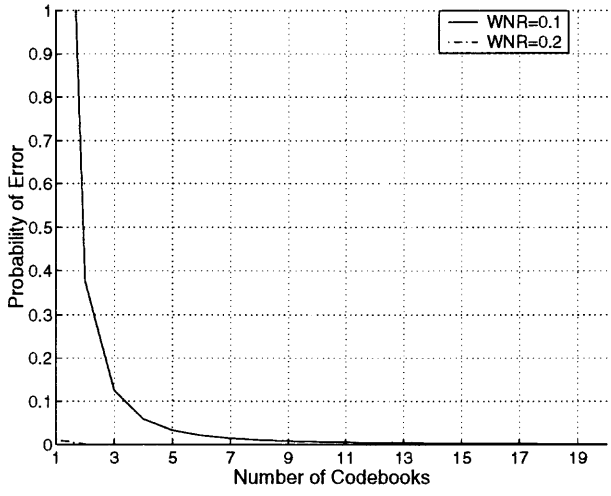


Figure 4.19 Probability of error performance for multiple codebook hiding based on maximum correlation criterion and distortion compensation type of processing for M=1000 and N=500.

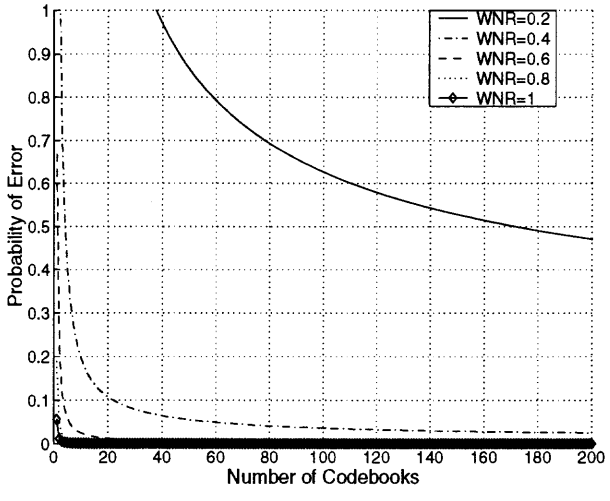


Figure 4.20 Probability of error performance for multiple codebook hiding based on minimum distance criterion and distortion compensation type of processing for M=100 and N=50.

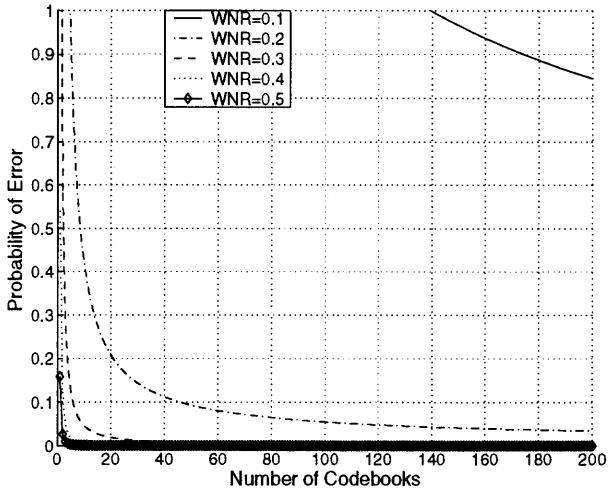


Figure 4.21 Probability of error performance for multiple codebook hiding based on minimum distance criterion and distortion compensation type of processing for M=200 and N=100.

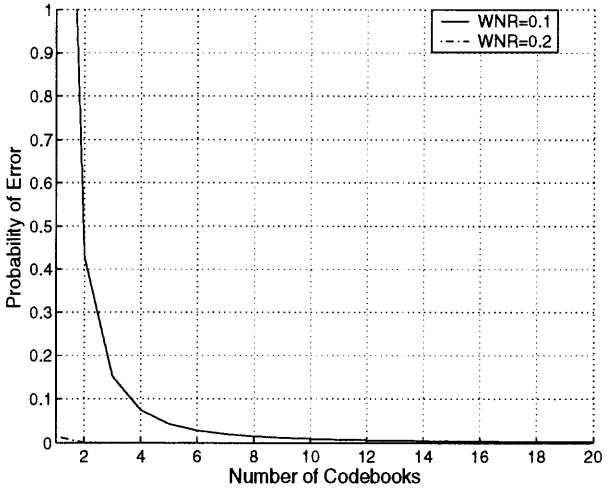


Figure 4.22 Probability of error performance for multiple codebook hiding based on minimum distance criterion and distortion compensation type of processing for M=1000 and N=500.

4.2.7 Implementation and Simulation Results

Optimum codeword selection in multiple codebook hiding depends on designing the set of transform bases $\mathbb{T}_1, \ldots, \mathbb{T}_L$ properly, (*i.e.* they should be able to generate maximally separated transformations of the host signal) Equation (4.13). One intuitive way of picking such a set of transform bases is by choosing them among rotation matrices so that each transformation of the host signal is a rotated version of the others. Multiple codebook hiding method is implemented by designing the transformation bases using Givens rotations [53]. Givens rotations provide orthogonal transformations in $\Re^N$ that can be employed to rotate a given vector with a chosen angle.

A particular transform basis $\mathbb{T}_k$ is obtained by the consecutive multiplication of $\frac{N(N-1)}{2}$ number of orthogonal matrices all with determinant one so that the resulting $\mathbb{T}_k$ is unitary. Each orthogonal matrix is derived from the identity matrix by introducing $\cos\theta_k$ terms at (i,i) and (j,j) locations along with $\sin\theta_k$ and $-\sin\theta_k$ terms at (i,j) and (j,i) locations in order to rotate (i,j) coordinate plane with the designated angle θ_k. The rotation angles θ_k, $k = 1, \ldots, L$ are chosen by uniformly sampling 2π, $\theta_k = (k-1)\frac{2\pi}{L}$.

By setting the signal size to N and number of messages to M, the size of the codebooks utilized by the embedder is fixed to $M \times N$. The watermark signals that are embedded into the host signal are generated using Hadamard transform matrix due to its simplicity. The Hadamard transform matrix of size $N \times N$ and its negated version are combined into a $2N \times N$ binary valued matrix. Every row of the combined matrix is indexed from 1 to $M = 2N$, scaled by $\frac{\Delta}{4}$ for maximum separation, and assigned to the watermark signal $\mathbf{W}_j$, $1 \leq j \leq M$, such that $E[\mathbf{W}_i^T \mathbf{W}_j] = 0$, $i \neq j$ and $i \neq j+N$. The host signal and channel noise are *iid* zero mean Gaussian vectors with $\sigma_C^2 \gg P_E$, σ_Z^2. Prior to embedding, the permitted embedding distortion P_E is fixed, and the optimal values for the embedding parameter Δ are derived for the considered WNRs.

The Δ values are also revealed to the detector. The parameters β and α, however, are properly adjusted for each embedding in order to ensure an embedding distortion of P_E and is not known to the detector. The simulations are done for different number of transformations L and signal sizes N by embedding and detecting randomly chosen message indices.

Multiple codebook hiding is implemented on the type-III scheme based on thresholding and distortion compensation types of post-processing using both maximum correlation and minimum distance criteria. Message embedding and detection with up to 25 codebooks is performed considering codebook sizes of 64×32, 128×64, 256×128 and the WNR range of 0.1 to 1. Figures 4.23 and 4.24 display the probability of success results obtained respectively for $L = 1, 3$ and $L = 1, 4$ with varying N values where the post-processing is thresholding. The increase in the embedding signal size N, at a fixed number of codebooks, improves the detection statistics since normalized correlation and mean squared distance give more reliable results with the larger signal sizes. On the other hand, Figures 4.25 and 4.26 display the performances for thresholding type of processing when $N = 128$ and $L = 1, 3, 5, 9, 14, 25$ using the two criteria. Corresponding results for distortion compensation type of processing are displayed in Figures 4.27 and 4.28 for both criteria. It is observed from these performance simulations that multiple codebook hiding method has superior performance than the corresponding single codebook method at the same N.

The computational complexity of the proposed method depends on the number of codebooks employed. Multiple codebook embedding when compared with single codebook embedding requires the embedding of the watermark signal into transformations of the host signal and a comparison based on the resulting signals, in order to select the transformation basis. On the other hand, at the detector, extraction should be repeated for each transformation basis. Therefore, the

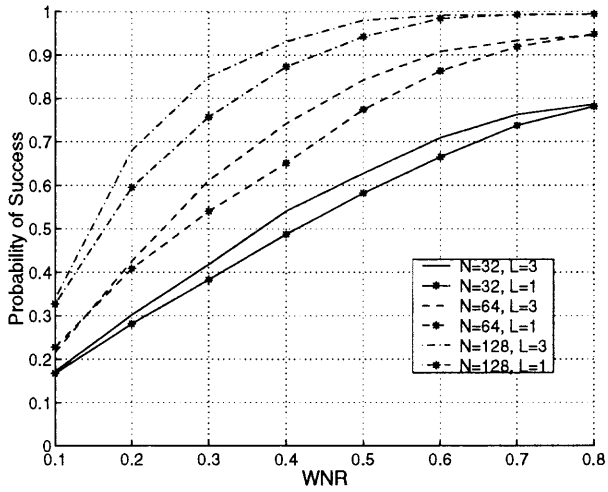


Figure 4.23 Probability of success performance for 3-codebook hiding based on thresholding processing and maximum correlation criterion for various watermark signal sizes of $N = 32$, $N = 64$ and $N = 128$.

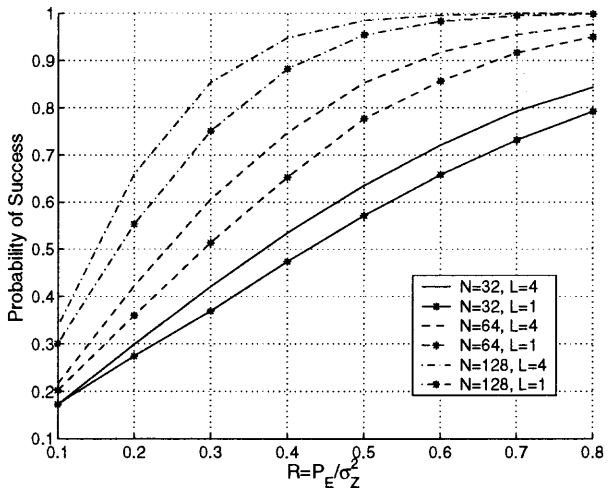


Figure 4.24 Probability of success performance for 4-codebook hiding based on thresholding processing and minimum distance criterion for various watermark signal sizes of $N = 32$, $N = 64$ and $N = 128$.

computational complexity increases almost linearly with the number of codebooks, Figure 4.10.

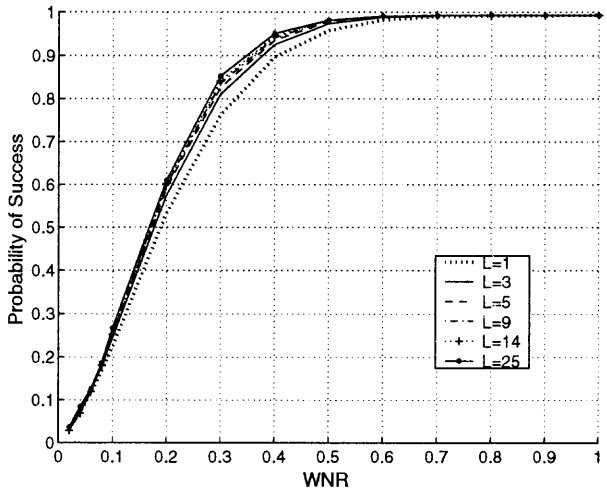


Figure 4.25 Probability of success performance for multiple codebook hiding based on thresholding type of processing and maximum correlation criterion for $L = 1, 3, 5, 9, 14, 25$ and $N = 128$.

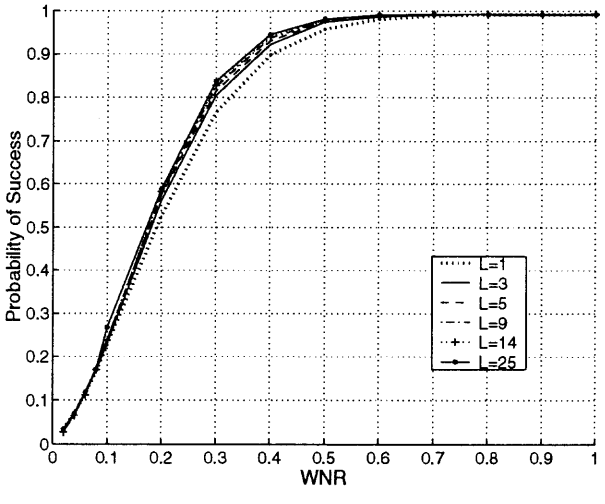


Figure 4.26 Probability of success performance for multiple codebook hiding based on thresholding type of processing and minimum distance criterion for $L = 1, 3, 5, 9, 14, 25$ and $N = 128$.

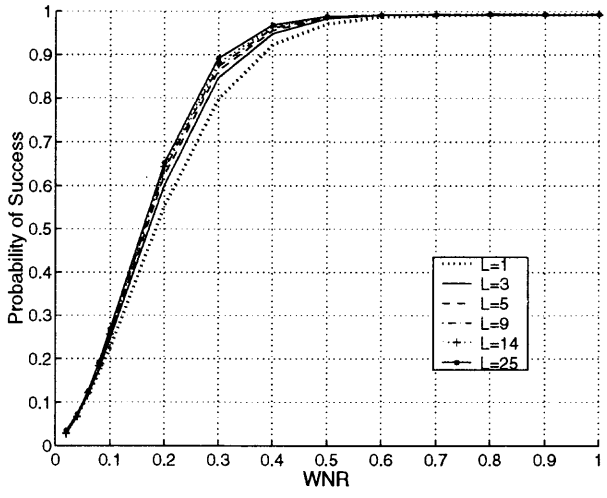


Figure 4.27 Probability of success performance for multiple codebook hiding based on distortion compensation type of processing and maximum correlation criterion for $L = 1, 3, 5, 9, 14, 25$ and $N = 128$.

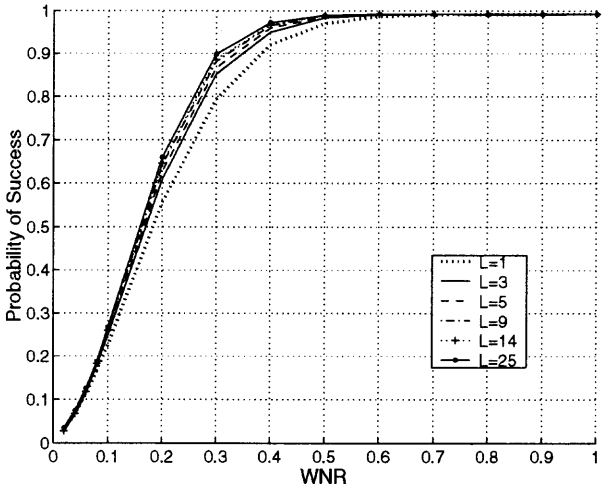


Figure 4.28 Probability of success performance for multiple codebook hiding based on distortion compensation type of processing using minimum distance criterion for $L = 1, 3, 5, 9, 14, 25$ and $N = 128$.

CHAPTER 5

WATERMARKING AGAINST NON-INVTERTIBLE ATTACKS

Two watermarking methods, that are based on type-III hiding methodology and are intended to resist non-invertible attacks are described. These attacks are the cropping & resizing, Section 5.1, and the lossy compression, Section 5.2.

5.1 Synchronization

In some data hiding applications like image, video and audio watermarking, preserving the synchronization between the embedding and detection operations becomes crucial. In such contexts, synchronization refers to accuracy of detector's information on spatial and temporal coordinates of the watermark signal in the stego signal. When the actual coordinates of the embedded watermark signal are different from the ones supposed by the extractor, detection performance may degrade significantly even though the traces of the watermark signal may be present in the stego signal. Therefore, removing the synchronization between embedder and detector becomes a more effective attack than say attempting to "erase" the watermark signal from the stego signal. Geometrical transformations like rotation, scaling, translation, warping and signal cropping are most common forms of desynchronization attacks [54, 55, 56]. For successful extraction of the watermark signal, data hiding methods require tools and techniques for restoring the synchronization efficiently, *e.g.* [57, 58, 59].

In the following sections, a hiding technique based on type-III methodology with thresholding type of post-processing is proposed for watermark recovery from stego signals consecutively subjected to cropping and resizing operations. These attacks pose a threat of poor watermark detection due to signal transformation and signal

loss. Hence, the detector has to be synchronized with the distorted stego signal prior to watermark extraction.

In general, if a particular desynchronizing attack can be modeled as a transformation, watermark detection could either depend on embedding in a domain which is invariant to that transform, or on the ability to estimate the applied transformation by the attacker, and invert it before detection. One particular technique which enables estimation of such transformation in the face of many different types of desynchronization attacks is by periodic embedding, and estimation of the transformation through cyclic autocorrelation.

It is shown that cyclic autocorrelation peak pattern (periodicity features of the signal) can specifically be used for calculating the resampling factor and estimating the amount of cropped data (*i.e.* number of deleted samples in a vector, number of pixels of line in an image). Therefore, the resampled signal can be restored to its original size.

The information loss due to cropping is countervailed by multiple embedding and redundancy coding of the watermark signal. Although, multiple embedding is not an ultimate remedy to cropping, the motivation is that all replicas can not be completely distorted simultaneously due to the perceptual constraints. Figure 5.1 is a representation of signal cropping and resampling. Erasures in the stego signal require reinstatement of synchronization. Synchronization is achieved by designing watermark signals in form of all-pass filters which are orthogonal to all their cyclic shifts, Section 5.1.2. The phase of the all-pass filter is modulated by the message to be conveyed. Reed-Solomon error correcting codes are used for both introducing redundancy and achieving synchronization.

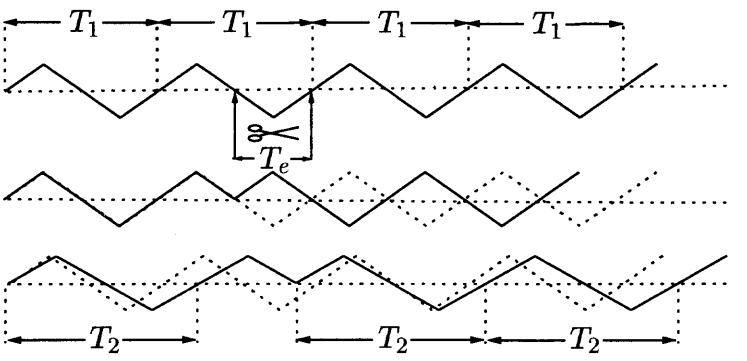


Figure 5.1 Representation of cropping and resampling consecutively.

5.1.1 Autocorrelation for Restoring the Cropped Signal

Let a periodic signal $\mathbf{V}$ be obtained by combining n replicas of the signal $\mathbf{W}$ of length T_1, Figure 5.1. $\mathbf{V}$ is arbitrarily cropped out, $\mathbf{V}_C$, and the resulting signal is resampled by the factor $\frac{1}{\tau} = \frac{T_2}{T_1}$, $\mathbf{V}_{CR}$. Then, T_2 is the size of the resampled $\mathbf{W}$. Let n be a large integer number, T_e be the amount of signal (number of coefficients) cropped from $\mathbf{V}$ where $T_e < T_1$, and $L = nT_2 - \frac{T_e}{\tau}$ be the length of $\mathbf{V}_{CR}$. The resampling factor can also be defined as $\frac{1}{\tau} = \frac{L}{nT_1 - T_e}$. The autocorrelation $R_{V_{CR}V_{CR}}(m)$ of $\mathbf{V}_{CR}$ is computed as

$$R_{V_{CR}V_{CR}}(m) = \sum_{k=1}^{L-|m|} V_{CR}()V_{CR}(k+m). \tag{5.1}$$

In order to recover $\mathbf{W}$, the cropped resampled signal $\mathbf{V}_{CR}$ of size $nT_2 - \frac{T_e}{\tau}$ has to be restored to cropped signal $\mathbf{V}_C$ with size $nT_1 - T_e$ by resampling with the factor τ. The autocorrelation function of $\mathbf{V}_{CR}$ is used to estimate $\frac{1}{\tau}$ depending on information about $\mathbf{V}$ available to extractor (*i.e.* size of $\mathbf{V}$, size of $\mathbf{W}$). It will also be seen that autocorrelation peak pattern provides insights into the nature of the croppings even when croppings occurs at multiple positions (note that if two or more consecutive samples in $\mathbf{V}$ are cropped, it will be considered a single cropping). The total amount of cropped signal is assumed to be much smaller than the size of $\mathbf{V}$, $T_e << nT_1$. The justification for this assumption is that in a typical attack scenario, due to perceptual

constraints, the attacker can not make radical changes on signal size $\mathbf{V}$. Therefore, all copies of $\mathbf{W}$ can not be cropped fatally at the same time. Consequently, in the corresponding autocorrelation function of $\mathbf{V}_{CR}$ the peaks observed at T_2 shifts of the origin, $R_{V_{CR}V_{CR}}(\pm iT_2)$ where $i \in \mathcal{Z}$, will be relatively greater in strength compared to other peaks irrespective of the number of croppings. Given T_1 is known at the extractor, resampling factor can be found by measuring T_2 through distances between the dominant peaks in the autocorrelation function and calculating $\frac{T_2}{T_1}$. Alternately, if the size of $\mathbf{V}$ prior to cropping, nT_1, is known rather than the size of $\mathbf{W}$, $\frac{1}{\tau}$ can be calculated using the relative peak locations of the autocorrelation function.

Considering the single cropping case of amount T_e, the autocorrelation function of the signal $\mathbf{V}_{CR}$ will indicate the presence of two periodic components with the same period, $T_2 = T_1\frac{1}{\tau}$. First component is identified by peaks at T_2 shifts of the origin. The second, on the other hand, generates peaks at the shift of $T_2 - T_e\frac{1}{\tau}$ with respect to zero-shift and at T_2 shifts thereafter. In other words, the first component is due to resampled copies of signal $\mathbf{W}$ in $\mathbf{V}_{CR}$ and second one is due to the cropping. In the autocorrelation, at every $T_2 - T_e\frac{1}{\tau}$ shift following a T_2 shift the incomplete signal period coincides with a copy of itself and generates a peak. The peaks corresponding to latter component are weaker in signal strength compared to the former due to the incomplete $\mathbf{W}$. Therefore, other than the peak at the zero shift, every peak at T_2 shifts (with respect to zero shift) is accompanied by a peak due to cropped $\mathbf{W}$ (assuming n is large enough). The distance d between the peak at kT_2, $k \leq n$, and $(k-1)T_2 + T_2 - T_e\frac{1}{\tau}$ is calculated as

$$\begin{aligned} d &= kT_2 - \left((k-1)T_2 + T_2 - T_e\frac{1}{\tau}\right), \\ &= T_e\frac{1}{\tau}. \end{aligned} \tag{5.2}$$

Being able to measure $\frac{T_e}{\tau}$ and T_2, the resampling factor τ is calculated as $\tau = \frac{nT_2}{nT_1}$ or $\tau = \frac{T_2}{T_1}$ based on availability of nT_1 or T_1. Then the total cropping amount T_e is

calculated using Equation (5.2). It should also be noted that given either of nT_1 or T_1, one can determine either using τ and T_2.

This approach can be expanded to double cropping case where T_{e1} and T_{e2} are the amounts of the non-overlapping cropped samples (T_{e1} and T_{e2} refer to croppings of $\mathbf{W}$ at different locations) from $\mathbf{V}$ with $T_{e1} + T_{e2} < T_1$. Autocorrelation function of $\mathbf{V}_{CR}$, for two-cropping case, may have up to four peaks in every T_2 interval that are $(k-1)T_2$, $k \leq n$, away from zero shift. These peaks may appear at $kT_2 - \frac{T_{e1}+T_{e2}}{\tau}$, $kT_2 - \frac{T_{e1}}{\tau}$, $kT_2 - \frac{T_{e2}}{\tau}$ and kT_2. The last one is due to resampled copies of $\mathbf{W}$ and has highest correlation value. Others are due to cropped-resampled copies of $\mathbf{W}$ and have smaller strengths. If no croppings are present in the first and last periods of $\mathbf{W}$, for relatively large n and T_1, the distance, d, between the first and last peak in any T_2 interval is measured as $\frac{T_{e2}+T_{e1}}{\tau}$. Similar to the single cropping case, nT_2 and $\frac{1}{\tau} = \frac{nT_2}{nT_1}$ are consequently computed.

For more number of croppings followed by resampling, similar analogy is applicable. If T_{e1}, ..., T_{em} are the amounts of the non-overlapping cropped signals and $T_{e1} + \cdots + T_{em} < T_1$, there may at most be $2m$ peaks at every shift based on how the signal $\mathbf{V}$ is cropped (*i.e.* the number of croppings in each period of $\mathbf{W}$, the location of a cropping in the period $\mathbf{W}$, neighborhood of the cropped periods). These croppings may yield correlation peaks at 2^m locations in a T_2 shift (assuming each cropping is non-overlapping with the others and considering first and last periods are not cropped). Corresponding peak locations in the autocorrelation function are at $kT_2 - \sum_{j=1}^{j=m} \frac{T_{ej}}{\tau}$, $kT_2 - \sum_{j=1,j\neq i}^{j=m} \frac{T_{ej}}{\tau}$ for $\forall i$, $kT_2 - \sum_{j=1,j\neq i,l}^{j=m} \frac{T_{ej}}{\tau}$ for $\forall i, l$ such that $i \neq l$, ..., $kT_2 - \frac{T_{ej}}{\tau}$ for $\forall j$, and at kT_2. Then, the distance d between the first and last peaks in a T_2 shift can be used to estimate the total erasure amount.

When the first and last periods of the signal $\mathbf{V}$ are cropped, autocorrelation function may not generate a peak at $kT_2 - \frac{T_{e1}+\cdots+T_{em}}{\tau}$. Therefore, the distance d, measured between the first and last peak at a T_2 shift of the autocorrelation function,

does not indicate $\frac{T_e}{\tau}$. However, as will be explained in Section 5.1.3, d may still be measured using cyclic autocorrelation features for such croppings. Further, if both T_1 and nT_1 are known at the extractor, the amount of cropping, T_e, can also be determined by measuring d and $\frac{1}{\tau}$ using Equation (5.2).

5.1.2 Practical Concerns

Calculating the resampling factor $\frac{1}{\tau}$ correctly depends on identifying correlation peaks and determining their relative locations in the autocorrelation function. However, some peaks may be buried in the correlation noise which makes peak detection unreliable. Designing white noise like **W** signals and using cyclic autocorrelation are two remedies available for measuring d reliably.

Watermark Signal Design. The design of the signal **W** is critical as autocorrelation properties of **W** characterize those of **V**. Designing **W** as an all-pass filter which is orthogonal to all its cyclic shifts, [60], gives one freedom to hide information by modulating the phase of the **W** as well as the improved autocorrelation properties, Section 5.1.2. An all-pass filter **W** of size T_1 gives $\frac{T_1-1}{2}$ degrees of freedom in modulating its phase, if T_1 is odd ($\frac{T_1-2}{2}$ degrees of freedom, if T_1 is even).

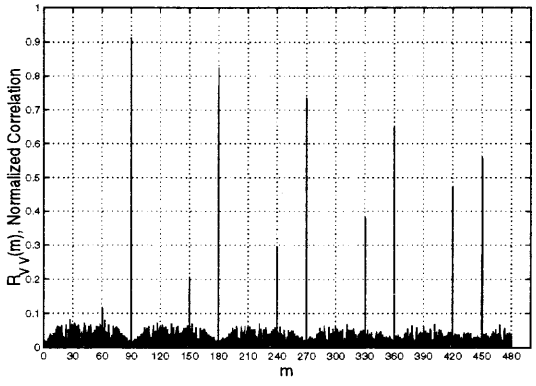
Cyclic autocorrelation. Cyclic autocorrelation enhances the correlation peaks due to signal wrapping in the autocorrelation function. Assuming $\mathbf{V}_{CR}$ has undergone multiple croppings of $T_{e1},\ldots,T_{em}$, the corresponding cyclic autocorrelation can be obtained from the autocorrelation function by flipping the signal range $(\frac{nT_2}{2}-\sum_{j=1}^{j=m}\frac{T_{ej}}{2\tau}, nT_2-\sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}]$ and adding it onto signal range $(0, \frac{nT_2}{2}-\sum_{j=1}^{j=m}\frac{T_{ej}}{2\tau}]$. After signal wrapping, the new coordinates for autocorrelation peaks in the range $(\frac{nT_2}{2}-\sum_{j=1}^{j=m}\frac{T_{ej}}{2\tau}, nT_2-\sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}]$ are found by subtracting their coordinates from $nT_2-\sum_{j=1}^{j=m}\frac{T_{ej}}{\tau}$ which always coincide with one of the 2^m peak locations. For instance, if $\mathbf{V}_{CR}$ has been cropped once by removing T_e samples, autocorrelation peaks at kT_2

and $kT_2 - \frac{T_e}{\tau}$ for $k > \frac{n}{2}$ translate to $(n-k)T_2 - \frac{T_e}{\tau}$ and $(n-k)T_2$ in the cyclic autocorrelation function. For the general case, the peaks at $kT_2 - \sum_{j=1}^{j=m} \frac{T_{ej}}{\tau}$, $kT_2 - \frac{T_{ei}}{\tau}$ for $i \leq m$ and kT_2 respectively translate to, $(n-k)T_2$, $(n-k)T_2 - \sum_{j=1, j\neq i}^{j=m} \frac{T_{ej}}{\tau}$ and $(n-k)T_2 - \sum_{j=1}^{j=m} \frac{T_{ej}}{\tau}$, making peak detection easier. Correspondingly, the autocorrelation peaks with highest strength (due to cropped and resampled $\mathbf{W}$) will be translated to $(n-k)T_2$ and $(n-k)T_2 - \sum_{i=1}^{m} \frac{T_{ei}}{\tau}$ irrespective of the cropping pattern. Then, the resampling factor $\tau = \frac{T_1}{T_2}$ (or $\frac{nT_1}{nT_2}$) is reliably calculated by measuring the distance d between the two peaks, $\frac{T_e}{\tau} = \frac{T_{e1}+\cdots+T_{em}}{\tau}$.

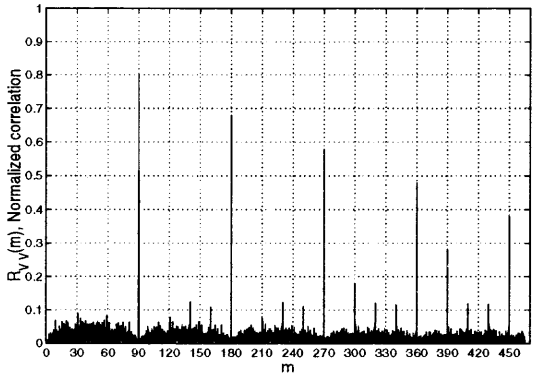
Figures 5.2 a-b display the cyclic autocorrelation functions, $R_{V_C V_C}(m)$, for single and double cropping cases. Signal $\mathbf{W}$ has a size of 90 and $\mathbf{V}$ is generated from 11 replicas of $\mathbf{W}$. In Figure 5.2-a, $\mathbf{V}_C$ is generated by cropping $\mathbf{V}$ once by removing first 30 samples of sixth period. On the other hand in Figure 5.2-b $\mathbf{V}$, is cropped twice by removing middle 40 samples of third period and last 20 samples of fifth period. In both figures, the peaks at multiple shifts of 90 (the size of $\mathbf{W}$) are easily identified, $\tau = 1$. Every shift of size 90, corresponding to size of $\mathbf{W}$, contains two peaks in 5.2-a and four peaks in 5.2-b. The distance $d = T_e$, the number of erased samples, between the peaks in the former is 30 and between the first and fourth in the latter is 60.

5.1.3 Synchronization

Restored cropped signal must be repartitioned to recover $\mathbf{W}$. Since it is not certain which partitions are affected from cropping, extractor needs some markers for re-establishing the synchronization. Most of the partitions contain signal $\mathbf{W}$ or a translated version of it. While some other partitions have cropped and translated versions of $\mathbf{W}$. Reed-Solomon error correcting codes for generating $\mathbf{W}$ and handling synchronization. Since, it is highly likely that most partitions will carry a cyclic shifted version of $\mathbf{W}$, errorless decoding will be possible when the partition is reordered. Thus, given enough redundancy, both robustness against signal loss and



(a)



(b)

Figure 5.2 Computing total cropped amounts using cyclic autocorrelation $R_{V_C V_C}$. **(a)** Cropping once, $T_e = 20$. **(b)** Multiple cropping, $T_{e1} = 40$ and $T_{e2} = 20$.

synchronization are achieved and errorless decoding of most of the partitions is possible at some cyclic shift of the partition.

5.1.4 Results

The methodology is implemented on 512×512 graylevel Lena image, Figure 5.3-a. Message m is assumed to be a sequence of 32 bits. The signal $\mathbf{W}$ takes the form of the watermark signal corresponding to m with a constraint on the correlation properties. Hadamard transform matrix is designated as the codebook and its orthogonal rows are mapped to codewords that are employed in watermark signal generation.

The message bit sequence is translated into words. Then, the message words are redundancy coded using Reed-Solomon error correcting codes. Using the codebook, encoded message words are BPSK modulated and ordered in a way that fulfills the frequency domain symmetry requirements for the phase of the all-pass filter in order to generate the watermark signal $\mathbf{W}$. Watermark signal is chosen to be 32×32 all-pass filter which provides the hider with $\frac{32 \times 32 - 4}{2} = 510$ phase samples to modulate by the coded message m. Then, 16 copies of the watermark signal is embedded throughout the whole image.

The watermarked image is cropped, and in order to compensate the reduction in size, it is resampled back to its original size. At the extractor a copy of the watermarked, cropped, and resampled image is divided into partitions of size $\mathbf{W}$. Watermark detection for each partition is followed by the two dimensional cyclic autocorrelation of the detected set of signals. Using correlation peak pattern resampling factor τ is estimated. Extractor, knowing an estimate of the total cropped amount but not their locations, resamples the image back to its size after cropping. Hence, the disturbing effects of the resampling can be reversed or at least minimized. This image is then re-partitioned and watermark extracted. Since extracted watermark signals may have been cropped and translated, an immediate detection of message m is not possible. Reed-Solomon codes are used to detect message m from the extracted watermark signal since they are capable of correcting burst error. Two-dimensional signal is shifted in rows and columns until an errorless decoding is possible. High redundancy coding helps detecting message m even under severe signal loss.

Figure 5.3 a-d display the results for the described method applied on Lena image, Figure 5.3-a. Watermarked Lena image is displayed in Figure 5.3-b where mean squared error per coefficient due to embedding is 6.9 (40 dB in PSNR). Figure 5.3-c is the watermarked image cropped twice in both dimensions to a size of 488×488.

Each cropping is the erasure of 12 lines of pixels in either horizontal or vertical dimension. Cropped image is resampled back to its original size of 512×512 in 5.3-d. Figures 5.3 e-f are the projections of the cyclic autocorrelation function onto horizontal and vertical dimensions. Distance between the first and last peaks in each period, corresponding to the size of the watermark signal enlarged by the resampling factor, of the cyclic autocorrelation function is $d = 25$ which has an estimation error of 1 line of pixels in both dimensions. T_2 is also measured using the Figures 5.3 e-f as 33 at some shifts and as 34 at most of the others, $\frac{32}{34} < \tau < \frac{32}{33}$. Image in Figure 5.3-d is resampled to a size shorter by 24 ($T_e = round(25 \times \frac{32}{34})$), lines of pixels in each dimension, partitioned in 32×32 blocks and watermark detected. Extracted signals from each block are averaged. Then, the averaged signal block is decoded in cyclic shifts of rows and columns until an errorless decoding is possible. For the presented implementation the redundancy rate is around $\frac{1}{15}$ ($\frac{32}{510}$). Reed-Solomon codes were successful in detecting the 32 bit message m with no errors.

5.2 Type-III Hiding for Lossy Compression

Data compression is the most common application that any multimedia content will undergo. Therefore, optimal design of a watermarking method for the given compression is a very practical requirement. Given the quantization tables utilized by the compression scheme, one will know the exact compression noise that a stego signal will undergo. Hence, compression may be considered as an attack where embedder has the ability to reduce its distorting effects on the stego signal.

As discussed in Chapter 3, the major advantage of quantization based methods over additive schemes is that the former enables hider to optimize the hiding rate at the given attack level unlike the latter. Due to this property of type-III methods, the embedding and detection parameters can be optimized in a way that takes into account compression distortion.

In this section, a type-III data hiding scheme that makes use of the compression scheme's quantization characteristics is presented, [61]. The method incorporates the embedding quantization with the quantization of compression. Results show that joint embedding and compression has better payload and lower compression bit rates when compared to independent compression and quantization. Hiding performance is evaluated under JPEG compression for thresholding type of processing, however, the proposed methodology is trivially applicable to any lossy compression scheme for all types of post-processing.

5.2.1 Joint Embedding and Compression

The motivation for modifying the embedder with respect to compression characteristics relies on the fact that content creator, as the distributor, has the control over both watermarking and compression. Under this circumstance, an optimal system is the one that handles watermarking and compression jointly rather than considering them independent.

Considering watermarking and compression apart from each other may reduce data hiding rate to remarkably low values or to zero. Among all possible cases, worst one occurs when the quantization step size specified by the compression scheme is much more greater than Δ, the distance between the reconstruction points of the embedding quantizers. This may remove all the watermark and lead to zero hiding rates. Moreover, low hiding rates may not be avoided even in moderate or high bit rate compression levels in such cases.

Embedding can be interpreted as introducing two forms of noise to the host signal, namely the distortion due to embedding quantization and the processing distortion. Quantization involved in compression will round embedded watermark signal values to discrete quanta values. Therefore, the compression distortion, the difference between the watermarked signal and the quantized watermarked

signal, is another source of noise that reduces the hiding rate. However, knowing the quantization characteristics in advance, embedder can adjust its embedding distortion and processing distortion to lessen the effects of compression distortion. This requires embedder to be modified in order to make comparisons between watermarked signal and its quantized version to decide on the proper embedding and detection parameters. Using the *a priori* information on the compression, embedder chooses among the $(\Delta,\ \beta)$ parameter pairs that maximizes the data hiding rate. (Note that, as discussed in Chapter 5, for a permitted amount of embedding distortion information hider has infinitely many choices of embedding-detection parameter pairs.)

The information hiding system is outlined below

$$\begin{aligned} \mathcal{W} &: \quad m \longrightarrow \mathbf{W}, \\ \mathbf{S} &= \mathcal{E}_Q(\mathbf{C}, \mathbf{W}) = \mathbf{S} + \mathbf{X_n}, \\ \mathbf{Y} &= \mathbf{S} + \mathbf{Q} + \mathbf{Z}, \\ \hat{\mathbf{W}} &= \mathcal{D}(\mathbf{Y}), \\ \mathcal{W}^{-1} &: \quad \hat{\mathbf{W}} \longrightarrow \hat{m}, \end{aligned} \tag{5.3}$$

where $\mathbf{W}$ is the watermark signal corresponding to message index m, $\mathbf{C}$ is the transformed cover signal coefficients, $\mathbf{X}_n$ is the type-III codeword, $\mathbf{Q}$ is the quantization noise due to compression and $\mathbf{Z}$ is the channel noise. Since quantization for lossy compression is generally performed in transform domain, embedder $\mathcal{E}_Q$ and the detector $\mathcal{D}$ operate on transform domain coefficients. The distortion introduced to $\mathbf{C}$ due to embedding, compression and channel noise are measured using mean square error distortion measure and are respectively denoted by P_E, P_Q and P_Z. Figure of merit used for evaluating the performance of the modified embedder is the normalized correlation between embedded watermark signal and the extracted signal at varying ratios of distortion introduced by embedding and compression to

channel noise distortion, $\frac{P_E+P_Q}{P_Z}$. Corresponding hiding rates are overestimated at a fixed $\frac{P_E+P_Q}{P_Z}$ through calculating the statistics of the Gaussian noise additive to the watermark signal vector so that the watermark signal vector and the extracted noised signal vector have the same correlation.

Comparing the joint and independent embedding-compression at the same distortion level of $P_Q + P_Z$, the hiding rate in the former will be higher as the mutual information between the $\mathbf{W}$ and $\hat{\mathbf{W}}$ is higher due to interrelated $\mathbf{X}_n$ and $\mathbf{Q}$. What is not so readily obvious is that better compression of the watermarked signal is possible when embedding is coordinated by the compression. As embedder tries to minimize quantization noise by changing the embedded signal value with respect to its reconstruction value at the output of the quantizer, entropy of the quantized watermark signal decreases.

Figure 5.4-a displays the hiding rate vs. robustness performance obtained for synthetically generated data using both joint and independent embedding-compression. The host signal, $\mathbf{C}$, is assumed to be an *iid* Gaussian vector. For compression, a quantization step size of 6Δ is assumed for all coefficients. Figure 5.4-b displays the entropies for the watermarked signal after quantization for the same set of data. Joint embedding and compression has higher payload and provides a better compression of the watermarked signal when compared with independent embedding and compression.

5.2.2 Results for JPEG Compression

The method is implemented on 256×256 sized test image where embedding is followed by JPEG compression scheme [62]. Quality factor concept introduced to compression standard enables provider to compress at various bit rate values by scaling the built in quantization tables. Transformed block coefficients are combined coherently into channels where the first channel (00-channel) corresponds to DC coefficients and the

rest of the 63 channels are for AC coefficients. The watermark signal is embedded into first 9 low frequency channels since the rest of the channels go through a coarser quantization which makes embedding extremely difficult.

Watermark signal embedded into transformed image coefficients is an *iid* uniformly distributed vector of length 1024. This vector is embedded into the preselected low frequency channels by the modified embedder making use of the quantization table for a particular quality factor. The attacker's intrusion is also modeled by *iid* Gaussian noise vector of length 1024. Performance results are obtained for a range of $0.2 \leq \frac{P_E+P_Q}{P_Z} \leq 0.8$.

Figures 5.5 a-b display the improvement in 00-channel's hiding rate with joint embedding and compression where embedding powers for JPEG-10 and JPEG-50 compression are restricted to be same. Similarly, Figures 5.6 a-b display the correctly detected number of bits among the embedded 1024 bits. Entropies of the watermarked images after quantization are displayed in figures 5.7 a-b. Modified embedder contributes less bits per pixel increase to the compression bit rate of the sample image.

Although the modification on the embedder for joint embedding and compression is a simple one, the resulting benefits are twofold. Based on the *a priori* information on the compression, it becomes possible to achieve higher embedding rates by embedding with appropriate Δ and β values. Additionally, as embedder aims to minimize quantization noise, resultant embedded signal is more friendly to the quantization.

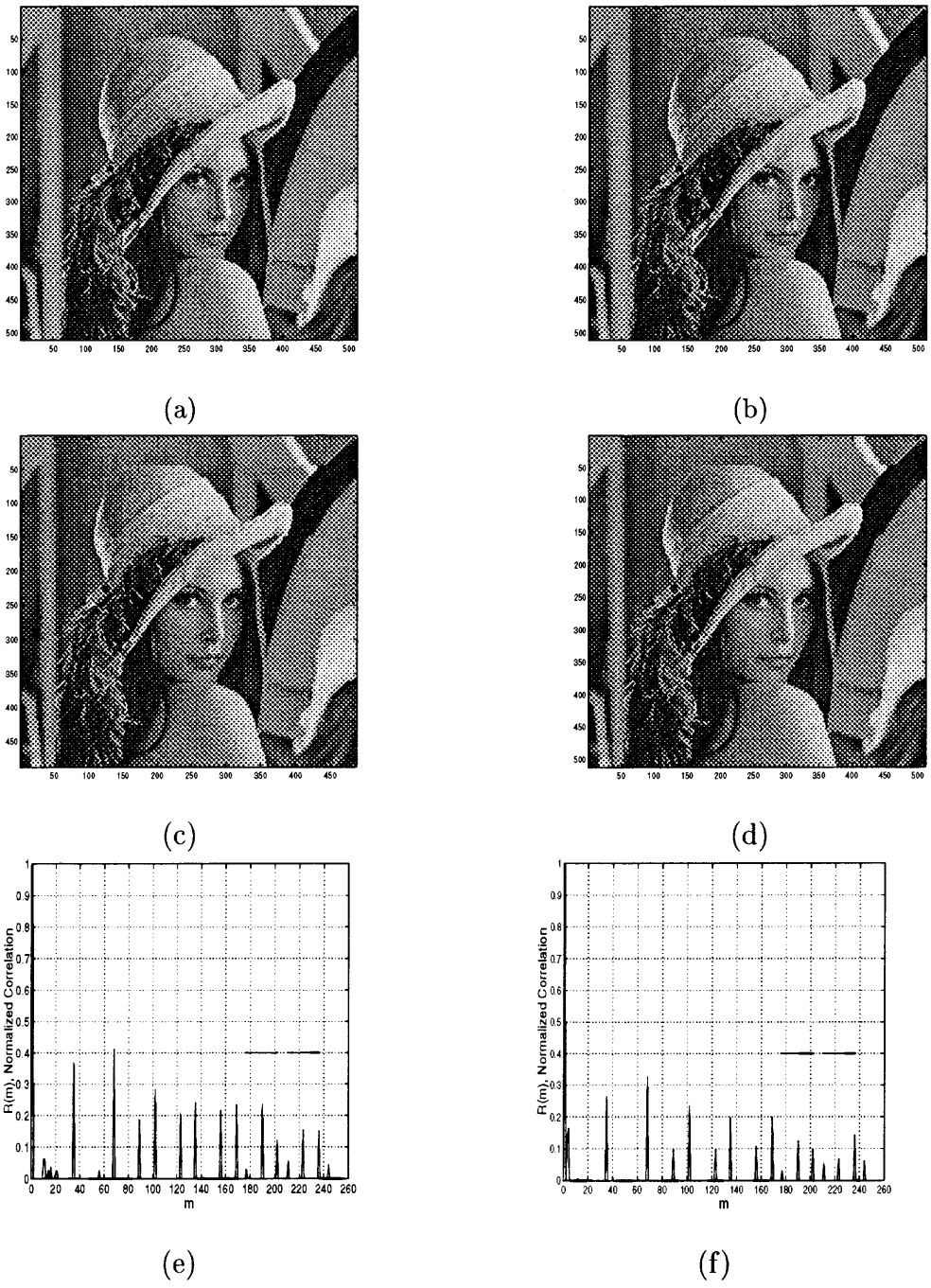
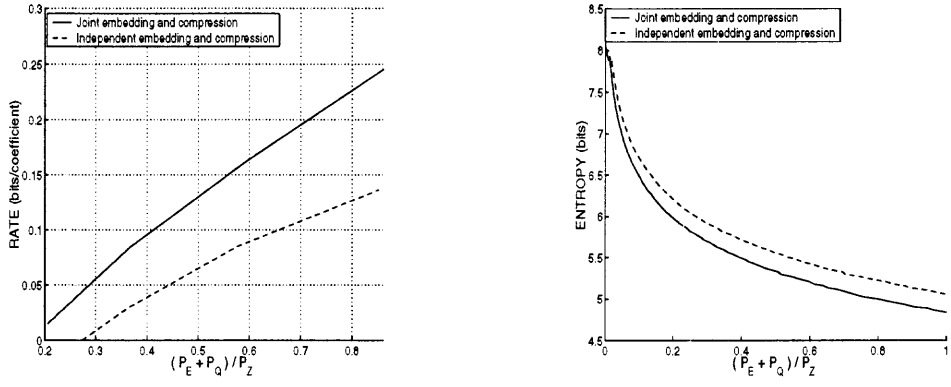
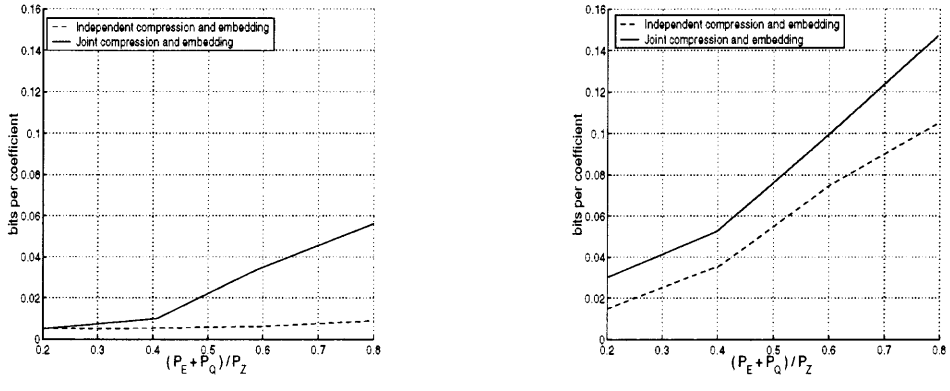


Figure 5.3 **(a)** Lena image. **(b)** Watermarked image. **(c)** Cropped image after watermarking. **(d)** Resampled image after cropping. **(e)** Estimation of cropped amounts from the resampled image **(e)** in horizontal dimension, **(f)** in vertical dimension.



(a) (b)

Figure 5.4 **(a)** Hiding rates for joint and independent embedding-compression. **(b)** Entropy of the quantized embedded signals.



(a) (b)

Figure 5.5 Hiding rates for 00-Channel with compression at quality factors ($40 \leq P_E \leq 170$) **(a)** JPEG-10 and **(b)** JPEG-50.

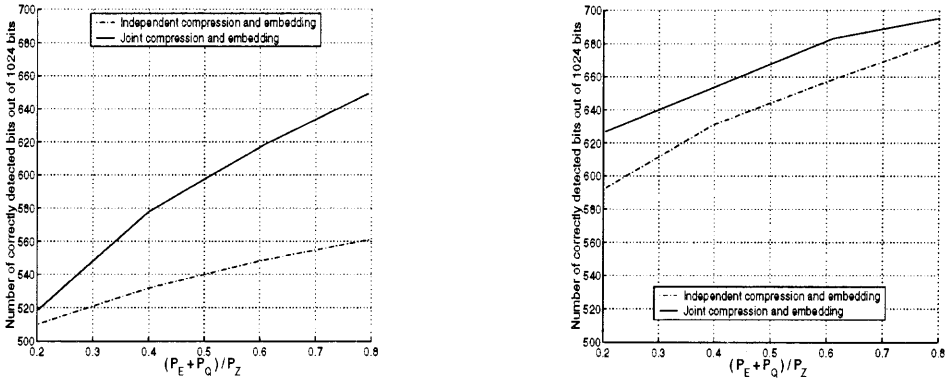


Figure 5.6 Number of correctly detected bits out of 1024 hidden bits for ($40 \leq P_E \leq 170$) **(a)** JPEG-10 and **(b)** JPEG-50.

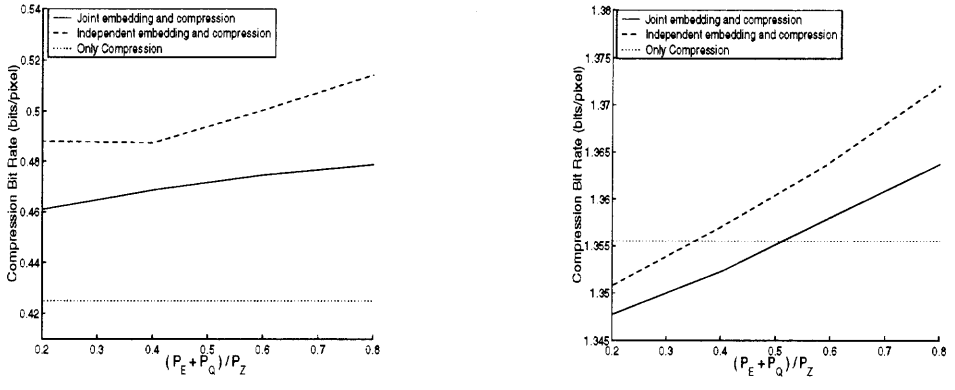


Figure 5.7 Entropy rates after quantization corresponding to **(a)** JPEG-10 and **(b)** JPEG-50.

CHAPTER 6

CONCLUSIONS

6.1 Contributions

This thesis studies oblivious data hiding with the emphasis on quantization based embedding-detection techniques. Contributions of the work can be summarized as follows.

- A communications framework based on channel adaptive encoding and channel independent decoding has been devised with a data hiding perspective.
- The performance evaluation criteria for quantization based embedding-detection techniques are laid out, and a formal treatment of post-processing, employed in practical data hiding methods, is provided.
- Practical embedding-detection techniques are compared in terms of rate, correlation, and probability of error performance merits.
- Multiple codebook data hiding method is introduced as a means of improving rate vs. robustness performance when the embedding signal size is relatively small.
- An oblivious embedding-detection scheme is proposed to cope with cropping and resampling attacks.
- A modification in embedder operation of quantization based methods is proposed to insure robustness against lossy compression by tuning the quantization of embedding with respect to quantization of the compression.

6.2 Remarks

CAE-CID framework: The proposed CAE-CID framework is equivalent to the communications framework introduced by Costa in [17]. However, the encoding of CAE-CID framework assumes the design of $\mathbf{U} = \mathbf{X} + \mathbf{C}$ and imposes the power constraint on the channel input as $\frac{1}{N}||\mathbf{X} - \mathbf{X}_t||^2 \leq P$ rather than $\mathbf{U} = \mathbf{X} + \alpha\mathbf{C}$ and $\frac{1}{N}||\mathbf{X}||^2 \leq P$ of [17], where $\mathbf{X}$, $\mathbf{X}_t$, and α are channel dependent. As a consequence of such codeword generation, decoding does not require channel noise information. When interpreted within the context of data hiding, the CAE-CID framework establishes better analytical model for embedding-detection schemes utilizing a form of post-processing like thresholding, distortion compensation, and Gaussian mapping, as the distortions introduced to host signal due to quantization of embedding and post-processing can be denoted by $\mathbf{X}$ and $\mathbf{X}_t$ in the formulation. Therefore, a better evaluation of the scheme depending on the employed processing is possible.

Performance evaluation: For AWGN attack and mean squared error distortion measure, results indicate that distortion compensation is the optimal embedding processing when X, X_t and C are Gaussian distributed. However, for uniform distribution of X the optimal processing depends on the channel noise level. Performance evaluation of the hiding methods based on probability of error, correlation, and mutual information metrics lead to the same conclusion. At the two extremes, "severe noise" and "no noise" cases, respectively additive schemes and dither modulation (no post-processing) achieve the optimal performance. However, for all other noise levels, the two scheme do not have preferable performances. At relatively high noise levels, techniques with thresholding processing performs best. While distortion compensation performs closely to thresholding, Gaussian mapping is not suited for high noise level applications. For low noise levels, on the other hand, both distortion compensation and Gaussian mapping types of processing yield comparable performances.

Performance Evaluation: Three major design issues for quantization based embedding-detection schemes are laid out and examined. These are the type of post-processing employed at the embedder, the form of detector, and the optimization criteria for embedding-detection parameters.

Multiple Codebook Hiding: It is shown how a practical oblivious information hiding scheme based on type-III embedding methodology with a fixed and *limited embedding signal size* can utilize multiple codebooks to improve its performance. The use of multiple codebooks provide the embedder with a codeword that better adapts to the host signal. The concept is applicable to all type-III data hiding schemes. The proposed method does not require any changes in the embedding and detection processes of a particular data hiding scheme. It merely requires the embedding to be performed multiple times in order to choose the codeword corresponding to a message index. Similarly, multiple extractions of the watermark signal are performed before making a decision on received message. Analytical results indicate that the upper bound on the probability of detection error decreases with the number of codebooks. Simulation results show that the use of multiple codebook hiding is indeed superior to single codebook hiding.

Embedding-detection schemes against some non-invertible attacks: Watermarking methods should have means of reducing the disturbing effects of non-invertible attacks by considering their nature. An oblivious data hiding scheme is proposed to enable watermark recovery from stego images subjected to cropping and resampling consecutively. At the embedder, multiple copies of the redundancy coded watermark signal is embedded in order to cope with the signal loss. Redundancy coding is also utilized to restore the synchronization between embedder and detector. It is shown that cyclic autocorrelation features of the cropped-resampled signal can be used to estimate the nature of the croppings and the cropped amount (in lines of pixels) in both dimensions up to the size of watermark signal.

A modification to the embedder operation is suggested to incorporate the quantization of embedding with the quantization of lossy compression. The embedding-detection parameters are selected to minimize the total distortion due to quantization of both embedding and compression. It is shown that embedder-detector sets making use of compression scheme's quantization characteristics have better payload and lower compression bit rates than independent embedding and compression.

APPENDIX A

CAE-CID FRAMEWORK UNDER VARYING CHANNEL NOISE

The optimal encoding and decoding described in Chapter 2 is achieved by the use of a shared collection of **U** sequences at the given channel noise level σ_Z^2. Consequently, when the channel noise level changes, successful operation can not be maintained due to the dependency on σ_Z^2. However in CAE-CID framework, if encoder is aware of this change, reliable transmission can be restored by adjusting the input power without updating the shared collection of **U** sequences.

Each **U** sequence is an *iid* vector with the Gaussian marginal distribution, $U \sim \mathcal{N}(0, \sigma_X^2 + \sigma_C^2)$. Since both encoder and decoder are bound to use the same sequences (*i.e.* σ_X^2 and σ_C^2 are both fixed) and σ_X and σ_Z are related to each other due to Equation (2.15) as

$$\sigma_X = \frac{P + \sigma_Z^2}{\sqrt{P}}, \tag{A.1}$$

encoder can adjust the input power in accordance with the new noise level $\hat{\sigma}_Z^2$. Using Equation (A.1), the new input power $\hat{P}$ is found as

$$\sqrt{\hat{P}_{1,2}} = \frac{\sigma_X \pm \sqrt{\sigma_X^2 - 4\hat{\sigma}_Z^2}}{2} \tag{A.2}$$

where $\hat{P}_1$ and $\hat{P}_2$ are both valid choices only if $\sigma_X^2 - 4\hat{\sigma}_Z^2 \geq 0$ is satisfied. This requires $\sigma_X \geq 2\hat{\sigma}_Z$ as stated in Section 2.2.1, Equation (2.19).

Consider $\hat{\sigma}_Z^2 = k\sigma_Z^2$, where $0 < k < \infty$, such that $0 < k < 1$ indicates a decrease in the channel noise and $1 < k < \infty$ indicates an increase compared to earlier state σ_Z^2. Since maximum communication rate is computed as $\frac{1}{2}\log_2\left(1 + \frac{P}{\sigma_Z^2}\right)$, Equation (2.14), the new rate will change as a function of $\frac{\hat{P}}{\hat{\sigma}_Z^2}$, or equivalently $\frac{\sqrt{\hat{P}}}{\sqrt{k}\sigma_Z}$. Using

Equation (A.2) the change in $\frac{\sqrt{\hat{P}}}{\sqrt{k}\sigma_Z}$ with respect to $\frac{\sqrt{P}}{\sigma_Z}$ can be expressed as

$$r = \frac{\frac{\sqrt{\hat{P}}}{\sqrt{k}\sigma_Z}}{\frac{\sqrt{P}}{\sigma_Z}} = \frac{\frac{\sigma_X}{\sqrt{k}} \pm \sqrt{\frac{\sigma_X^2}{k} - 4\sigma_Z^2}}{\sqrt{P}}. \tag{A.3}$$

Since, for the given σ_Z^2, $\sqrt{P}$ needs to be a solution of Equation (A.2), $\sqrt{P} = \frac{\sigma_X \pm \sqrt{\sigma_X^2 - 4\sigma_Z^2}}{2}$ will be true for one of the $\pm$. Then, the ratio given in Equation (A.3) can be viewed as

$$r = \frac{\frac{\sqrt{\hat{P}}}{\sqrt{k}\sigma_Z}}{\frac{\sqrt{P}}{\sigma_Z}} = \frac{\frac{\sigma_X}{\sqrt{k}} \pm \sqrt{\frac{\sigma_X^2}{k} - 4\sigma_Z^2}}{\sigma_X \pm \sqrt{\sigma_X^2 - 4\sigma_Z^2}}. \tag{A.4}$$

Depending on the choice of $\hat{P}_1$ or $\hat{P}_2$ and k, the expression given in Equation (A.3) will either be greater or smaller than 1. Therefore, when the channel noise changes from σ_Z^2 to $k\sigma_Z^2$, embedder and detector will be able to resume communication with the same set of $\mathbf{U}$ sequences at a, lower or higher, rate of $\frac{1}{2}\log_2\left(1 + r^2\frac{P}{\sigma_Z^2}\right)$ depending on the choice of input power, as given in Equation (A.2), and k.

APPENDIX B

STATISTICS OF $\rho_{DEP}|P$ AND $D_{DEP}|P$

The mean m_{ρ^*} of the random variable $\rho_{dep}|P$ can be computed by deriving the joint and marginal moments of W and $\hat{W}$. The random variable $\hat{W}$ is expressed in terms of Z_{eff} and W in Equation (3.30), where W is a binary random variable with the density function $f_W(w) = \frac{1}{2}\delta(w - \frac{\Delta}{4}) + \frac{1}{2}\delta(w + \frac{\Delta}{4})$. The pq-th joint moment of W and $\hat{W}$ is defined as

$$E[W^p\hat{W}^q] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w^p\hat{w}^q f_{W,\hat{W}}(w,\hat{w})\mathrm{d}w\mathrm{d}\hat{w}. \tag{B.1}$$

The joint pdf in the above integral can be expressed in terms of marginal and conditional pdfs, $f_{W,\hat{W}}(w,\hat{w}) = f_{\hat{W}}(\hat{w}|w_m)f_W(w)$, thus, Equation (B.1) can be written as

$$\begin{aligned} E[W^p\hat{W}^q] &= \mathrm{P}(w = \frac{\Delta}{4})\int_{-\infty}^{\infty}\left(\frac{\Delta}{4}\right)^p \hat{w}^q f_{\hat{W}}(\hat{w}|w = \frac{\Delta}{4})\mathrm{d}\hat{w} \\ &\quad +\mathrm{P}(w = -\frac{\Delta}{4})\int_{-\infty}^{\infty}\left(-\frac{\Delta}{4}\right)^p \hat{w}^q f_{\hat{W}}(\hat{w}|w = -\frac{\Delta}{4})\mathrm{d}\hat{w}. \end{aligned} \tag{B.2}$$

Since the expectation of a function of a random variable can be expressed in terms of the pdf of the random variable itself rather than of the function [63], $E[\hat{W}] = \int_{-\infty}^{\infty}\hat{w}(z_{eff})f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff}$, and since all pdfs are assumed to be symmetric, Equation (B.2) may be rewritten as

$$\begin{aligned} E[W^p\hat{W}^q] &= \frac{1}{2}\sum_{i=-\infty}^{i=\infty}\int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}}\left(\frac{\Delta}{4}\right)^p\left(\left(\frac{(2i+1)\Delta}{4} - z_{eff}\right)(-1)^i\right)^q \times \\ &\quad f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff} \\ &\quad +\frac{1}{2}\sum_{i=-\infty}^{i=\infty}\int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}}\left(-\frac{\Delta}{4}\right)^p\left(\left(-\frac{(2i+1)\Delta}{4} + z_{eff}\right)(-1)^i\right)^q \times \end{aligned}$$

$$
\begin{aligned}
& f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff}, \\
= & \left(\frac{1}{2}+\frac{(-1)^{p+q}}{2}\right)\left(\frac{\Delta}{4}\right)^{p} 2 \times \\
& \sum_{i=0}^{i=\infty} \int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}} \left(\left(\frac{(2i+1)\Delta}{4}-z_{eff}\right)(-1)^{i}\right)^{q} f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff}, \\
= & \left(\frac{1}{2}+\frac{(-1)^{p+q}}{2}\right)\left(\frac{\Delta}{4}\right)^{p} R(q), \qquad \text{(B.3)}
\end{aligned}
$$

where $R(q)$ is as defined in Equation (3.32). Hence, the joint moment of W and $\hat{W}$ is generalized, based on Equation (B.3), as

$$
E[W^{p}\hat{W}^{q}] = \begin{cases} (\frac{\Delta}{4})^{p} R(q), & \text{if } p,q \text{ are both even or odd,} \\ 0, & \text{otherwise.} \end{cases} \qquad \text{(B.4)}
$$

Marginal moments of W are derived straightforwardly, due to the binary distribution, as

$$
E[W^{p}] = \begin{cases} 0, & \text{if } p \text{ is odd,} \\ (\frac{\Delta}{4})^{p}, & \text{if } p \text{ is even.} \end{cases} \qquad \text{(B.5)}
$$

The moments of the r.v. $\hat{W}$ depend on W and Z_{eff} through Equation (3.30) and can be computed by using the properties employed in deriving Equations (B.2) and (B.3) as

$$
\begin{aligned}
E\left[\hat{W}^{p}\right] &= \mathrm{P}(W=\frac{\Delta}{4})E\left[\hat{W}^{p}|w=\frac{\Delta}{4}\right]+\mathrm{P}(W=-\frac{\Delta}{4})E\left[\hat{W}^{p}|w=-\frac{\Delta}{4}\right], \\
&= \frac{1}{2}\sum_{i=-\infty}^{i=\infty}\int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}}\left(\left(\frac{(2i+1)\Delta}{4}-z_{eff}\right)(-1)^{i}\right)^{p} f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff} \\
&\quad +\frac{1}{2}\sum_{i=-\infty}^{i=\infty}\int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}}\left(\left(-\frac{(2i+1)\Delta}{4}+z_{eff}\right)(-1)^{i}\right)^{p} f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff}, \\
&= \left(\frac{1}{2}+\frac{1}{2}(-1)^{p}\right)2\sum_{i=0}^{i=\infty}\int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}}\left(\left(\frac{(2i+1)\Delta}{4}-z_{eff}\right)(-1)^{i}\right)^{p} \times \\
&\quad f_{Z_{eff}}(z_{eff})\mathrm{d}z_{eff}, \\
&= \left(\frac{1}{2}+\frac{1}{2}(-1)^{p}\right)R(p). \qquad \text{(B.6)}
\end{aligned}
$$

Finally, $E\left[\hat{W}^p\right]$ can be summarized as

$$E[\hat{W}^p] = \begin{cases} 0, & \text{if } p \text{ is odd,} \\ R(p), & \text{if } p \text{ is even.} \end{cases} \tag{B.7}$$

Based on Equations (B.1)-(B.7), m_{ρ^*} is derived as

$$\begin{aligned} m_{\rho^*} &= \frac{E[W\hat{W}]}{\sqrt{E[W^2]E[\hat{W}^2]}}, \\ &= \frac{\frac{\Delta}{4}R(1)}{\sqrt{(\frac{\Delta}{4})^2 R(2)}}, \\ &= \frac{R(1)}{\sqrt{R(2)}}. \end{aligned} \tag{B.8}$$

The variance $\sigma^2_{\rho^*}$ is the variation of the correlation coefficient ρ^* around its mean m_{ρ^*} when m_{ρ^*} is estimated from N *iid* samples of $\hat{\mathbf{W}}_m$ and $\mathbf{W}_m$. For the case when $\hat{W}$ and W are from a bivariate Gaussian distribution, the variance is as given in [64]. However, when the samples are from non-Gaussian distributions, derivation of σ_{ρ^*} is not straightforward. Therefore, Monte-Carlo simulations are performed to obtain the $\sigma^2_{\rho^*}$ values for the considered N by computing the correlations between the embedded $\mathbf{W}_m$ and extracted $\hat{\mathbf{W}}_m$ at the assumed WNR and then measuring the deviation from m_{ρ^*}. However, for minimum distance criterion the corresponding variance values can be calculated in a straightforward manner.

For the minimum distance criterion the statistics of $d_{dep}|P$ are computed in terms of the statistics of the random variable $\lambda = W^2 + \hat{W}^2 - 2W\hat{W}$.

When the noise level is very high so that it can be considered uniformly distributed over all quantization intervals W and $\hat{W}$ become independent of each other, and $\hat{W}$ is extracted as a uniformly distributed signal in $[-\frac{\Delta}{4}, \frac{\Delta}{4}]$. The mean $m_\lambda = E[\lambda]$ and the variance $\sigma^2_\lambda = E[\lambda^2] - m^2_\lambda$ of λ is calculated in terms of the

moments

$$\begin{aligned} E[\lambda] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(w^2+\hat{w}^2-2w\hat{w})f_{W,\hat{W}}(w,\hat{w})\mathrm{d}w\mathrm{d}\hat{w}, \\ &= \int_{-\infty}^{\infty} w^2\, f_W(w)\,\mathrm{d}w + \int_{-\infty}^{\infty}\hat{w}^2\, f_{\hat{W}}(\hat{w})\,\mathrm{d}\hat{w} \\ &\quad -2\int_{-\infty}^{\infty} w\, f_W(w)\,\mathrm{d}w \int_{-\infty}^{\infty}\hat{w}\, f_{\hat{W}}(\hat{w})\,\mathrm{d}\hat{w}, \\ &= Var[W]+Var[\hat{W}], \\ &= \frac{\Delta^2}{16}+\frac{\Delta^2}{48}=\frac{\Delta^2}{12}, \end{aligned} \tag{B.9}$$

$$\begin{aligned} E[\lambda^2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(w^2+\hat{w}^2-2w\hat{w})^2 f_{W,\hat{W}}(w,\hat{w})\mathrm{d}w\mathrm{d}\hat{w}, \\ &= \int_{-\infty}^{\infty} w^4 f_W(w)\mathrm{d}w + \int_{-\infty}^{\infty}\hat{w}^4 f_{\hat{W}}(\hat{w})\mathrm{d}\hat{w} \\ &\quad +6\int_{-\infty}^{\infty} w^2 f_W(w)\mathrm{d}w \int_{-\infty}^{\infty}\hat{w}^2 f_{\hat{W}}(\hat{w})\mathrm{d}\hat{w}, \\ &= \frac{\Delta^4}{2^8}+\frac{\Delta^4}{2^85}+\frac{\Delta^4}{2^7}=\frac{\Delta^4}{80}, \end{aligned} \tag{B.10}$$

as $m_\lambda = \frac{\Delta}{12}$ and

$$\sigma_\lambda^2 = \frac{\Delta^4}{80}-(\frac{\Delta^2}{12})^2=\frac{\Delta^4}{180}. \tag{B.11}$$

When W and $\hat{W}$ are dependent on each other, the statistics of $d_{dep}|P$ can be similarly computed in terms of the individual and joint moments of W and $\hat{W}$, Equations (B.4)-(B.7). Consequently the mean m_{d^*} and the variance $\sigma_{d^*}^2$ are computed as

$$\begin{aligned} m_{d^*} &= E\left[\frac{1}{N}\sum_{l=1}^{l=N}\left(W^2+\hat{W}^2-2W\hat{W}\right)\right], \\ &= E[W^2]+E[\hat{W}^2]-2E[W\hat{W}], \\ &= \left(\frac{\Delta}{4}\right)^2+R(2)-\left(\frac{\Delta}{4}\right)R(1), \end{aligned} \tag{B.12}$$

$$
\begin{aligned}
\sigma_{d^*}^2 &= E\left[\left(\frac{1}{N}\sum_{l=1}^{l=N}\left(W^2+\hat{W}^2-2W\hat{W}\right)\right)^2\right]-m_{d^*}^2, \\
&= \frac{1}{N}\left(E[W^4]+E[\hat{W}^4]+6E[W^2\hat{W}^2]-4E[W^3\hat{W}]-4E[W\hat{W}^3]\right) \\
&\quad +\frac{N-1}{N}\Big(E[W^2]^2+E[\hat{W}^2]^2-4E[W^2]E[W\hat{W}]-4E[\hat{W}^2]E[W\hat{W}] \\
&\quad +2E[W^2]E[\hat{W}^2]+4E[W\hat{W}]^2\Big)-m_{d^*}^2, \\
&= \frac{1}{N}\left(\left(\frac{\Delta}{4}\right)^4+R(4)+6\left(\frac{\Delta}{4}\right)^2R(2)-4\left(\frac{\Delta}{4}\right)^3R(1)-4\left(\frac{\Delta}{4}\right)R(3)\right) \\
&\quad +\frac{N-1}{N}\left(\left(\frac{\Delta}{4}\right)^4+R(2)^2-4\left(\frac{\Delta}{4}\right)^3R(1)-4\frac{\Delta}{4}R(2)R(1)+2\left(\frac{\Delta}{4}\right)^2R(2)\right. \\
&\quad \left.+4\left(\frac{\Delta}{4}\right)^2R(1)^2\right)-m_{d^*}^2. \qquad \text{(B.13)}
\end{aligned}
$$

REFERENCES

[1] Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery, *IEEE Transactions on Signal Processing,* vol. 51, no. 4, 2003.

[2] B. Chen, *Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[3] M. Ramkumar, *Data Hiding in Multimedia - Theory and Applications.* PhD thesis, New Jersey Institute of Technology, Newark, NJ, 2000.

[4] A. Cohen, *Information Theoretic Analysis of Watermarking Systems.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2001.

[5] J. J. Eggers, *Information Embedding and Digital Watermarking as Communication with Side Information.* PhD thesis, Lehrstuhl fur Nachrichtentechnik I, Universitat Erlangen-Nurnberg, Erlangen, Germany, 2001.

[6] J. Chou, *Channel coding with side information: theory, practice and applications.* PhD thesis, University of California, Berkeley, California, 2002.

[7] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing,* vol. 6, no. 12, pp. 1673–1687, 1997.

[8] B. Chen and G. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking," in *Proc SPIE: Security and Watermarking of Multimedia Contents II,* vol. 3971, pp. 48–59, 2000.

[9] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communication with side information," *Proc. of IEEE,* vol. 87, pp. 1127–1141, 1999.

[10] J. Chou, S. S. Pradhan, L. E. Ghaoui, and K. Ramchandran, "On the duality between data hiding and distributed source coding," in *Proc. of 33rd Annual Asilomar conference on Signals, Systems, and Computers,* 1999.

[11] R. J. Barron, B. Chen, and G. W. Wornell, "The duality between information embedding and source coding with side information and its implications - applications," *IEEE Transactions on Information Theory.*

[12] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal,* vol. 35, no. 3-4, pp. 313–336, 1996.

[13] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "A secure, robust, watermark for multimedia," in *Proc. of 1st int. Information Hiding Workshop,* pp. 185–206, 1996.

[14] J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Proc. of 1st int. Information Hiding Workshop*, pp. 207–226, 1996.

[15] F. Hartung and B. Girod, "Digital watermarking of raw compressed video," in *Proc. of European Conference of Advanced Imaging and Network Technologies*, 1996.

[16] J. R. Hernandez, F. Perez-Gonzalez, J. M. Rodriguez, and G. Nieto, "Performance analysis of a 2-d multipulse amplitude modulation scheme for data hiding and watermarking of still images," *IEEE J. Select. Areas Commun.*, vol. 16, no. 4, pp. 510–524, 1998.

[17] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, pp. 439–441, 1983.

[18] B. Chen and G. W. Wornell, "Provably robust digital watermarking," in *Proc SPIE: Multimedia Systems and Applications*, vol. 3845, 1998.

[19] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, pp. 1423–1443, May 2001.

[20] M. Ramkumar and A. N. Akansu, "Self-noise suppression schemes for blind image steganography," in *Proc SPIE International Workshop on Voice, Video and Data Communication, Multimedia Applications*, vol. 3845, Sept. 1999.

[21] J. J. Eggers, J. K. Su, and B. Girod, "A blind watermarking scheme based on structured codebooks," *IEE Colloq. Secure Images and Image Authentication*, vol. 4, pp. 1–6, Apr. 2000.

[22] J. Chou, S. S. Pradhan, L. E. Ghaoui, and K. Ramchandran, "A robust optimization solution to the data hiding problem using distributed source coding principles," in *Proc SPIE: Image and Video Communications and Processing*, vol. 3974, 2000.

[23] B. Chen and G. W. Wornell, "Digital watermarking and information embedding using dither modulation," in *IEEE Second Workshop on Multimedia Signal Processing*, pp. 273–278, 1998.

[24] B. Chen and G. W. Wornell, "Dither modulation: A new approach to digital watermarking and information embedding," in *Proc. of SPIE: Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 342–353, 1999.

[25] K. Tanaka, Y. Nakamura, and K. Matsui, "Embedding secret information into a dithered multi-level image," in *Proc. of IEEE International Conference On Image Processing*, pp. 216–220, 1990.

[26] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Proc. of IEEE International Conference On Image Processing*, vol. 2, pp. 86–90, 1994.

[27] G. Caronni, "Assuring ownership rights for digital images," in *Proc. of Reliable IT Systems*, vol. VIS-95, Vieweg Publishing Company, 1995.

[28] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Data hiding for video-in-demand," in *Proc. of IEEE International Conference On Image Processing*, vol. 2, pp. 676–679, 1997.

[29] H.-J. M. Wang, P.-C. Su, and C.-C. J. Kuo, "Wavelet-based digital image watermarking," *Optics Express*, vol. 3, pp. 491–496, Dec. 1998.

[30] M. Wu and B. Liu, "Watermarking for image authentication," in *Proc. of IEEE International Conference On Image Processing*, vol. 2, pp. 437–441, 1998.

[31] B. Chen and G. Wornell, "Achievable performance of digital watermarking systems," in *Proc. of IEEE Int. Conference on Multimedia Computing and Systems*, vol. 1, pp. 13–18, 1999.

[32] J. J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, "Scalar costa scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, 2003.

[33] F. Perez-Gonzalez, F. Balado, and J. R. Hernandez Martin, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 960–980, 2003.

[34] H. Malvar and D. A. F. Florencio, "Improved spread spectrum: a new modulation for robust watermarking," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 898–905, 2003.

[35] C. E. Shannon, "Channels with side information at the transmitter," *IBM Journal of Research and Development*, vol. 2, pp. 289–293, 1958.

[36] A. V. Kusnetsov and B. S. Tsybakov, "Coding in a memory with defective cells," *Translation from Problemy Peredachi Informatsi*, vol. 10, pp. 52–60, 1974.

[37] S. I. Gelfand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[38] C. Heegard and A. A. El Gamal, "On the capacity of computer memory with defects," *IEEE Transactions on Information Theory*, vol. 29, pp. 731–739, 1983.

[39] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," *IEEE Transactions on Information Theory*, vol. 49, pp. 563–593, Mar. 2003.

[40] A. S. Cohen and A. Lapidoth, "The gaussian watermarking game," *IEEE Transactions on Information Theory*, vol. 48, pp. 1639–1667, June 2002.

[41] T. M. Cover and J. A. Thomas, *Elements of Information Theory, Second Edition.* John-Wiley and Sons Inc, New York, NY, 1991.

[42] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "Efficient codebook structures for practical information hiding systems," in *Proc. of CISS*, Mar. 2001.

[43] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "A new perspective for embedding-detection methods with distortion compensation and thresholding processing techniques," in *Proc. of IEEE-ICIP Conference*, 2003.

[44] M. Ramkumar and A. N. Akansu, "Capacity estimates for data hiding in compressed images," *IEEE Transaction on Image Processing*, vol. 10, no. 8, pp. 1252–1263, 2001.

[45] M. Ramkumar and A.N. Akansu, "Information theoretic bounds for data hiding in compressed images," in *IEEE Second Workshop on Multimedia Signal Processing*, pp. 267–272, 1998.

[46] J. Bloom, M. Miller, and I. Cox, *Digital Watermarking: Principles and Practice.* Morgan Kaufmann, 2001.

[47] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "An overview of scalar quantization based data hiding methods," *submitted to IEEE Transactions on Signal Processing*, 2003.

[48] R. G. Gray and T. M. Stockham, "Dithered quantizers," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 805–812, 1993.

[49] J. J. Eggers, J. K. Su, and B. Girod, "Public key watermarking by eigenvectors of linear transforms," in *European Signal Processing Conference (EUSIPCO 2000)*, Sept. 2000.

[50] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "Multiple codebook information hiding," in *Proc. of CISS*, Mar. 2002.

[51] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "Multiple codebook information hiding based on minimum distortion criterion," in *Proc. of CISS*, Mar. 2003.

[52] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "An embedding-detection technique for data hiding with small host signal sizes," *submitted to IEEE Transactions on Signal Processing*, 2003.

[53] D. S. Watkins, *Fundamentals of Matrix Computations.* John Wiley & Sons, New York, 1991.

[54] F. A. P. Petitcolas, R.J. Anderson, and M.G. Kuhn, "Attacks on copyright marking systems," in *Second Workshop on Information Hiding, Lecture Notes in Computer Science*, vol. 1525, pp. 218–238, 1998.

[55] I. J. Cox and J.P. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 587–593, 1998.

[56] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," in *Proc. of SPIE Security and Watermarking of Multimedia Contents*, vol. 3657, pp. 226–239, 1999.

[57] M. Kutter, "Watermarking resistent to translation, rotation, and scaling," in *Proc SPIE Multimedia Systems and Applications*, vol. 3528, pp. 423–431, Nov. 1998.

[58] E. T. Lin and E. J. Delp, "Temporal synchronization in video watermarking," in *Proc SPIE Security and Watermarking of Multimedia Contents IV*, vol. 4675, Jan. 2002.

[59] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "A robust type-iii data hiding technique against cropping and resizing attacks," in *IEEE International Symposium Circuits and Systems*, vol. 4, pp. 3449–3452, 2002.

[60] M. Ramkumar, G. V. Anand, and A. N. Akansu, "On the implementation of the 2-band cyclic filter banks," in *IEEE International Symposium Circuits and Systems-ISCAS'99*, vol. 3, pp. 520–523, 1999.

[61] H. T. Sencar, M. Ramkumar, and A. N. Akansu, "Improvements on data hiding for lossy compression," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 444–447, 2002.

[62] G. K. Wallace, "The jpeg still picture compression standard," *Communications of the ACM*, vol. 34, pp. 31–44, Apr. 1991.

[63] A. Papoulis, *Probability, Random Variables, and Stochastic Processes, Third Edition.* McGraw Hill Inc., Singapore, 1991.

[64] R. A. Fisher, *Statistical Methods for Research Workers.* Hafner Press, New York, 1970.