

POLITECNICO DI BARI

SCUOLA INTERPOLITECNICA DI DOTTORATO

Research Doctorate Course in Information Engineering
SSD: ING-INF/03

Final Dissertation

**Digital watermarking methods
applied to non-additive channels**



Michele Scagliola

Tutor
prof. P. Guccione

Co-ordinator of the Research Doctorate Course
prof. M. N. Armenise

25 February 2010

Supervisor: Prof. Pietro Guccione

1. Reviewer: Prof. Fernando Pérez-González

2. Reviewer: Prof. Carlos Mosquera

Day of the defense: February 25, 2010

Abstract

In the design of a data hiding method, one of the basic requirements that has to be accounted for is robustness. Thus, a great effort has been spent by the data hiding community to develop algorithms to achieve robustness against the modifications that watermarked contents can undergo during their life cycle. The core of this thesis is devoted to the analysis of the robustness for data hiding methods with side information available at the embedder, i.e., where the embedder takes advantage of deterministically knowing the content to be watermarked. These methods have been extensively studied in the literature in case of additive noise and in case of quite simple non-additive channels, such as gain scaling. In this work we have studied side informed data hiding methods in case of more complex non-additive channels, that model most of the processings commonly applied to media contents. Exploiting the properties of rational dither modulation, which is a gain invariant quantization-based data hiding method, novel methods robust to non-additive channels have been proposed and theoretical analysis has been provided to identify their achievable performance. Our study focused on the power-law channel and on the linear time invariant filtering channel. For the former attack channel, a novel class of data hiding methods theoretically invariant to this channel has been proposed and a deep analysis has been provided for the so called *Hyperbolic RDM*. The filtering on the channel has been faced by discrete Fourier transform - rational dither modulation, but this technique exhibits a severe loss of performance for non-white hosts. In this thesis the analysis of discrete Fourier transform - rational dither modulation has been generalized to non-white hosts and a reasonable explanation of the loss of performance has been provided. Then, departing from this analysis, a novel extension has been developed to fill the gap between the data rates achieved for white and non-white hosts.

The thesis summarizes the research activity developed during the three years Ph.D. course, which was founded by the "Scuola Interpolitecnica di Dottorato", a special project whereby the three Italian Technical Universities, the Polytechnic of Torino, the Polytechnic of Bari and the Polytechnic of Milano, aim to offering a joint PhD program of high qualification.

Acknowledgements

I wish to thank my supervisor for having directed towards the study of exciting and fascinating research fields in signal processing. I am very thankful for the guidance, advices and encouragements he provided since the beginning of my Ph.D.

I am grateful to Prof. Fernando Pérez-González for creating the opportunity of staying as visiting student at the University of Vigo and for the interesting research activities I took part to.

I'd like to thank all the colleagues of the Telematics Lab at the Politecnico di Bari and the colleagues of the Signal Processing in Communications Group at the University of Vigo, for the interesting discussions and, above all, for their friendship.

I would also like to express my gratitude to the reviewers of the thesis, Prof. Fernando Pérez-González and Prof. Carlos Mosquera.

Michele Scagliola

Contents

List of Figures	v
List of Tables	ix
Acronyms and notations	x
1 Introduction	1
1.1 A brief history	1
1.2 Applications	3
1.3 General requirements	6
1.4 Classification of attacks	7
1.5 Summary and thesis objectives	8
2 Watermarking as communications: channel models and embedding strategies	11
2.1 Watermarking as communications	12
2.2 Distortion measures	14
2.3 Spread Spectrum watermarking	15
2.4 Watermarking as communications with side information at the transmitter	18
2.5 Desynchronization attack channel	20
2.5.1 Performance of basic decoders	22
2.5.2 Achieving robustness to desynchronization attacks	23
2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images	25
2.6.1 Image pre-processing	27
2.6.2 Embedding and detection	28

2.6.3	Results	31
2.6.3.1	Effectiveness of direction extraction	31
2.6.3.2	Imperceptibility	32
2.6.3.3	Robustness	35
2.7	Concluding remarks	35
3	Quantization-based embedding and scaling vulnerability	38
3.1	Dither modulation	39
3.2	Gain attack and countermeasures	42
3.2.1	Angular quantization index modulation	46
3.2.2	DM in the logarithmic domain	47
3.3	Rational Dither Modulation	50
3.4	Concluding remarks	55
4	Providing invariance to nonlinear volumetric scaling for quantization based data hiding	56
4.1	The power-law attack	57
4.2	Hyperbolic transformation providing invariance to power-law attack	59
4.2.1	Embedder	60
4.2.2	Decoder	61
4.2.3	Experimental results	62
4.3	Hyperbolic RDM	64
4.3.1	Embedding distortion	65
4.3.2	Analytical derivation of the bit error rate	68
4.3.3	Alternative scheme for memory reduction	72
4.3.4	Experimental results	73
4.4	Concluding remarks	81
5	High-rate data hiding robust to linear filtering	83
5.1	LTI filtering attack	84
5.2	Discrete Fourier transform RDM	86
5.2.1	Analytical derivation of the bit error probability	89
5.2.2	Improvements	93
5.3	Performance analysis for colored Gaussian hosts	97

5.3.1	Experimental results	101
5.4	Whitened DFT-RDM	110
5.4.1	Experimental results for colored Gaussian host	113
5.4.2	Experimental results for audio tracks	116
5.5	Concluding remarks	123
6	Conclusions	125
6.1	Future research lines	127
	References	130

List of Figures

2.1	Watermarking as communication channel.	12
2.2	Gaussian channel for spread spectrum watermarking.	17
2.3	Watermarking as communication channel with side information at the encoder.	18
2.4	DC-QIM embedder.	20
2.5	Model for desynchronization attacks.	21
2.6	Block diagram of the direction extraction method.	26
2.7	Block diagram of the pre-processing.	27
2.8	Embedder block scheme.	29
2.9	An intuitive insight of DM embedding.	30
2.10	Example of pre-processing on Boat image with extraction of invariant direction for both original and rotated cases.	32
2.11	Comparison of a detail of the standard image Boat marked using constant gain and variable gain.	34
3.1	An intuitive insight of DM embedding.	40
3.2	Illustration of volumetric distortions on standard image Man.	43
3.3	Illustration of white Gaussian noise addition on standard image Man.	44
3.4	Fixed gain attack channel.	45
3.5	Empirical values of the error probability of AQIM as a function of σ_n for different values of M	48
3.6	Empirical values of the error probability for differential and non-differential logarithmic DM (DWR=25 dB).	49
3.7	Block diagram of L th-order RDM.	51

LIST OF FIGURES

3.8	Empirical values of error probability of RDM against FGA ($\rho = 1.05$) for different values of L , Gaussian host and DWR=25 dB.	53
3.9	Empirical values of error probability for different gain invariant quantization-based techniques, for Gaussian host and DWR=25 dB.	54
4.1	Power-law attack.	57
4.2	Block diagram of the hyperbolic watermarking system.	61
4.3	Empirical values of the error probability for power-law attack with $\rho = .7$ and $\gamma = 1.2$ (DWR=25 dB).	63
4.4	Block diagram of the hyperbolic RDM embedding/decoding scheme.	64
4.5	Comparison of the empirical values of DWR for different values of L_h with the analytical approximation.	67
4.6	Comparison of the empirical values of DWR for different values of L with the analytical approximation.	67
4.7	Empirical embedding PAR, as a function of memory size L , for a Gaussian host and $L_h = 100$	69
4.8	Illustration of original and equivalent channel at the decoder.	70
4.9	Empirical and analytical decoding error probabilities for hyperbolic RDM (DWR=25 dB).	72
4.10	Alternative scheme for memory reduction at the encoding side.	73
4.11	Analytical approximation and empirical values of the error probability for different values of L (here $L_h = 50$).	74
4.12	Analytical approximation and empirical values of the error probability for different values of L_h (here $L = 50$).	75
4.13	Analytical approximation and empirical values of the error probability for DM, RDM, logarithmic DM and hyperbolic RDM (DWR=25 dB).	76
4.14	Empirical values of the error probability for power-law attack with $\rho = .7$ and $\gamma = 1.2$ (DWR=25 dB).	77
4.15	Empirical values of the error probability for the alternative scheme and different values of L_h	77
4.16	Empirical values of the error probability for different ordering of Lena host vector (DWR=25 dB).	78

LIST OF FIGURES

4.17 Empirical values of the error probability for scrambled Lena host vector and scrambled Baboon host vector under gamma compression with $\gamma = 1.3$ (DWR=25 dB).	79
4.18 Comparison of original image Baboon and hyperbolic RDM watermarked image Baboon (DWR = 25 dB and Watson distance of 17).	80
4.19 Comparison of original image Lena and hyperbolic RDM watermarked image Lena (DWR = 25 dB and Watson distance of 24).	80
5.1 LTI filtering attack channel.	84
5.2 Block scheme of the whole embedding/decoding chain for DFT-RDM. . . .	88
5.3 Equivalent model for the k th RDM-like channel in DFT-RDM.	90
5.4 BER versus DFT channel for an ideal low-pass filter with $\omega_c = 0.8\pi$ rad, $N = 256$ and DWR = 25 dB.	92
5.5 Comparison of optimal window with rectangular and hamming window. . .	94
5.6 BER vs. DFT channel for DFT-RDM with $N = 256$ and DWR = 25 dB using different windows against an ideal low-pass filter with $\omega_c = 0.8\pi$ rad. . .	95
5.7 BER vs. DFT channel for DFT-RDM with $N = 256$ and DWR = 25 dB using different spreading factors against an ideal low-pass filter with $\omega_c = 0.8\pi$ rad.	96
5.8 BER vs. DFT channel for DFT-RDM applied to different hosts with $N = 256$, $DWR \approx 25$ dB, spreading factor $M = 4$ and optimal window against an audio equalizer.	97
5.9 $R_A(\omega, k)$ versus discrete frequency for $k = 25$ and $k = 250$ ($N = 512$). . . .	100
5.10 Magnitude of the frequency response of the filter $A_{av}(e^{j\omega})$ with order $Q = 10$.102	102
5.11 Per-channel watermark signal power in dB ($\sigma_{x_0}^2 = 1000$ and $N = 256$). . . .	103
5.12 Magnitude of the correlation coefficient $ \rho_{k,t} $ for white and colored watermarked signals evaluated at channels $k = 5$, $k = 50$, $k = 150$ and $k = 250$. . .	104
5.13 Magnitude of the frequency responses of the low-pass filters used in the experiments.	105
5.14 Analytical and experimental results for colored host and low-pass filter with $\omega_c = 0.8\pi$	105
5.15 Analytical WNR versus DFT channel for different order AR filters.	106

LIST OF FIGURES

5.16	Analytical and experimental results for colored host and low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad.	107
5.17	Magnitude of the frequency response of the band-pass filter used in the experiments.	108
5.18	Analytical and experimental results for colored host and band-pass filter.	108
5.19	Magnitude of the ten-band audio equalizer used in the experiments.	109
5.20	Analytical and experimental results for colored host and ten-band equalizer attack.	109
5.21	Block scheme of the whole embedding/decoding chain for W-DFT-RDM.	111
5.22	Experimental power spectral densities of host and watermark signal after reconstruction filtering for DWR = 25 dB.	112
5.23	BERs versus discrete frequency for low-pass filter with $\omega_c = 0.8\pi$ rad.	113
5.24	BERs versus discrete frequency for low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad.	114
5.25	BERs versus discrete frequency for the band-pass filter.	115
5.26	BERs versus discrete frequency for the ten-band equalizer attack.	115
5.27	Magnitude of the frequency response of the whitening filter and the filters $A_{av}(e^{j\omega})$ evaluated for the tracks "Spff" and "Trpt.	119

List of Tables

2.1	Invariant directions extracted from attacked images.	33
2.2	Comparison of DWRs and Watson's distances for constant gain and variable gain.	34
2.3	Detection results on the attacked image Boat (DWR ≈ 40.94 and $D_{wat} \approx 13.36$).	36
5.1	Overall error probabilities for the low-pass filter with $\omega_c = 0.8\pi$ rad ($M = 1$ and rectangular window)	117
5.2	Overall error probabilities for the low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad ($M = 1$ and rectangular window)	117
5.3	Overall error probabilities for the band-pass filter ($M = 1$ and rectangular window)	118
5.4	Overall error probabilities for the ten-band equalizer attack ($M = 1$ and rectangular window)	118
5.5	Overall error probabilities for the low-pass filter with $\omega_c = 0.8\pi$ rad ($M = 8$ and optimal window)	120
5.6	Overall error probabilities for the low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad ($M = 8$ and optimal window)	121
5.7	Overall error probabilities for the band-pass filter ($M = 8$ and optimal window)	121
5.8	Overall error probabilities for the ten-band equalizer attack ($M = 8$ and optimal window)	122
5.9	Overall error probabilities for MP3 compression attacks ($M = 8$ and optimal window)	122

Acronyms and notations

Acronyms

AQIM	Angular Quantization Index Modulation
AR	Autoregressive
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CPTWG	Copy Protection Technical Working Group
DC-DM	Distortion Compensation - Dither Modulation
DC-QIM	Distortion-Compensated Quantization Index Modulation
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DFT-RDM	Discrete Fourier Transform - Rational Dither Modulation
DM	Dither Modulation
DNR	Document to Noise Ratio
DRM	Digital Rights Management
DVD	Digital Versatile Disk
DWA	Digital Watermarking Alliance
DWR	Document to Watermark Ratio
DWT	Discrete Wavelet Transform
ECRYPT	European Network of Excellence for Cryptology
FGA	Fixed Gain Attack
FHI	Filtered Host Interference
FT	Fourier Transform

ACRONYMS AND NOTATIONS

HVS	human visual system
i.i.d.	independent and identically distributed
IDFT	Inverse Discrete Fourier Transform
IEEE	Institute of Electrical and Electronic Engineers
JPEG	Joint Photographic Experts Group
LTI	Linear Time Invariant
MPEG	Moving Picture Experts Group
MSE	Mean Squared Error
PAR	Peak to Average Ratio
pdf	probability density function
psd	power spectral density
PSNR	Peak Signal to Noise Ratio
QIM	Quantization Index Modulation
RDM	Rational Dither Modulation
SDMI	Secure Digital Music Initiative
SS	Spread Spectrum
ST-DM	Spread Transform - Dither Modulation
W-DFT-RDM	Whitened Discrete Fourier Transform - Rational Dither Modulation
WAVILA	WATERmarking VIRTUAL LAB
WIR	Watermark to Interference Ratio
WNR	Watermark to Noise Ratio

Notations

α	Distortion compensation parameter.
Δ	Quantization step-size.
δ_l	Kronecker's delta.
γ	Exponent on the channel for power-law attack.
$\hat{\mathbf{a}}$	Estimate of the vector \mathbf{a} .
Λ	It denotes a generic lattice.
λ	Scaling factor for SS.
\mathbb{R}	Set of reals.
\mathbb{Z}	Set of integers.
\mathbf{A}	M -dimensional random vector.
\mathbf{a}	M -dimensional vector.
\mathbf{a}^*	Complex conjugate of the vector \mathbf{a} .
\mathbf{a}^T	Transpose of the vector \mathbf{a} .
\mathbf{N}, \mathbf{n}	Channel noise signal signal (random variable, deterministic).
$\mathbf{s}^{(m)}$	Spreading sequence for SS.
\mathbf{W}, \mathbf{w}	Watermark signal (random variable, deterministic).
\mathbf{X}, \mathbf{x}	Host signal (random variable, deterministic).
\mathbf{Y}, \mathbf{y}	Watermarked signal (random variable, deterministic).
\mathbf{Z}, \mathbf{z}	Received signal (random variable, deterministic).
\mathcal{M}	Message alphabet.
$\mathcal{N}(\mu, \sigma)$	Gaussian distribution with variance σ^2 and mean μ .
$\mathcal{U}(Z)$	Uniform distribution over Z .
χ_n^2	Chi-square distribution with n degrees of freedom.
ρ	Gain scaling on the channel.
σ_A^2	Variance of a random variable V .
$\tilde{a}_{m,k}$	k th coefficient of the DFT computed on the m th block of the signal \mathbf{a} .
A_k	k th component of vector \mathbf{A} .
a_k	k th component of vector \mathbf{a} .
D_c	Average attacking distortion power.

ACRONYMS AND NOTATIONS

D_w	Average embedding distortion power.
$E[g(A)]$	Expectation of the function $g(A)$, where A is a random variable.
$F_A(a)$	Probability density function of the random variable A .
$I(X; Y)$	Mutual information between X and Y .
j	Imaginary unit.
M_{ap}	p th absolute moment of the random variable A .
$Q_b(\cdot)$	Quantizer corresponding to the symbol b .
$\arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$	It provides the element/s of the set \mathcal{X} that minimizes $f(\cdot)$.

1

Introduction

Digital data hiding can be defined as the practice of embedding information in a digital signal that acts as a "host" or "cover" signal. Since the second half of the 1990's, digital data hiding has received increasing interest from the information technology community. This attention was first triggered by its potential use in applications to protect the intellectual property rights for digital multimedia contents, which is usually referred as digital watermarking. However the range of applications using data hiding techniques soon became largely increased and today there exists a number of companies successfully exploiting data hiding technologies for different purposes.

1.1 A brief history

The invention of digital watermarking can be ascribed to Emil Hembrooke of the Muzak Corporation, who in 1954 filed a patent [1] describing a method for the embedding of inaudible codes into music signals with the objective of proving their ownership. However, it was not until the eighties that the first discussions about digital watermarking can be traced back. The term *digital watermarking* appears to be first used in 1988 by Komatsu and Tominaga [2], even if the interest in digital watermarking technologies started to grow significantly from the mid-nineties, how it is disclosed by the number of papers published on watermarking by IEEE [3]. This increasing attention was motivated by the beginning of Internet pervasiveness and by the wide spread of digital media distribution, which revealed the need of protection for intellectual property rights of digital contents and watermarking seemed to be a feasible option. Consequently, first digital watermarking

techniques was mainly devoted to this aim, as originally conceived in [1]. However, the potential applications rapidly expanded to include a wide range of new applications, such as copy protection, metadata annotation, etc., and at the same time the watermarking methods were applied to an increasing variety of digital content. All these applications were contained in the general definition of "information hiding" technologies.

At the end of nineties, the great effort of information technology community was followed by the attention of several organizations, which began considering watermarking technologies for inclusion in standards. The Copy Protection Technical Working Group (CPTWG) [4] tested the adoption of watermarking technology in DVD to protect video contents. The European Union sponsored some projects [5, 6] to test watermarking for broadcast monitoring. The Secure Digital Music Initiative (SDMI) [7] developed a system for protecting music based on watermarking. In spite of this intense activity, in the late nineties, digital watermarking fell into several failures (the SDMI challenge [8] is the most famous) and some skeptical voices expressed serious concerns in opposition to the effectiveness of watermarking technologies in real-world applications [9]. This criticism brought to a deep reflection about watermarking within the scientific community [10, 11], that reacted in consequence [12], clarifying that digital watermarking was still an open topic and far from being a mature technology and identifying new challenges for information hiding. Some of these were focused in [10, 11], where the scientific community foresaw relevant interest in watermarking security [13] and in side information schemes [14]. In fact these were some of the topics of the Watermarking Virtual Lab (WAVILA) of the European Network of Excellence for Cryptology (ECRYPT) [15], a project funded by the European Community.

From the on, research activity on information hiding continues reaching unmistakable results. Even if digital watermarking technologies have not successfully addressed some of the challenges in the Digital Rights Management (DRM) field [16], they have found market opportunities in other scenarios. For the moment, watermarking technologies are successfully exploited by some companies in a wide range of applications, such as broadcast monitoring, audience metering, video surveillance solutions, or audio and video watermarking for forensic applications. Digimarc [17], Teletrax [18], Philips [19], Thomson [20], Cinea [21], Verimatrix [22], Verance [23], GeoVision [24], MediaSec [25], TRedess [26], Datamark [27], Civolution [28], Markany [29], Widewine [30], MSI Copy Control [31]

and Aquamobile [32] are some examples of companies with business models based on information hiding technologies.

Furthermore, the Digital Watermarking Alliance (DWA) [33] has been recently constituted by "a group of companies that share a common interest in furthering the adoption of digital watermarking and which are actively involved in commercialization of digital watermarking-based applications, systems and services" [33]. This renewed interest is justified by some reports [34] that foresee in the next years a rapid grow of identification technologies, such as digital watermarking, surpassing US \$ 500 million worldwide by 2012.

1.2 Applications

The first digital watermarking techniques were developed for digital images in the early nineties, exploiting elementary image processing manipulations. During the last years, a huge variety of digital data hiding methods have been proposed for a wide range of digital contents. In fact in any kind of digital content a hidden information can be inserted as long as the given work is suitable for being represented in two (or more) different ways without significant differences perceived by the user. In literature, data hiding techniques have been proposed for video, images but also for software and electronic text documents, exploiting the properties of the different contents and formats; an overview can be found in [35] or [6].

To get a general idea of what is under the wide term of digital data hiding, in this section some applications of data hiding in digital contents will be briefly described:

- **Demonstration of rightful of ownership:** the author of a work embeds in it, as soon as he creates the work, a watermark to prove unambiguously the ownership. Obviously, the watermark has to be designed to survive the attacks that pirates could perform to remove it [3,6,36]. Even if this application was one of the first to be faced, it is also one of the most challenging, since the opponents are supposed to be aware of the watermark existence and a wide range of attacks has to be considered [37].
- **Fingerprinting:** it is a way to achieve copy deterrence by providing a mechanism to trace unauthorized copies of a protected work. In each copy of data that is distributed, the owner inserts a distinct watermark, called fingerprint, which unambiguously identifies the buyer. In this way, if an unauthorized copy of the protected

work is found, the owner of the copyright can retrieve the identity of the buyer that illegitimately distributed the content [3, 35, 36].

- **Copy control:** it aims at preventing users from making copies of a protected content, when this is not allowed. The embedded watermark describes the rights of copying owned by the user and every recording device is supposed to be equipped with a watermark detector, so that the device can prohibit recording whenever the watermark that prevents copying is detected in the content. Similar systems were studied by CPTWG and SDMI, for video DVDs and audio respectively [3, 36, 38].
- **Authentication:** the availability of cheap and effective digital multimedia editing tools makes easy the modification of original contents without leaving any perceptible traces of tampering. In this way, a digital data can never be trusted to be authentic. In this application, the embedded watermark, which is modified together with the host signal, reveals when it has been tampered, even after small changes, so that modifications on the watermarked content can be detected [3, 36, 39]. These schemes are said to use fragile watermarks, in contrast to robust watermarks which must remain unchanged by processing. As an example, they are used to guarantee the integrity of automatic video surveillance data.
- **Data compression:** combining data compression and information hiding has been investigated to increase the effectiveness of the overall compression. The basic idea is to compress only a part of the content and to hide the remaining part into the compressed content itself [36]. As an example the chrominance component of an image can be embedded into the luminance component or the audio stream can be embedded into the relative video stream [40]. This application is quite interesting because, traditionally, data hiding and data compression are considered contradictory operations, since ideal perceptually lossless compression aims at removing all the information that cannot be perceived by a human user. But, since perceptual equality does not mean mathematical equality, the effectiveness of data hiding in presence of ideal perceptually lossless compression is an open issue.
- **Error recovery:** transmission of data in compressed formats, such as JPEG or MPEG, is vulnerable to transmission errors. A typical way to cope with these errors is channel coding at cost of introducing a controlled amount of redundancy. A

similar approach can be performed through data hiding, which allows to hide into the content redundant informations that can be possibly used to recover from errors at the decoder side [41, 42]. As an example, the hidden redundant information can be simply a low quality version of the content.

- **Annotation watermarks:** the data hiding schemes can be used as a way to insert an annotation within the work to enhance its value. The capability of the embedded information to survive digital to analog and analog to digital conversion allows the annotation data to be associated to the work itself and format-independent [3, 36, 43, 44]. As an example, a label describing a video object belonging to a MPEG-4 stream can be hidden within each object. In this way, the label is indissolubly tied to the object it refers to and it can be used by retrieval systems to automatically recognize the content of the video object.
- **Covert communications:** the ultimate goal of covert communications is to hide the very existence of the hidden message, so that the major requirement for this applications is the undetectability of the presence of the message. The term steganology is used to refer to both steganography and steganalysis, which are respectively the practice of undetectably altering a content to embed a secret message and the practice of discovering the presence of steganographic channels [3, 36].
- **Broadcast monitoring:** a label, which identifies a multimedia content to be broadcasted, can be hidden within the content itself to systematically detect when it is on air and to provide immediate reporting [5, 45]. This application can be used by both broadcasters, to track their playlists and copyright royalties, and by advertisers, to verify the exact execution of their campaigns' media planning.
- **Link quality estimation:** a fragile watermark has been also used to provide a method for blind estimation of the quality of multimedia communication links at the receiver side. This quality assessment system is based on the evaluation of the mean-squared-error between the received and the actual watermark signal, providing a quality measure of the effective status of the link without increasing the bit rate [46].

As a final comment, this brief overview cannot be clearly exhaustive and exact boundaries between applications are quite difficult to be drawn, since some of them can be overlapping.

1.3 General requirements

The development of a data hiding system, depending on the particular application, involves several trade offs between unlike requirements. Generally, the watermark should be imperceptible, the amount of conveyed information should be as large as possible, it should be robust to content modifications and, in most watermarking applications, it should be secure. As a consequence, in literature the most relevant requirements considered are:

- **Payload:** the amount of information (the number of bits) that the watermark is able to convey. This requirement is strictly related with the application, as an example annotation watermarks usually needs larger payload than fingerprinting, where robustness and security are much more important requirements.
- **Imperceptibility:** a fundamental property of any data hiding system is that the watermarked signal should be perceptually indistinguishable from the original one. In other words, the embedding process has not to downgrade the quality of the digital content. Perceptual analysis has been widely exploited to find the optimal embedding domains and perceptual masks that maximize the power of the watermark without impairing the perceived quality of the content.
- **Robustness:** it accounts for the capability of the hidden data to survive host signal manipulations. Attacks to robustness are those whose target is to increase the probability of error of the data hiding channel [47]. In data hiding are considered both non-malicious manipulations, which do not explicitly aim at removing the watermark or at making it unreadable, including conventional signal processings and digital format conversion, and malicious manipulations, which precisely aim at making the hidden information no more retrievable.
- **Security:** it is defined as "the inability by unauthorized users to access (i.e., to remove, to read, or to write the hidden message) the communication channel" [48]. Security is usually related with the ability of a data hiding scheme to protect some secret parameter, so that an attacker can not use it to access the watermark contents. According to this definition, attacks to security are those aimed at gaining knowledge (measurable as information) about the secrets of the system (e.g. the embedding and/or detection keys) [47].

It is worth noting that the above defined requirements are usually colliding [39]. As an example, increasing the payload necessarily yields a decreased imperceptibility or a decreased robustness of the watermark. Depending on the particular application, a trade-off has to be reached to balance among the colliding requirements to best suit the faced problem.

1.4 Classification of attacks

An attack can be every processing applied to the marked content, as long as the attacked content has an acceptable perceptual quality. So defined, the attack can be both common signal processing operations, that can occur during the normal life cycle of a digital content, and ad-hoc attacks performed by malicious users aiming at removing, reading or modifying the hidden data. A possible classification of the various attacks is here outlined. We can distinguish between malicious and non malicious attacks:

- **Non malicious:** non malicious attacks are considered those attacks that can occur during the normal use of the marked content and consequently their nature and strength is strongly application dependent. Among these we have lossy compression, digital to analogue conversion , processing aimed at enhancing the quality, etc. .
- **Malicious:** malicious attacks are processing performed with the explicit aim of remove the watermark or impair its retrieving. These attacks can be divided in blind attacks, where the attacks is performed without exploiting any knowledge about the particular watermarking method, and informed attacks, where the effectiveness of the watermarking system is impaired using some information about the algorithm.

A more fitting categorization has been given in [49], where the attacks are classified in

- **Watermark removal:** in this class of attack, the attacker is not necessarily concerned with the semantics of the bits carried by the watermark. He is only concerned with removing the message, without aiming at knowing it, also because he can already know the message (i.e. 'you may not copy'). Hence, the purpose is to modify the marked content such that the perceptual quality of the media is retained, but that the watermark detection is impaired. From the above definitions, the removal attacks are the typical attacks to robustness.

- **Unauthorized watermark estimation or detection:** the aim of this type of attacks is to estimate the watermark signal. Very often, convenient filtering is used to estimate a good approximation of the original content, so that the difference between the watermarked item and the estimated original is taken. Removal attacks and estimation attacks are strictly related, since any estimation attack can be used to build a removal attack and any removal attack can be used to build an estimation attack.
- **Unauthorized watermark writing attack:** the Copy Attack [50] is an example of this class. The watermark signal is firstly estimated from the watermarked content and then it is embedded in unmarked content. This attack typically succeeds for watermark signals that are independent of the content to be marked, such as simple additive spread-spectrum technique.

However, any classification of attacks cannot be exhaustive because the distinction between different classes is sometimes a fuzzy one and because the attacks are highly dependent on the particular application. First of all, different applications need to be resistant to different attacks, as an example annotation watermark will not have to care about malicious attacks. Moreover the possible attack are strictly related with the content type where the information is embedded. In image data hiding techniques, geometric manipulations have to be taken into account, while a relevant attack for video watermarking can be temporal scaling, such as temporal resampling or frame insertion and deletion.

An overview of attacks on generic data hiding systems can be found in [36] and [3], while a classification of attacks to security can be found in [13]. Finally, a brief categorization of existing attacks for images, video and audio can be found respectively in [37], [51] and [52]

1.5 Summary and thesis objectives

As described above, robustness is a fundamental requirement for watermarking/data hiding schemes. Even if a great effort has been spent in this topic by the research community, there are still open problems and this is not surprising. In fact a multimedia content can undergo a huge variety of processing and each of them can impair in a different way the correct hidden information retrieving. On the other hand, the data hiding methods are

constrained by the imperceptibility requirement to embed low-power signals in order to slightly modify the content.

As a consequence, the main topic of the thesis is the study of data hiding methods that provide robustness against non-additive attack, guaranteeing at the same time high capacity capability and preserving the content fidelity. This study has been motivated by the fact that, for the class of quantization-based embedding methods, there is much less studied about non-additive attack channels, which model most of the processings in a real context. Consequently, achieving robustness to non-additive attacks is an essential requirement to develop practical data hiding applications based on this class of methods.

This thesis concerns the development of novel data hiding methods, within the class of quantization-based, intrinsically invariant to the non-additive attacks taken into account. The proposed methods have been tested with real signals, such as images and audio tracks, and theoretical analysis are also developed for them in order to prove the achievable performance.

The remaining chapters are organized as follows.

- Chapter 2 describes how the information hiding problem is, in essence, a classical communications problem. The spread spectrum-based and the quantization-based embedding methods are shortly described, highlighting the respective properties. Then the wide class of desynchronization attacks is presented and the ability of the different embedding approaches in coping with these attacks is discussed. A spread spectrum-based watermarking method robust to geometric distortions, such as rotations, is here presented showing that robustness against complex distortions can be quite straightforwardly achieved using these techniques. This method for image watermarking has been published in a conference paper [53] and in a journal paper [54].
- Chapter 3 presents the dither modulation, which is the basic quantization-based data hiding algorithm, and its high robustness to additive white Gaussian noise channel is demonstrated. Then the weakness of these methods against volumetric distortions, such as gain scaling, is discussed and the different approaches to achieve robustness to this class of attacks are presented. Eventually, the Rational Dither Modulation is described as a method to achieve asymptotically the performance of DM against additive noise and, at the same time, theoretical invariance to constant gain scaling.

- Chapter 4 studies the power-law attack, which consists of a constant exponentiation and a constant gain scaling. Then it is presented a class of data hiding methods theoretically invariant to this nonlinear valumetric attack. This property is achieved developing a proper extension of any gain invariant quantization-based algorithm which consists of performing the embedding in a convenient invariant domain. Among this class, the method called Hyperbolic RDM is deeply analyzed to identify the lower bound of the decoding error probability and to quantify the embedding distortion. The obtained results have been published so far in a conference papers [55], and in a journal paper [56].
- Chapter 5 addresses the problem of coping with filtering for quantization-based techniques. The discrete Fourier transform - rational dither modulation was proposed to achieve robustness against filtering without assuming any previous knowledge about the attack filter. Here the analysis of this method is generalized for non-white hosts to provide a reasonable explication of the loss of performance when it is applied to this class of signal. Departing from this generalized analysis, an extension of discrete Fourier transform - rational dither modulation is presented to improve its performance for non-white hosts. Moreover, the expected performance improvement is also verified for real audio signals, that are nonstationary, non-Gaussian and colored hosts. A conference paper [57] and a journal paper [58] are to be published.
- Finally, Chapter 6 summarizes the main conclusions that can be extracted from this thesis and presents a series of improvements and open topics that deserve further analysis in the future.

2

Watermarking as communications: channel models and embedding strategies

Abstracting from the particular content type where the information has to be hidden, the information hiding problem is, in essence, a classical communications problem. In fact, information hiding aims at transmitting an information (the watermark) through a channel (the host document) by means of a modulator (the embedder) and at recover it at the receiver (the watermark detector).

Consequently digital communications played a fundamental role in the development of watermarking technology, providing the modeling of watermarking as a communications problem, inspiring the most relevant embedding techniques and suggesting the approach to perform rigorous performance analyses.

First, Section 2.1 presents the general communications model for digital data hiding that will be used throughout this thesis. Section 2.2 defines a common set of objective measures to evaluate the attack strength and the distortion introduced by embedding functions. In Section 2.3 the spread spectrum paradigm for data embedding is presented, while in Section 2.4 the class of side-informed data hiding schemes is introduced. The desynchronization attacks, their effect on basic data hiding decoding performance and the strategies to cope with such impairments are addressed in Section 2.5. A spread spectrum-based watermarking technique for still images robust to geometric transformations is presented in Section 2.6 to show the resilience of this class of embedding methods against complex

desynchronization transformations. Some final considerations are summarized in Section 2.7.

2.1 Watermarking as communications

During the first years of research on data hiding, algorithms flooded the literature, in most cases, with weak theoretical grounds and with little concern about robustness or performance [59, 60]. A relevant milestone in data hiding research is a paper by Cox et al. [61], where it is recognized that data hiding itself was in fact a particular case of communications, suggesting the way for using well-known techniques and facilitating subsequent performance analyses. In [49], Kalker define watermarking as "a mechanism to create a communication channel that is multiplexed into original content". In Fig. 2.1 the model of a digital data hiding system as a digital communications problem is shown.

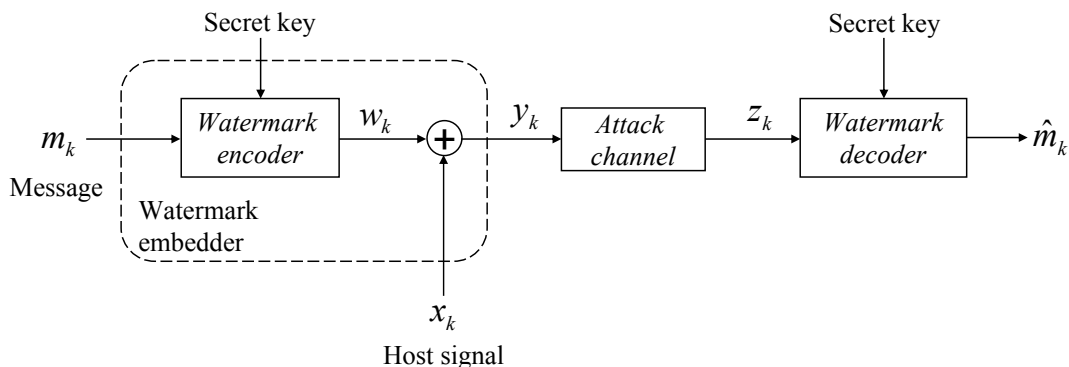


Figure 2.1: Watermarking as communication channel.

In a generic scheme, the watermark is embedded in a host signal, which is here arranged in an one-dimensional vector denoted by \mathbf{x} . The host signal can be assumed to be computed from the digital content to be marked in order to have a suitable set of features for the embedding. As an example, the host signal can be a collection of coefficients of the digital content in a transformed domain, such as DCT, DWT or DFT. The message samples to be embedded are denoted by m_k and they are assumed to belong to a p -ary alphabet. The other input to the watermark embedder is the secret key, which is usually a secret shared with the decoder, and it allows the randomization of the embedding function in order to guarantee the security of the hidden information.

The watermark embedder, from the knowledge of the message and of the secret key, modifies properly the host signal \mathbf{x} to give the watermarked signal \mathbf{y} , that is mapped back to the original domain (if it is needed) to have the actual watermarked content. The watermarked signal \mathbf{y} produced by any data hiding scheme can be seen as the summation of the host signal \mathbf{x} and the watermark signal \mathbf{w} , which conveys the hidden information, so that we can write

$$\mathbf{y} = \mathbf{x} + \mathbf{w} \tag{2.1}$$

Once the marked content is generated, it undergoes the channel, where it is subject to attacks, which are usually modeled by means of a probabilistic channel. An attack can be every processing applied to the marked content, as long as the attacked content has an acceptable perceptual quality. So defined, the attack can be both common signal processing operations, that can occur during the normal life cycle of a digital content, and ad-hoc attacks performed by malicious users aiming at removing, reading or modifying the hidden data.

After the possible attacks have taken place, the attacked content is presented to the decoder. Here the attacked feature host vector \mathbf{z} is computed performing the same operations that are performed at the embedder to have the host signal departing from the original content. From the attacked host \mathbf{z} , the decoder produces an estimate $\hat{\mathbf{m}}$ of the the embedded message. We remark that only blind decoding [36] will be here taken into account, which means that the decoder is assumed to not have access to the original host signal \mathbf{x} .

It is worth introduce here some notations. The host vectors are assumed to be arranged in one-dimensional vectors and, as it has been shown previously, boldface fonts will be used to denote vectors or sequences, so that the host signal is $\mathbf{x} = (x_1, x_2, \dots, x_M)^T$ and its k th sample is x_k . Moreover, in order to analyze the data hiding problem from a theoretical point of view, the signals involved in the system will be modeled as random variables. Heretofore, the random variables will be denoted by capital letters X and random vectors by $\mathbf{X} = (X_1, X_2, \dots, X_M)^T$, while their instantiations will be denoted by respective lowercase letters. The probability density function of a random variable X will be referred as $f_X(x)$.

2.2 Distortion measures

As it has been discussed in Section 1.3, one of the fundamental requirements of data hiding is the imperceptibility of the watermark, which is equivalent to require that, even if the information embedding introduces a distortion, the host signal and the watermarked signal have to be close enough, according to a given metric. The definition of a common set of objective measure to evaluate the distortion introduced by the embedding function, as well as the attack strength, is needed to compare data hiding systems based on different embedding and decoding rules. According to the digital communication approach, the distortion measures are usually mean square error (MSE) based.

To evaluate the embedding distortion of the data hiding method, the Data to Watermark Ratio (DWR) is used, which is defined as the ratio between the power of the host signal and that of the watermark signal. Expressing the distortion measure as a ratio has been motivated by the consideration that a stronger signal can accommodate a stronger hidden signal, preserving the perceived quality. Hence the DWR for the host signal \mathbf{x} and the relative watermark signal \mathbf{w} is defined as

$$DWR \triangleq \frac{(1/M) \sum_{k=1}^M x_k^2}{(1/M) \sum_{k=1}^M w_k^2} \quad (2.2)$$

However a so computed distortion measure is strictly inherent to the particular host and watermark signal. To quantify the distortion introduced by the embedding method itself, it is more useful to get a global measure of the embedding distortion for a given algorithm by averaging the power of the host signal and the power of the watermark signal over all possible hosts and watermarks. Considering the average power for both the signals and modeling the sample as random variable, the average embedding distortion is defined as

$$DWR \triangleq \frac{(1/M) \sum_{k=1}^M E[X_k]^2}{(1/M) \sum_{k=1}^M [W_k]^2} \quad (2.3)$$

where $E[\cdot]$ denotes the statistical expectation. Throughout the thesis, the average power of the watermark signal will be also referred as *embedding distortion*, which is by definition equal to $D_w \triangleq (1/M) \sum_{k=1}^M E[W_k]^2$.

Similarly, the strength of the attack is usually measured by the Watermark to Noise Ratio (WNR), which is the ratio between the power of the watermark signal and the *attacking distortion*, which is generally defined as the power of the noise introduced by

the attack. Abstracting from the particular attack performed on the channel, for noise it is intended the signal difference between the attacked signal and the watermarked signal $\mathbf{N} = \mathbf{Z} - \mathbf{Y}$. For a particular watermark signal and a particular noise vector, the WNR is given by

$$WNR \triangleq \frac{(1/M) \sum_{k=1}^M w_k^2}{(1/M) \sum_{k=1}^M n_k^2} \quad (2.4)$$

Assuming the noise samples modeled as random variables, the average WNR is defined as

$$WNR \triangleq \frac{(1/M) \sum_{k=1}^M E [W_k]^2}{(1/M) \sum_{k=1}^M [N_k]^2} \quad (2.5)$$

that is obtained by averaging over all watermarks and all noise vectors to measure the attack strength on the random additive channel. The attacking distortion is then given by $D_c \triangleq (1/M) \sum_{k=1}^M E [N_k]^2$

Another measure of the attack strength is the Data to Noise Ratio (DNR), which is defined as the ratio between the power of the host signal and that of the noise introduced by the attack. This measure is useful in scenarios where it is important to evaluate the maximum allowable distortion that an attacker can introduce without degrading the marked content.

However, these simple MSE-based measures have the advantage that can be used independently of the particular content type and the particular domain where the embedding is performed, even if they does not take into account the perceptual characteristics of the introduced distortions, which are strictly related with the content type. In fact, MSE describes the total power devoted to the watermark and it should be handled with care when dealing with perceptual constraints, since masking phenomena affecting perception and exploited for the hiding of information generally respond to local effects. On the other hand, defining constraints on the power of watermark signal is very convenient from an analytical point of view, as it will be shown in the following.

2.3 Spread Spectrum watermarking

In the data hiding channel modeled as in Fig. 2.1, the watermark signal is directly mixed with the host so that the host itself is treated as an additional noise source impairing the transmitted signal, which is the watermark. Moreover, given the imperceptibility

2.3 Spread Spectrum watermarking

constraint, the watermark signal has a much lower strength than the host, which acts as a noise source. In digital communications, systems which have to cope with very noisy channels, possibly affected by intentional disturbs such as jamming or interferences, are usually based on spread spectrum (SS) technology [62]. Hence, in early years, many of the watermarking algorithms used a similar technique to code the to-be-hidden information, as it was firstly proposed in [61].

In agreement with the spread spectrum paradigm, the message \mathbf{m} is transformed into a pseudo random sequence $\mathbf{s}^{(\mathbf{m})}$, named spreading sequence, whose length is usually much larger than that of \mathbf{m} and whose elements are random variables drawn from a given pdf. Then $\mathbf{s}^{(\mathbf{m})}$ and the selected host feature sequence are mixed to perform the embedding. Among the different embedding functions [36], the most common and simple approach is the additive one:

$$\mathbf{y} = \mathbf{x} + \lambda \mathbf{s}^{(\mathbf{m})} \quad (2.6)$$

where λ is a scaling factor which allows to control the watermark strength, being the introduced distortion $D_w = \lambda^2(1/M) \sum_{k=1}^M [S_k]^2$.

All the possible spreading sequences are assumed to be known to the decoder. The decoder's task is to determine whether a watermark sequence is present in the received signal and/or which sequence is within it. A correlation measure of the received signal with all the possible spreading sequences is usually employed to estimate the watermark presence and to make an estimate of the embedded message $\hat{\mathbf{m}}$. Under the assumption that the host features are drawn from i.i.d. Gaussian random variables and an Additive White Gaussian Noise (AWGN) channel is take into account as the only attack, correlation-based decoding is optimum, in both the minimum error probability and missing watermark detection sense [63]. In Fig. 2.2, the watermarking channel model under the above assumptions is depicted describing the involved signals as random variables.

As in communications [64, 65], information theory is a very useful tool in establishing the fundamental limits of the considered communications problem. Consequently watermarking schemes can be evaluated with respect to the channel capacity, which is defined as the maximum amount of information that can be transmitted through the channel and decoded without any errors.

Hence, being $X \sim \mathcal{N}(0, \sigma_x)$ and $N \sim \mathcal{N}(0, \sigma_n)$, by definition the capacity of the channel in Fig. 2.2 is defined as

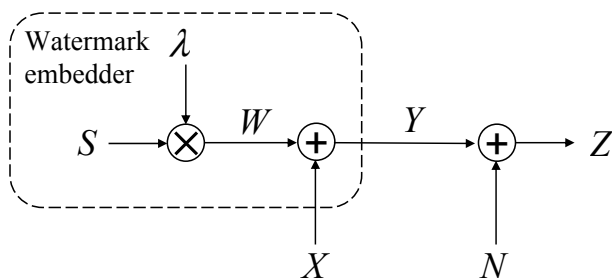


Figure 2.2: Gaussian channel for spread spectrum watermarking.

$$C = \max_{f_W(w)} I(W; Z) \quad (2.7)$$

and it can be shown [66] that for $W \sim \mathcal{N}(0, \sigma_w)$ the maximum is achieved

$$C = \frac{1}{2} \log \left(1 + \frac{\sigma_w^2}{\sigma_x^2 + \sigma_n^2} \right) \quad (2.8)$$

leading to the result that Gaussian watermarks are optimal when the host signal and attack channel are both Gaussian.

However the main consideration that can be drawn from eq. (2.8) is that, treating the host as an additional noise source, spread spectrum watermarking schemes can achieve good performance in case the host variance is small in comparison to the watermark signal variance. Actually in real applications, since the watermark distortion is constrained to preserve the quality of the marked signal, we have $\sigma_x^2 \gg \sigma_w^2$. Moreover, assuming the attack strength also constrained to not degrade the content quality, $\sigma_x^2 \gg \sigma_n^2$ can be reasonably assumed. From these considerations the main limitation of SS-based watermarking against AWGN channel dramatically comes to light, since a small capacity is reasonably expected for $\sigma_x^2 \gg \sigma_w^2, \sigma_n^2$ [67, 68].

It is worth remarking that Gaussian channel is commonly used in the analyses of data hiding schemes not only in analogy with communications, but also since it can be easily applied to any watermarking method allowing to identify the upper capacity bound [69]. Moreover Gaussian noise is a good model for those applications in which robustness against unintentional attacks is required [67].

2.4 Watermarking as communications with side information at the transmitter

The second fundamental milestone in watermarking research can be identified in a paper by Cox et al. [14], where it was realized that the host, even if it is totally unknown at the decoder side, is perfectly known by the embedder. Then, a power-constrained communication channel with side information at the encoder, shown in Fig. 2.3, was recognized as a more fitting model for the watermarking channel, since the transmitted signal is impaired by two noise sources and one of them is known by the encoder; this paradigm is usually termed as *side-informed data hiding* and it was further developed in other works too [70, 71].

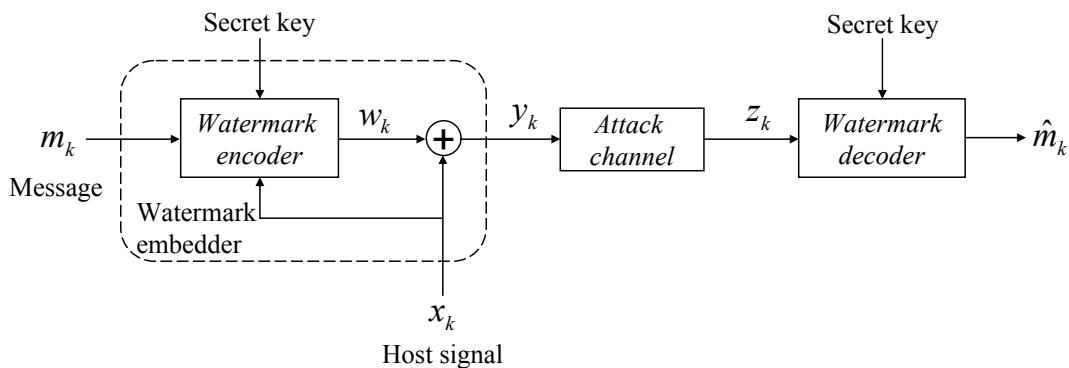


Figure 2.3: Watermarking as communication channel with side information at the encoder.

The research effort spent in the past years studying side-informed data hiding has been motivated by a surprising result by Costa [72], that was rescued by Chen and Wornell. In [72] it is proven that for the case of i.i.d. Gaussian host features, that are non-causally known by the encoder and unknown by the decoder, and i.i.d. Gaussian noise channel independent of the host signal, the channel capacity is given by

$$C = \frac{1}{2} \log \left(1 + \frac{\sigma_w^2}{\sigma_n^2} \right) \quad (2.9)$$

which is the same capacity that can be obtained for the i.i.d. Gaussian noise channel without the additional noise source known by the encoder. This result proves that is possible, at least theoretically, to transmit the information data rejecting the interference of the host signal, even if it is totally ignored at the decoder side, by applying channel coding

2.4 Watermarking as communications with side information at the transmitter

with side information at the encoder. The main problem of Costa's scheme is that the capacity in eq. (2.9) is reached with a channel code obtained by random construction, so its complexity makes its implementation unfeasible in real applications. In fact, the information theoretic analysis of communications with side information is not a constructive one: the existence of codes able to achieve the capacity is demonstrated but their practical construction is not provided.

However, the theoretical results shown above gave inspiration to look for practical schemes able to reject, at least partially, the interference between the host and the hidden signal exploiting the side information at the encoder. So far, the proposed side-informed data hiding schemes can be summarized in those based on structured codebooks [67, 73], named quantization-based methods, and those based on Trellis [74, 75], named dirty-paper trellis code watermarking methods. The quantization-based framework was firstly proposed by Chen and Wornell in [67], where the Distortion Compensated-Quantization Index Modulation (DC-QIM) was presented. The basic procedure of DC-QIM involves the quantization of the host signal using a multidimensional quantizer $\mathbf{Q}_{\mathbf{m}}(\cdot)$ belonging to a finite set of available quantizers and indexed by the message $\mathbf{m} \in \mathcal{M}$ to be embedded. Then the watermarked signal is obtained by adding back to the quantized host signal the quantization error scaled according to a tunable parameter α ; this operation, which is referred as distortion compensation, makes DC-QIM equivalent to Costa's scheme. The DC-QIM embedding rule is thus given by

$$\mathbf{y} = \mathbf{Q}_{\mathbf{m}}(\alpha\mathbf{x}) + (1 - \alpha)(\mathbf{x} - \mathbf{Q}_{\mathbf{m}}(\alpha\mathbf{x})) \quad (2.10)$$

The parameter α has to be properly chosen to yield the achievable rate under additive white Gaussian noise independent of the host signal [67, 72]. The block scheme of the DC-QIM embedder is shown in Fig. 2.4.

The DC-QIM decoder acts by quantizing the received signal using the same set of multidimensional quantizers that are used at the encoder side. The estimated embedded message is the one that indexes the quantizer having the minimum distance from the received signal, according to a convenient distance metric:

$$\hat{\mathbf{m}} = \arg \min_{\mathbf{m} \in \mathcal{M}} \text{dist}(\alpha\mathbf{z}, \mathbf{Q}_{\mathbf{m}}(\alpha\mathbf{z})) \quad (2.11)$$

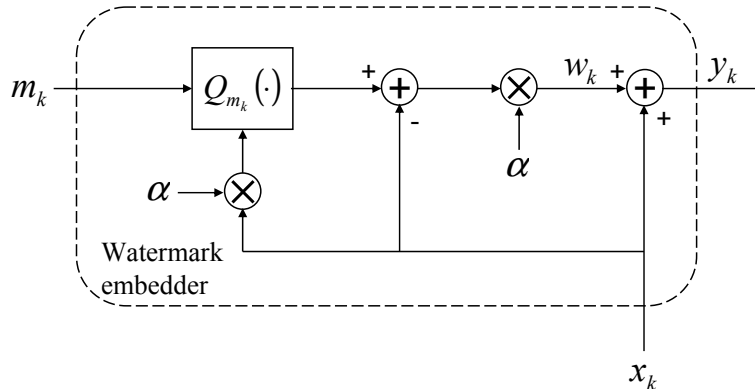


Figure 2.4: DC-QIM embedder.

Moreover, Chen and Wornell also gave the first proposal to put DC-QIM into practice with Distortion Compensated-Dither Modulation (DC-DM) [67, 73]. In DC-DM the set of quantizers are dithered (shifted) versions of a basic one and, due to the implementation and design issues associated to multidimensional quantizers, this basic quantizer usually relies on a lattice. In most of practical implementations, the lattice is obtained as the Cartesian product of scalar uniform quantizers, that makes the DC-DM both straightforward to be implemented and analyzable with probabilistic tools. Full description of DM-based embedding will be given in Section 3.1.

2.5 Desynchronization attack channel

In a communication system, the synchronization issue is a key point and the same applies also in information hiding systems. Indeed, a signal bears information only relatively to its original references and modifying the references results in a damage of the conveyed information. In the classification of attack channels that can be considered in a data hiding system, the class of *desynchronization attacks* has particular relevance since commonly employed encoders and decoders have often a limited resilience against them [39, 76–78]. On the other hand, among the desynchronization attacks, very common and simple processing operations can be included, such as filtering, amplitude scaling, gamma correction, time varying delays and spatial warping. The perceptual effects of such operations are normally quite weak, but the effects on the hidden data retrieving performance can be devastating due to the introduced desynchronization [79, 80].

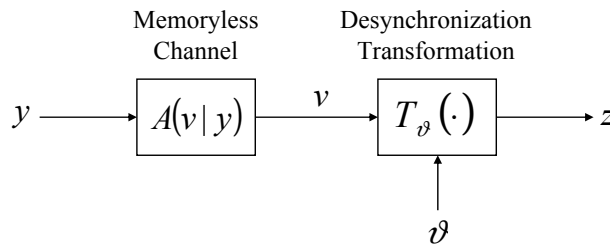


Figure 2.5: Model for desynchronization attacks.

From a theoretical point of view, any attack channel can be modeled by a conditional distribution $p(\mathbf{z}|\mathbf{y})$. To take into account both memoryless impairments and more complex transformation, the attack channel can be decomposed in a memoryless channel $A(\mathbf{v}|\mathbf{y})$ followed by a parametric desynchronization transformation $\mathbf{T}_\vartheta(\cdot)$ [39]. The vector ϑ contains the desynchronization parameters whose dimensionality is independent of the dimensionality of \mathbf{y} . The described model for desynchronization attack channel is depicted in Fig 2.5. The memoryless channel can model the addition of a white Gaussian noise independent of both the the watermarked signal and the desynchronization parameter. Oppositely the parametric desynchronization transformation can model a variety of attacks, such as:

- **Offset:** $y_k = \vartheta + x_n$.
- **Amplitude scaling:** $y_k = \vartheta x_n$.
- **Gamma correction:** $y_k = x_n^\vartheta$.
- **Temporal shifts:** $y_k = \sum_i h_i(\vartheta)x_{k-i}$, where ϑ is generally not an integer and $h_i(\vartheta)$ are the taps of an interpolation filter.
- **Linear Time Invariant (LTI) filtering:** $y_k = h_k(\vartheta) * x_k$, where $*$ denotes convolution and $h(\vartheta)$ is the impulse response of a real valued LTI filter.

Other than these quite simple attacks, in the class of desynchronization attacks, more complex transformations of the watermarked content can be included. For instance, in the case of still images, any geometric transformation which the watermarked data can undergo is an attack that could desynchronize the decoder and the hidden information [81]. Achieving robustness for watermarking schemes against geometric transformations is a

very difficult task and in the last ten years in the watermarking community a great research effort has been devoted to look for techniques able to guarantee the synchronization between the embedder and the decoder against complex geometric distortions [82–87] as well as to propose universal decoders for coping with them [88].

2.5.1 Performance of basic decoders

We focus now on the performance of the basic decoders for spread spectrum-based and quantization-based data hiding techniques against desynchronization attack according to [39], where the attacked signal is defined as $\mathbf{z} = \mathbf{T}_\vartheta(\mathbf{y}) + \mathbf{n}$, being \mathbf{n} a vector of i.i.d. Gaussian noise samples independent of the host signal. This channel can be considered a particular case of the channel model shown in Fig. 2.5. The aim of this Section is to give an intuitive understanding of the impact of desynchronization on basic decoders, without trying to be exhaustive due to the variety of both attacks and data hiding algorithms [3,36].

- **Spread Spectrum:** according to the the SS embedding formula and to the above described channel model, the attacked signal can be written as

$$\mathbf{z} = \mathbf{T}_\vartheta(\mathbf{x} + \lambda\mathbf{s}) + \mathbf{n} \quad (2.12)$$

The basic SS decoder computes the correlation statistic between the received vector \mathbf{z} and all the possible spreading sequences \mathbf{s} . For desynchronization attacks such as amplitude modulation or time warping, according to [39], the following linear approximation can be considered

$$\mathbf{z} = \mathbf{T}_\vartheta(\mathbf{x} + \lambda\mathbf{s}) \approx \mathbf{T}_\vartheta(\mathbf{x}) + \lambda\mathbf{T}_\vartheta(\mathbf{s}) \quad (2.13)$$

To mitigate the effects of on the correlation statistics, we would like to have $\mathbf{T}_\vartheta(\mathbf{s}) \approx \mathbf{s}$. Hence, for the basic correlation decoder to perform as intended, the watermark itself should be nearly invariant against desynchronization attacks.

- **Quantization Index Modulation:** using minimum distance based decoding, QIM-based systems essentially depend on the strength of the overall additive noise at the decoder. In the considered scenario, a further distortion term $\mathbf{e}(\vartheta) = (\mathbf{T}_\vartheta(\mathbf{y}) - \mathbf{y})$ is added to the white Gaussian noise independent of the signal. Consequently the attacked signal is defined as

$$\mathbf{z} = \mathbf{T}_\vartheta(\mathbf{y}) + \mathbf{n} = \mathbf{y} + \mathbf{e}(\vartheta) + \mathbf{n} \quad (2.14)$$

Hence, depending on the particular distortion attack, different impairments will be produced. As an example, for an offset attack, we have $\mathbf{e}(\vartheta) = \vartheta$ and the mean-squared error (MSE) of the attack noise is increased from σ_n^2 to $\sigma_n^2 + \vartheta^2$, which can be significant if $|\vartheta| > \sqrt{\sigma_n^2}$. Otherwise, for an amplitude scaling attack, a signal dependent noise is produced $\mathbf{e}(\vartheta) = (\vartheta - 1)\mathbf{y}$ and, assuming $E[\mathbf{N}^T \mathbf{Y}] = 0$, the MSE of the attack noise becomes $\sigma_n^2 + (\vartheta - 1)^2 \sigma_y^2$, whose effect is significant if $(\vartheta - 1)^2$ exceeds the noise-to-host power ratio σ_n^2 / σ_y^2 . In the following chapters other examples and experimental results about the extreme sensitivity of QIM decoding against desynchronization transformations will be shown.

Therefore, the effect of even mild desynchronization attacks on basic QIM decoders can cause a dramatic increase of the error probability and that is why robust quantization-based algorithms are needed to employ them in real applications.

It is worth recalling that a dramatic performance decrease for desynchronization attacks is not surprising, since already in [71] it was shown that, under squared-error constraints, the data hiding system capacity when the attacker is restricted to additive attacks is strictly larger than the capacity for an attack channel specified by a conditional distribution.

2.5.2 Achieving robustness to desynchronization attacks

In the literature various approaches have been proposed to better cope with desynchronization attacks. Even if SS-based methods and QIM-based methods are profoundly different, the strategies to have a decoding resilient to the lack of synchronization between embedder and decoder are essentially the same. We can roughly classify them in three main classes:

- **Two-steps decoders:** the desynchronization parameter ϑ is firstly estimated, possibly using a large search over its space. Then, having the estimate available, the desynchronization attack can be inverted and the resulting sequence is fed into the basic decoder [?, 89–95]. The main problem of this approach is the potential computational complexity of the search.

- **Pilot sequences:** the basic idea is to embed a known sequence in the host (in addition to the watermark message) that can be used at the decoder to estimate the desynchronization parameters from the received data [73, 96–100]. If the dimensionality of ϑ is small relative to the dimensionality of the host, reliable estimates can be obtained using the method of maximum likelihood or some other consistent estimators. The main weakness of this approach is that also the pilot sequence can be subjected to custom attacks aimed at removing it. Moreover the pilot insertion will unavoidably modify the host signal without conveying information, so less power is available for the information-bearing signals. On the other hand, it has been proven that pilot-based schemes are theoretically suboptimal [101].
- **Embedding in invariant domain:** according to this approach, the embedding and the decoding are performed both on the host signals transformed in a domain which is proven to be invariant to such distortions [82, 102–107]. Consequently, the difficulty with these methods is to construct suitable invariants for the considered desynchronization attacks. Even if this has been done for operations such as scaling, translation, and rotation, extend this approach to more complex transformation is a problem still open.

While promising results have been achieved in limited settings, focusing on particular attacks and/or particular content type, the gap between theory and practice is still significant.

However, as it has been outlined in Section 2.5.1, the deployment of data hiding algorithms resilient to desynchronization attacks is quite simpler for SS-based methods, due to the adopted decoding strategy, allowing the development of systems invariant to complex transformations [82–84, 86, 87], as it will be also shown in Section 2.6. Obviously, using SS-based techniques, the side information at the transmitter is not exploited, renouncing to the advantages in terms of embedding distortion and achievable capacity.

Consequently, much research is going into the design of QIM-based algorithms that can survive desynchronization attacks. This effort is motivated by the fact that QIM is proven to be robust against AWGN channel and to have the high capacity capability, always preserving the host signal fidelity.

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

We present here an example of an SS-based watermarking method resilient to geometric distortions for still images with the aim to show that robustness even against complex transformations can be achieved using this embedding approach. The spreading sequence is embedded in a conveniently selected subset of coefficients from the image Fourier transform (FT). Will show here that, using a correlation-based detector, to correctly identify the inserted spreading sequence is sufficient to feed the detector with the received coefficients in the Fourier domain where the watermark has been embedded, even after desynchronization attacks.

Recalling the well known property of the 2D Fourier spectra, which states that the FT of a rotated image is the rotated version of the FT applied on the not-rotated image, the key idea is to properly define an embedding region in the Cartesian double transformed Fourier domain able to achieve rotation invariance avoiding the need of a log-polar mapping, unlike in techniques which are based on embedding in the Fourier-Mellin domain, such as [103] and [82]. As a consequence, the computational complexity is considerably reduced with respect to these methods. A single invariant direction is extracted from the image spectrum to synchronize the detector and the watermark.

The invariant direction is defined as the straight line, passing for $(f_x = 0, f_y = 0)$, along which the function $|I(f_x, f_y)|$ has its maximum cumulated value, denoting by $I(f_x, f_y)$ the Fourier transform of an image $i(x, y)$ (heretoeafter it is intended that the "zero" frequency location is $(f_x = 0, f_y = 0)$). Hence the invariant direction is uniquely identified by the angle θ_{inv} formed by the extracted line with a reference direction.

The placement of the invariant direction in the Fourier domain is motivated by two reasons. Since the embedding is performed in the Fourier domain, it is reasonable to extract a resynchronization feature in the same domain. Moreover, a watermarked image can undergo attacks modifying either the whole image or circumscribed part of it, hence the Fourier domain has the advantage that local modifications in the spatial domain are always spreaded.

The invariant direction is used as resynchronization feature, which enables the watermarking system to identify the same direction from every distorted $I'(f_x, f_y)$; in this way,

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

even if the image has been rotated by an angle α , the invariant direction ($\theta_{inv} + \alpha$) is extracted, so that the detector and the watermark signal will be always synchronized.

As it has been discussed in Section 2.5.2, the strategy to achieve robustness to desynchronization attacks for SS systems is to make the watermark nearly invariant against desynchronization attacks. Hence, recognizing exactly the embedding region at the decoder through the invariant direction, the correlation detector is expected to identify the correct watermark sequence.

Actually the problem is that the extraction of the invariant direction is performed both at the embedding and detection sides; between the two operations the cover image could have been modified by the channel. In order to make the extraction method as robust as possible, a pre-processing step is applied to the image both at embedding and decoding side to get a spectrum which is as less dependent as possible on these modifications. The pre-processing operation will be described in Section 2.6.1.

In the diagram depicted in Fig. 2.6 the processing chain to identify the invariant direction from on the pre-processed image is shown.

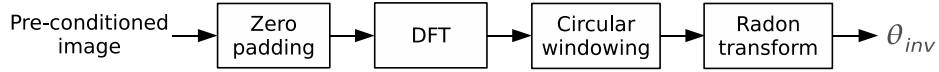


Figure 2.6: Block diagram of the direction extraction method.

The image is firstly checked to have the same dimensions; if $N \neq M$, where N and M denote the image dimensions, the shortest dimension is padded with zeroes to obtain a square matrix before the computing of the DFT.

Invariant direction extraction is performed using the Radon transform [108] on $I(f_x, f_y)$. According to the definition of invariant direction, the Radon transform is computed only on directions passing through the zero frequency:

$$R(0, \varphi)[I(f_x, f_y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(f_x, f_y) \delta(f_y - f_x \tan(\varphi)) df_x df_y \quad (2.15)$$

Actually a Discrete Radon Transform is computed on the image DFT, using a simple sum approximation. The Radon transform on a finite rectangular domain does not consider that, on different lines, different amounts of pixels lie and consequently diagonals are

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

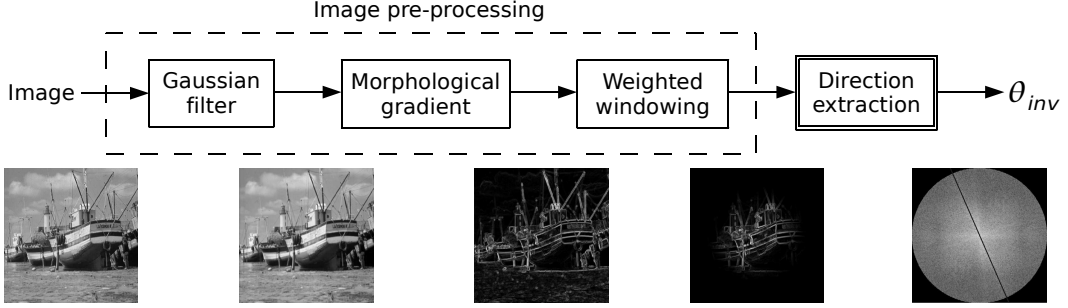


Figure 2.7: Block diagram of the pre-processing.

privileged directions, causing mistaken estimation. The circular windowing before the Discrete Radon Transform can prevent misleading results since along every line a fixed amount of non-zero terms lies [109].

Computing of Radon transform in a discrete context implies that a finite set of direction equally spaced between $[0, \pi[$ has to be defined as $\varphi_i = i \cdot \Delta\varphi$, $i = 0, 1, \dots, (\lfloor \pi/\Delta\varphi \rfloor - 1)$ and consequently eq. (2.15) is written as

$$R(0, i)[|I(f_x, f_y)|] = \Delta f_x \sum_{k=0}^{N-1} |I(k \Delta f_x, \tan(\varphi_i) k \Delta f_x)| \quad (2.16)$$

where Δf_x and $\Delta\varphi$ set the sampled lines along which the cumulated value of Radon transform are computed. The invariant direction angle θ_{inv} is then given by

$$\theta_{inv} = \arg \max_{\varphi_i \in [0, \pi[} \{R(0, i)[|I(f_x, f_y)|]\} \quad (2.17)$$

2.6.1 Image pre-processing

The pre-processing is needed to get a spectrum which is less dependent on the modifications that the image can undergo, aiming at providing robustness to invariant direction extraction method.

To get an image representation invariant to channel modifications, the edge feature is pointed out, since edges generally survive (even if distorted) to both geometric distortions and usual image processing. In Fig. 2.7 the pre-processing chain scheme is fully depicted.

The basic tool used for edge extraction is the cascade of a Gaussian low-pass filtering and a morphological gradient, i.e. a morphological operator revealing sharp luminance

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

transitions. This cascade is quite similar to a Laplacian of a Gaussian filtering, usually adopted for edge extraction, being the Laplacian operator here substituted by the morphological gradient operator [110]. Morphological operator has been here preferred to Laplacian one since the latter associates zero-crossing of second derivatives to an edge, whose location is as close as possible to the real edge [111], and it is typically followed by a thresholding to convert a gray-scale image into a binary edge image. In this application, rather than in an accurate location of edges with a binary image, we are interested in an invariant (and possibly robust against noise) feature extraction through edge enhancement; moreover we aim to get edges with thickness as invariant as possible to modifications on image. For this purpose a Gaussian low-pass filtering together with a suitable morphological gradient operator, able to enhance and to widen the edges, has been proven to fit better.

The circular symmetric windowing is performed on the enhanced edge image to reduce aliasing effects in the discrete Fourier domain. In fact the FT and the rotation are operations that do not exactly commute in a discrete domain [112], so we do not expect a perfect identity between the rotated version of the DFT of the image and the DFT computed from the the rotated image. Hence, aliasing effects can be just reduced smoothing the boundaries with a circular and symmetric window [112].

2.6.2 Embedding and detection

The spreading sequence is inserted into a sorted subset of the middle frequency coefficients of the doubly transformed domain, obtained as the intersection of a circular crown region with a finite subset of straight lines belonging to a sheaf:

$$(f_x, f_y) t.c. \begin{cases} R_{min} \leq (f_x^2 + f_y^2) \leq R_{max} \\ f_y = \tan(\theta_{inv} + i \cdot \Delta\theta) \cdot f_x \\ i = 0, \dots, (\lfloor \pi/\Delta\theta \rfloor - 1) \end{cases} \quad (2.18)$$

where θ_{inv} is the invariant direction extracted in (2.17). In the embedding process it is essential to keep the complex conjugate symmetry of the FT of the image, since it is a real signal. Therefore the same spreading sequence is properly embedded both in the real part, keeping the even symmetry, and in the imaginary part, keeping the odd symmetry. The whole embedding scheme is depicted in Fig. 2.8.

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

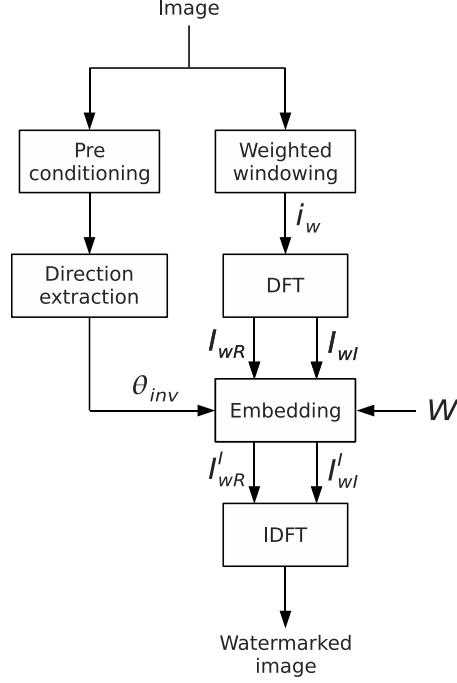


Figure 2.8: Embedder block scheme.

It is worth remarking here that the detector and the embedder can be assumed synchronized until they extract the same invariant direction, indexed by θ_{inv} , so that the straight lines belonging to the sheaf are considered in an ordered sequence starting with the same one.

Also, we remark that the angle step size $\Delta\theta$ between two subsequent lines belonging to the sheaf is not related with the angle step size used by Discrete Radon Transform $\Delta\varphi$ in direction extraction. In fact, while the latter affects uniquely the accuracy of the Radon Transform, the former sets the number of lines which compose the embedding region and consequently the length of the message that can be accommodated.

Denoting by \mathbf{x} the whole host sample sequence, which consists of the DFT coefficients corresponding to the frequencies described by eq. (2.18), and being $\mathbf{s} = \{s_1, s_2, \dots, s_L\}$ a pseudo-random spreading sequence where $S \sim \mathcal{N}(0, 1)$, according to [113], the multiplicative embedding rule is given by

$$y_k = x_k + g s_k^{(\mathbf{m})} |x_k| \quad (2.19)$$

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

where g is a constant gain factor modulating the embedding strength. However this is not the optimal choice. In fact usually in image spectra the energy is not uniformly distributed: the magnitude of low frequencies is usually much larger than the magnitude of middle and high ones. On the other hand the Human Visual System (HVS) is more sensitive to the low frequencies than the higher ones [36]. Thus, using a multiplicative embedding rule with a constant gain, the distortions undergone by low frequency coefficients are considerably much larger than higher frequencies variations, decreasing the perceptual quality of the watermarked image as demonstrated in Section 2.6.3.2. In order to bound this effect, it is reasonable to use a variable gain factor g_k , changing its value along every insertion direction in inverse relation to frequency. For this purpose a very simple linear law is proposed:

$$g_k = g \cdot \left(\frac{1}{4} + \frac{3}{4} \cdot \frac{k}{L} \right) \quad (2.20)$$

where $k = 1, 2, \dots, L$ spans the samples on every line of the embedding region. However more complicated functions can be thought to improve the visual masking effect without losing a sufficient level of robustness against attacks.

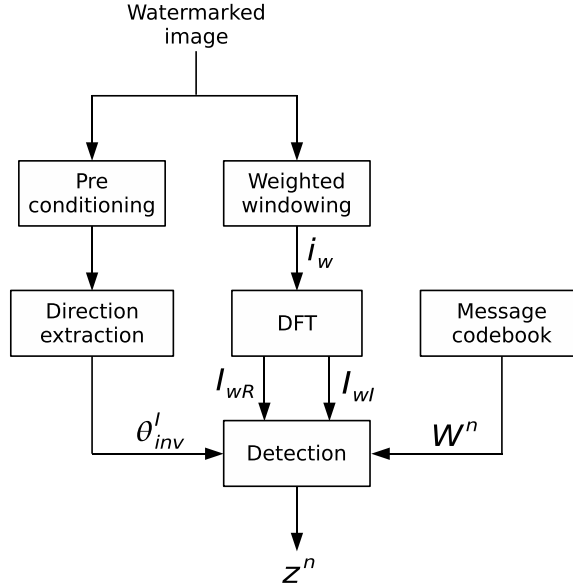


Figure 2.9: An intuitive insight of DM embedding.

The detector, which is depicted in Fig. 2.9 is built so that, given a possibly attacked image in input, once the invariant direction has been extracted, it can retrieve the marked

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

sample sequence from DFT domain. Then the detector verifies what message has been inserted computing a correlation coefficient between the marked, and possibly corrupted, sample sequence extracted from the image and every codeword belonging to a shared watermark codebook, known both at embedder and at detector sides. Being C the number of codewords in the spreading sequence codebook \mathbf{C} and being M the length of each sequence, the estimated one is given by

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s} \in \mathbf{C}} \left(\frac{1}{M} \sum_{k=0}^{M-1} z_k s_k^{(\mathbf{m})} \right) \quad (2.21)$$

2.6.3 Results

Herein some experimental results are shown, obtained processing standard images, whose size is typically 512x512 pixels, and embedding a mark sequence of length 9800 samples along 70 directions spaced by $\Delta\theta \approx 2.57 \text{ deg}$ into the DFT domain.

2.6.3.1 Effectiveness of direction extraction

The effectiveness of the direction extraction method is a necessary condition for the decoder and the mark to be synchronized. The following results were obtained computing the Radon transform of the windowed DFT of a pre-processed image along a finite set of directions equally spaced with a step $\Delta\varphi = 0.5 \text{ deg}$ and choosing the direction with the maximum cumulated value, as described in eq. (2.17).

In Figs. 2.10(a)–2.10(d) the right working of the proposed direction extraction method is exhibited (for an original picture of Boat image see Fig. 2.7). In Fig. 2.10(a) the pre-processed Boat image is depicted while in Fig. 2.10(b) its windowed spectrum is shown and the extracted invariant direction is highlighted. Rotating by an angle of 30 deg the Boat image, we had the pre-processed image and the windowed spectrum depicted respectively in Figs. 2.10(c) and 2.10(d). Comparing the highlighted directions in Figs. 2.10(b) and 2.10(d), the rotation of the invariant direction according to the rotation of the original image is noticeable.

In order to measure the effectiveness of the direction extraction method, we compared the extracted invariant directions before and after embedding and attacks. We tested the algorithm against both common signal processing techniques (Additive White Gaussian Noise, JPEG compression and smoothing) and geometric distortions (rotation, scaling

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

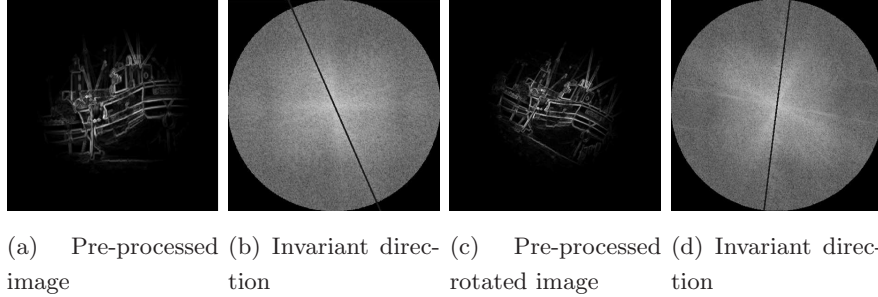


Figure 2.10: Example of pre-processing on Boat image with extraction of invariant direction for both original and rotated cases.

and cropping). Cropping attack was performed cutting away both symmetrically and asymmetrically the framing part of the marked image, but preserving the image center (where the most relevant information is supposed to be located).

The experimental results are listed in Table 2.1, where $\Delta\theta_{inv} = \theta'_{inv} - (\theta_{inv} + \alpha) \bmod \pi$ is the difference of extracted directions at the detection and embedding sides. The angle values are expressed in *deg*, while the results for cropping attacks are indexed by the percentage of pixels constituting the cropped image with respect to the original one.

Moreover, the direction extraction method performs well against random line removal attack too (experimental results are here not shown).

2.6.3.2 Imperceptibility

To assess the embedding distortion, the fidelity of the watermarked image has been evaluated. For this purpose in watermarking literature the mean square error based metrics are widely used, even if these metrics does not exploit the properties of the HVS and they can give measures that depart from the perceived quality, as it is shown in [114] and as it has been discussed in Section 2.2. In order to assess the imperceptibility of the watermarking method, between the various existing quality assessment methods based on known characteristics of the HVS, the Watson's distance [115] is here used and its results are compared with DWR.

The introduced distortion has been evaluated performing the embedding in the standard images with 1000 different watermarks and computing the average values of the adopted measures. Moreover, to verify the enhancement in the perceptual quality resulting from the use of a variable gain factor, the embedding distortion has been established

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

Table 2.1: Invariant directions extracted from attacked images.

Image	Lena	Peppers	Boat	
Invariant direction	145	173.5	113.5	
Attack	$\Delta\theta_{inv}$	$\Delta\theta_{inv}$	$\Delta\theta_{inv}$	
AWGN	$\sigma = 10$	-0.5	0	0
	$\sigma = 15$	-0.5	0	0
	$\sigma = 20$	0	0	0
JPEG	quality = 50	-0.5	0	0
	quality = 30	0	0	0
	quality = 20	0	0	0
	quality = 10	0	0	0
Smoothing	Average	0	0	0
	Gaussian	0	0	0
	Median	0	0	0
Rotation	$\alpha = 1$	0	0	0
	$\alpha = 5$	-0.5	-0.5	0
	$\alpha = 45$	-0.5	-1	-0.5
	$\alpha = 60$	-0.5	-0.5	-0.5
	$\alpha = 90$	-0.5	0	0.5
	$\alpha = 120$	-0.5	-0.5	0
	$\alpha = 160$	0	-0.5	-0.5
	$\alpha = -1$	0	0	0
	$\alpha = -5$	-0.5	-0.5	0
	$\alpha = -45$	-0.5	-1	-0.5
	$\alpha = -60$	-0.5	-0.5	0
	$\alpha = -90$	-0.5	0	0.5
	$\alpha = -120$	-0.5	-0.5	-0.5
$\alpha = -160$	0	1	0	
Scaling	scale = 0.5	1.5	0.5	0
	scale = .75	1.5	0	0
	scale = 1.5	0	0	0.5
	scale = 2	0	0.5	0
Cropping	rem% = 85	1.5	0	0
	rem% = 78	1.5	0	0
	rem% = 65	-2.5	0	0
	rem% = 53	-2.5	0	0

2.6 Spread Spectrum based watermarking resilient to geometric distortions for still images

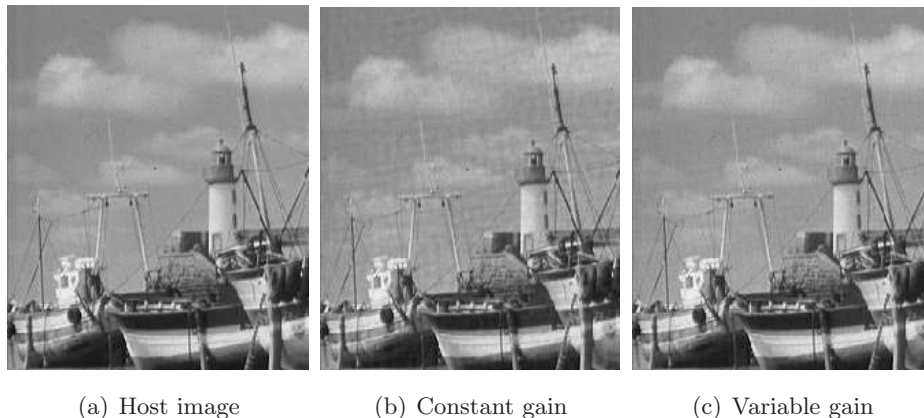


Figure 2.11: Comparison of a detail of the standard image Boat marked using constant gain and variable gain.

in case of both a constant gain and of a variable gain. To have a fair comparison in terms of mean square error, we have set the constant gain $g_k = \sqrt{2}/2$, while in the variable gain function (2.20) we have set $g = 1$. In this way, under the unreal hypothesis that the image spectrum is white, the two different gain functions would produce watermark signals having the same power.

Table 2.2: Comparison of DWRs and Watson’s distances for constant gain and variable gain.

Image		Lena	Peppers	Boat
Constant gain	DWR	34.13	32.25	36.42
	D_{wat}	24.15	45.94	26.02
Variable gain	DWR	38.35	35.91	40.94
	D_{wat}	12.82	25.32	13.36

From Table 2.2, where the measured embedding distortions are presented, the quality enhancement deriving from the use of a variable gain function is confirmed. In fact the Watson’s distance is considerably higher in case of constant gain for all the standard images. As a confirmation, in Fig. 2.6.3.2 it is shown a detail of the standard image Boat which reveals as the visible artifacts, that are evident in case of constant gain, are masked using a variable gain. Hence in the forthcoming results, it is implied the use of the variable gain function.

2.6.3.3 Robustness

Afterwards the robustness of the whole watermarking system has been checked; for this purpose some gray scale standard images were watermarked and both signal processing and geometric attacks were applied to these images. The detector output reveals the estimated embedded message as the one having the greatest correlation coefficient with the sample sequence extracted from the marked and possibly corrupted image. Thus, from the detector response to all the codewords belonging to the codebook, we measured the first-to-second peak ratios (P_1/P_2) in decibel.

In the experimental results, listed in Table 2.3, the index \hat{n} represents the codeword retrieved from the attacked image by comparison of the correlation coefficients. These results were obtained processing the standard image Boat, but similar results have been obtained with other standard images. Here the results for the cropping attack are related only to symmetrical cuts of the framing part of the marked image, since for asymmetrical cropping the watermark retrieving have shown to fail.

By inspection of the results listed in Table 2.3 it is verified the robustness of the proposed scheme against almost all the tested attacks. In particular the correct mark retrieving is guaranteed even if a slight error in extracted direction occurs. Hence, even if a rough synchronization is guaranteed by the invariant direction, the detector is still robust to complex geometric distortions, i.e. rotations, due to the intrinsic robustness of correlation detector, as it has been discussed in Section 2.5.2. On the other hand, even if the invariant direction is exactly retrieved, geometric distortions and the consequent interpolation in the pixel domain considerably modify the image spectrum, demonstrating that robustness to desynchronization transformations relies also on the detector.

2.7 Concluding remarks

In this chapter we have presented the information hiding as a classical communications problem, showing as digital communications played a fundamental role in the development of watermarking technology inspiring the most relevant embedding techniques and suggesting the approach to perform rigorous performance analyses. This has been widely discussed for spread spectrum and quantization based data hiding methods, comparing their characteristics and the theoretical achievable capacity.

Table 2.3: Detection results on the attacked image Boat (DWR ≈ 40.94 and $D_{wat} \approx 13.36$).

Embedded codeword index		$n = 259$	
		Correlation detector	
Attack		\hat{n}	$P_1/P_2(dB)$
AWGN	$\sigma = 10$	259	3.99
	$\sigma = 15$	259	4.61
	$\sigma = 20$	259	4.38
JPEG	QF= 50	259	3.64
	QF= 30	259	3.38
	QF= 20	259	2.61
Smoothing	Average	259	3.35
	Gaussian	259	3.55
	Median	259	3.15
Rotation	$\alpha = 1$	259	3.73
	$\alpha = 5$	259	3.82
	$\alpha = 45$	259	3.69
	$\alpha = 60$	259	2.22
	$\alpha = 90$	259	3.69
	$\alpha = 120$	259	3.75
	$\alpha = 160$	259	1.98
	$\alpha = -1$	259	3.83
	$\alpha = -5$	259	3.56
	$\alpha = -45$	259	4.05
	$\alpha = -60$	259	4.66
	$\alpha = -90$	259	3.69
	$\alpha = -120$	259	4.18
$\alpha = -160$	259	4.26	
Scaling	scale = 0.5	Fail	-
	scale = .75	Fail	-
	scale = 1.5	Fail	-
	scale = 2	Fail	-
Cropping	rem% = 85	259	4.50
	rem% = 78	259	4.08
	rem% = 65	259	3.35
	rem% = 53	259	2.95

Then the class of desynchronization attacks has been introduced, the robustness of basic decoding strategy against these attacks has been discussed and the main strategies to cope with them have been outlined.

This chapter ends with the description of a proposed spread spectrum watermarking method for images robust to some geometric distortions. The experimental results show how, even if the marked image undergoes complex desynchronization attacks, basic SS detector is able to retrieve the embedded information. The ease with which SS based data hiding methods guarantees robustness against desynchronization attacks is even more evident in comparison with the vulnerability of quantization based. The existing gap in these scenarios between different embedding approaches stimulates research activity in the field of quantization based methods, developing new strategies to provide them invariance to such attacks.

3

Quantization-based embedding and scaling vulnerability

In this chapter, a deep description of basic quantization-based data hiding methods is given and the problem of vulnerability to volumetric distortions is faced. We start from the description of Dither Modulation (DM), which is a scalar and binary QIM-based algorithm, since its theoretical analysis allows to identify the achievable performance of quantization-based methods against additive noise. However, even if these embedding techniques exhibit high robustness to additive white Gaussian noise (AWGN) channel and high capacity capability while preserving the host signal fidelity, their main weakness is the extreme sensitivity to volumetric distortions, such as gain scaling. Since very common processings, such as brightness change or loudness change, are modeled by gain scaling of the host signal, the development of data hiding schemes able to cope with this attack has been an interesting challenge for the research community and at the moment the gain attack can be considered somewhat solved.

Among the various methods that have been proposed, a deep description is given of Rational Dither Modulation (RDM). The interest in this algorithm is motivated by the fact that here theoretical invariance to constant gain scaling is achieved preserving the fundamental properties of DM. Also, it is proven that RDM is able to asymptotically achieve the performance of DM, so that the same error probability can be reached under the same conditions. This behavior reveals the reason why RDM has been exploited in several subsequent algorithms and it can be adopted as basic component to develop data hiding methods to cope with more complex distortions.

This chapter is organized as follows: in Section 3.1 dither modulation is presented, while in Section 3.2 the problem of the vulnerability to gain scaling is discussed and an overview of the different approaches to cope with this attack is given. Then, rational dither modulation is described in Section 3.3 and finally some considerations are drawn in Section 3.4.

3.1 Dither modulation

To introduce the practical methods belonging to the wide class of QIM-based [67], whose principles have been outlined in Section 2.4, the basic approaches for the binary unidimensional case will be firstly presented. We will focus on the binary dither modulation (DM) [116], in which it is assumed that only one binary digit of information $m_k \in \{-1, +1\}$ per host sample is hidden by the adopted scalar quantizer. Hence the basic procedure is to quantize the current host sample x_k using one of two quantizers, depending on the binary value to be embedded. The interest received by DM is due to the fact that it can be easily analyzed, clearly identifying the achievable data rate against additive noise. Moreover the same kind of analysis developed for DM can be extended to other quantization approaches.

DM is essentially based on the definition of an unidimensional shifted lattices

$$\Lambda_m = 2\Delta\mathbb{Z} - m\frac{\Delta}{2}, \quad m = -1, 1 \quad (3.1)$$

which describes the centroids for a couple of Euclidean distance based quantizer $Q_m(\cdot)$, being Δ a fixed quantization step size. Given the k th symbol $m_k \in \{-1, 1\}$ to be hidden, the indexed dithered quantizer is used to perform the embedding into the k th sample, so that the watermarked sample is obtained using the rule

$$y_k = Q_{m_k}(x_k) = 2\Delta \left\lfloor \frac{x_k - m_k\Delta/2}{2\Delta} \right\rfloor + m_k\frac{\Delta}{2} \quad (3.2)$$

where $\lfloor \cdot \rfloor$ denotes rounding to the nearest integer. Thus, the watermarked sample y_k turns out to be the quantization centroid closest to the corresponding host sample x_k and the watermark signal is just the quantization error

$$w_k = y_k - x_k = Q_{m_k}(x_k) - x_k \quad (3.3)$$

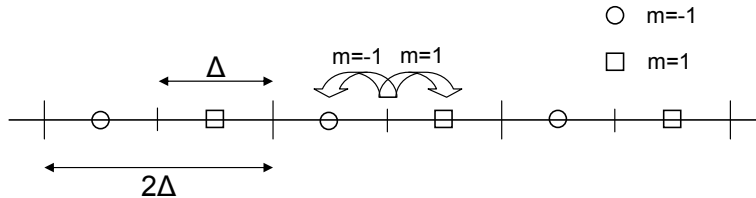


Figure 3.1: An intuitive insight of DM embedding.

According to the shifted lattices construction, Fig. 3.1 gives an intuitive insight of DM embedding: the sum of the k th dither sample $m_k \Delta/2$ and of the host sample x_k is quantized with step-size 2Δ , then the quantized value is perturbed to move the watermarked sample y_k to the closer quantization centroid related with the symbol m_k to embed. The quantization step-size Δ is usually chosen to produce the desired embedding distortion on the watermarked signal in term of mean square error.

Assuming that the quantization bins are small and the host pdf $f_x(x)$ is smooth enough, leading to the validity of Schuchman’s condition [117], the host pdf can be assumed flat and the watermark can be also considered to have a pdf that is roughly a constant within each individual bin. Hence, the host pdf in each quantization bin $x \in (2\Delta i, 2\Delta i + 2\Delta)$ $i \in \mathbb{Z}$ can be assumed constant and equal to $f_X(x) \approx f_i$. Consequently, the watermark samples W_k can be assumed to be uniformly distributed in $(-\Delta, \Delta)$, so that the embedding distortion for a given cell corresponds by definition to the variance of this uniformly distributed random variable, which is equal to $\Delta^2/3$. As both quantizers $Q_{m_k}(\cdot)$ are uniform, the average embedding distortion will be

$$D_w = \frac{1}{M} \sum_{k=1}^M E[W_k] = \frac{\Delta^2}{3} \tag{3.4}$$

At the decoding side, having observed the possibly attacked sample z_k , the estimated k th bit \hat{m}_k is retrieved quantizing the received sample with a quantizer with fixed step-size Δ . This is equivalent to decide \hat{m}_k according to a minimum Euclidean distance rule

$$\hat{m}_k = \arg \min_{-1,1} |z_k - Q_{m_k}(z_k)| \tag{3.5}$$

The decoding error probability can be easily defined for such a decoding rule and an additive random noise on the channel, whose pdf is denoted by $f_N(n)$. Being the received

3.1 Dither modulation

samples $z_k = y_k + n_k$ and denoting respectively by \mathcal{R}_{-1} and \mathcal{R}_1 the decision regions associated with $m_k = -1$ and $m_k = 1$, the error probability is given by

$$\begin{aligned} P_e &= P\{|z_k - Q_1(z_k)| < |z_k - Q_{-1}(z_k)| \mid m_k = -1\} = \\ &= P\{z_k \in \mathcal{R}_1 \mid m_k = -1\} \end{aligned} \quad (3.6)$$

Recalling that $f_X(x)$ is assumed constant in each quantization bin, for high DWRs the random variables Y_k are expected to have a flat pdf in each quantization bin. Thus, given the periodicity in \mathcal{R}_1 , the error probability can be reasonably assumed independent of the particular quantization bin of $Q_{-1}(\cdot)$ in which the sample lies. Consequently, to obtain an integral expression of P_e , it is sufficient to compute P_e by conditioning x_k to lie in a given bin, which is equivalent to know the watermarked sample $y_k = Q_{-1}(x_k)$. Eventually, being by definition $N_k = Z_k - Y_k$, the error probability can be computed as

$$\begin{aligned} P_e &= \int_{\mathcal{R}_1} f_Z(z \mid b = -1; Q_{-1}(x_k)) = \\ &= \int_{\mathcal{R}_1} f_N(n) dn = \sum_{i=-\infty}^{\infty} \int_{(4i+1)\Delta/2}^{(4i+3)\Delta/2} f_N(n) dn \end{aligned} \quad (3.7)$$

In case of the additive random noise on the channel is a white Gaussian noise with variance σ_n^2 , the decoding error probability of DM can be analytically computed and is given by

$$\begin{aligned} P_e &= \sum_{i=-\infty}^{\infty} \int_{(4i+1)\Delta/2}^{(4i+3)\Delta/2} \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left(-\frac{n^2}{2\sigma_n^2}\right) dn = \\ &= \sum_{i=-\infty}^{\infty} \left[Q\left(\frac{(4i+1)\Delta}{2\sigma_n}\right) - Q\left(\frac{(4i+3)\Delta}{2\sigma_n}\right) \right] \end{aligned} \quad (3.8)$$

As it is stated in [67], QIM, and consequently DM too, has been defined as a "provably good" method. This is motivated by the fact that, for certain amplitude bounded attacks, the decoding error probability is zero. As an example, in the case of N_k have a $f_N(n) = 0$ for $|n| > \Delta/2$, no decoding errors occur.

To provide security to DM, a secret dither sequence can be added to randomize the shifted lattices giving the quantization centroids [73]. Being d_k a random sequence drawn

from a random variable having value in the range $[0, 1)$, randomized shifted lattices are defined by

$$\Lambda_m = 2\Delta (\mathbb{Z} + d) - m\frac{\Delta}{2}, \quad m = -1, 1 \quad (3.9)$$

without modifying any properties of the DM embedding method. On the other hand, only the knowledge of the secret dither sequence is supposed to allow the correct decoding of the hidden information, which relies on the knowledge of the quantization centroids. For more details on security issues for quantization based algorithms, refer to [118].

The above described algorithm is the simplest within the class QIM-based methods, but it can be easily extended to more general construction. The distortion compensation capability can be added to standard DM giving rise to DC-DM [67], aka scalar Costa scheme [73], whose embedding rule is given by

$$y_k = x_k + \alpha (Q_{m_k}(x_k) - x_k) = x_k + w_k \quad (3.10)$$

where the distortion compensation $\alpha \in [0, 1]$ is a parameter to be optimized. In this way, a scalar implementation of QIM, which has been outlined in Section 2.4, is obtained. It is worth noting that for $\alpha = 1$ DC-DM reduces to DM. In DC-DM with $\alpha \in [0, 1]$, it can be demonstrated that the watermark signal is uniformly distributed in $(-\alpha\Delta, \alpha\Delta)$, so that the embedding distortion is equal to $D_w = \alpha^2\Delta^2/3$ and the error probability can be analytically derived [119].

Also, DM can be extended to the vector case, as it was proposed by Chen and Wornell [67], by replacing the scalar quantizer of DM with a convenient vector quantizer. The more general construction of quantization-based embedding, which encompasses most of the proposed QIM formulations and it is usually referred to as "lattice data hiding", is based on L -dimensional nested lattice codes [39]. Other related family of quantization-based methods is Spread Transform - Dither Modulation (ST-DM) [67, 73, 119], which combines certain characteristics of spread spectrum and QIM methods, although it is not specifically addressed in this thesis.

3.2 Gain attack and countermeasures

From the beginning of the research on quantization-based data hiding it was recognized the extreme sensitivity of such methods to valumetric distortions [73], which are encompassed in the wide family of desynchronization attacks presented in Section 2.5. A valumetric

distortion can be defined as a generic function $f(\cdot)$ applied pointwise to all of the host samples modifying their original values [120]. The distortion function can be either linear or nonlinear.

Lattice-based schemes are vulnerable to amplitude varying attacks because these attacks make the encoder and decoder lattice volumes no more matching. In other words, modifying the amplitude of the watermarked samples is equivalent to modify accordingly the quantization bins used by the embedder, but since the quantization bins at the decoder are fixed, an amplitude desynchronization between embedder and decoder is produced that dramatically increases the BER [78]. Referring to basic DM, it is reasonably expected that, if $y_k - f(y_k) > \Delta/2$, an error will probably occur for the k th sample.

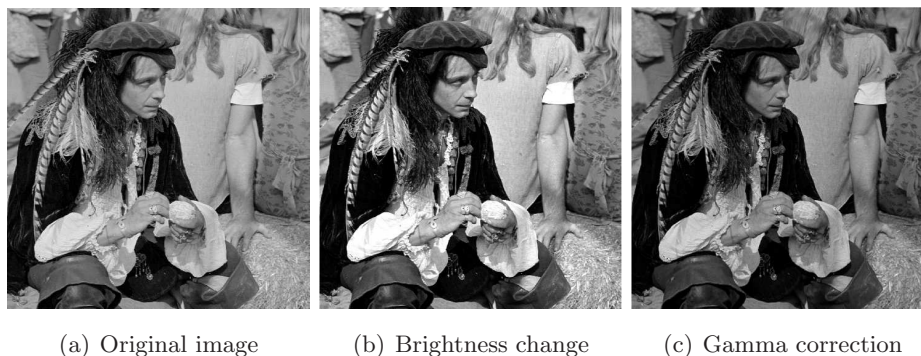


Figure 3.2: Illustration of valumetric distortions on standard image Man.

The relevance of valumetric distortions in data hiding relies on the fact that a variety of processing that are commonly applied to image, video and audio signals lie on pointwise amplitude scaling. For instance brightness and contrast change for images and video are a gain scaling of the host samples amplitude as well as loudness change in audio. On the other hand these kinds of distortions usually do not introduce significant fidelity degradation or even they can improve the perceived pleasantness of the signal. An example of the perceived quality of an image after brightness change and gamma correction is shown in Fig. 3.2, where it can be appreciated the limited perceptual degradation that is introduced by valumetric distortions.

It is worth noting that non-additive attacks are not efficiently taken into account by MSE-based measures, such as the attack strength as it has been defined in Section 2.2. In fact, even if significant variations of the host samples amplitude may produce very large MSE values, the resulting perceptual modifications cannot be significant. As an

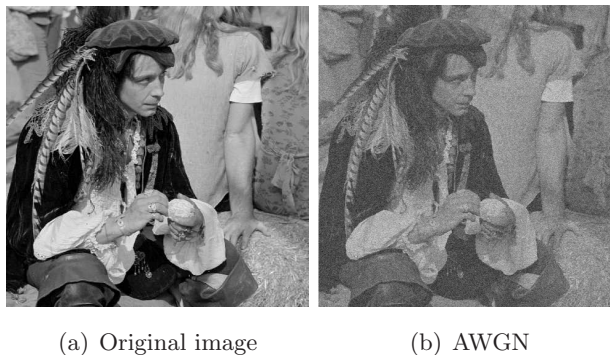


Figure 3.3: Illustration of white Gaussian noise addition on standard image Man.

example, it can be noticed that, in spite of the fidelity of Fig. 3.2(b) and Fig. 3.2(c), the measured DNRs are equal to 10.5 dB and 12 dB, respectively. On the other hand MSE-based measures are well-fitted to additive noise as it can be seen in Fig. 3.3. Here the Man image is corrupted by white Gaussian noise addition to have DNR=11 dB and the resulting perceptual quality is dramatically downgraded. Thus, in a similar framework, the valumetric distortion is often considered as imperceptible and then it is reasonable to disregard it when the effects of noise addition are evaluated using a MSE-based measure.

The simpler and more studied valumetric distortion is the the gain attack, consisting of the multiplication of the host feature sequence by a gain factor ρ , which is unknown at the decoder. If ρ is constant throughout the watermarked signal, the so called *Fixed Gain Attack* (FGA) is obtained, as illustrated in Fig. 3.4. On the channel also a zero-mean additive white noise \mathbf{N} with variance σ_n^2 and independent of the watermarked signal is considered to compare gain invariant schemes with other DM-based methods. The attacked signal is then given by

$$\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N}) \tag{3.11}$$

where it is assumed $\rho > 0$. According to the considerations drawn above, the valumetric distortion is considered as imperceptible and as a consequence the attacking distortion has to be made independent of ρ , so that the attacking strength D_c is defined as

$$\begin{aligned}
D_c &= \frac{1}{M} \sum_{k=1}^M E \left[|\rho^{-1} Z_k - Y_k|^2 \right] = \\
&= \frac{1}{M} \sum_{k=1}^M E [N_k^2] = \sigma_n^2
\end{aligned} \tag{3.12}$$

For sake of completeness, we remark here that in literature the FGA channel has been also defined as $\mathbf{Z} = \rho \mathbf{Y} + \mathbf{N}$ [73].

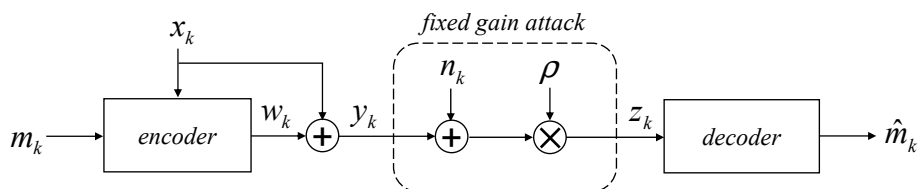


Figure 3.4: Fixed gain attack channel.

Motivated by the relevant drawback represented by the vulnerability of quantization-based techniques against gain scaling, a great effort has been spent in the last years by the data hiding community to cope with this attack and at the moment it can be considered somewhat solved. The various techniques that have been proposed can be roughly classified into three categories: a class of watermarking systems tries to estimate the amplitude scaling factor using or auxiliary pilot signals [121] or blind estimation methods [90,92,122]. In the former approach, the scaled quantization step-size is estimated at the decoder analyzing the histograms of the received pilot samples; in the latter approach the amplitude scaling factor is blindly estimated from the pdf of the received, and possibly attacked, host. Another possible solution is the adoption of codes having the codewords placed on a high-dimensional hypersphere [74, 123], which were proposed as a different way to realize the side informed paradigm and which are invariant to gain scaling by construction. The latter class encompasses algorithms performing the embedding in a convenient gain invariant domain. In the method proposed in [106], which will be briefly described in Section 3.2.1, a possible gain invariant domain has been recognized in the angle formed by a vector of host signal samples in a hyperspherical coordinates system. Comesaña et al. have lately proposed in [107] to perform standard DM embedding in the logarithmic domain. Here, a scaling resistant scheme is proposed by embedding the watermark into the difference between two successive samples in the logarithmic domain. A brief overview

of DM in the logarithmic domain will be given in Section 3.2.2. Belonging to this class can be considered the methods proposed by Oostveen et al. [124] and Cox et al. [114] for image watermarking, in which the quantization step size is scaled by a factor computed according to a perceptual model. In [114] the Watson’s model [115] is modified so that the slacks, which determine how much each DCT coefficient can be altered, scale linearly with valumetric gain. The slacks are then used to adaptively adjust the quantization step sizes for QIM-based embedding into the DCT coefficients, achieving gain scaling robustness in conjunction with high perceptual quality of the marked image. Finally, Rational Dither Modulation (RDM) has been proposed in [104] providing a simple and effective gain-invariant method. The relevance of RDM lies on the fact that it retains the main properties of DM and, under some conditions, it has been proven to asymptotically achieve the same performance of DM.

3.2.1 Angular quantization index modulation

In communications, a relevant performance improvement is obtained modulating the carrier’s phase (FM) instead of amplitude (AM). Similarly, phase modulation schemes can be used in data hiding to improve the performance in case of gain scaling on the channel.

In Angular Quantization Index Modulation (AQIM) the hidden information is embedded by quantization of the angle formed by the host signal vector with the origin of a hyperspherical coordinate system. In the 2-dimensional case, given two successive samples of the host signal \mathbf{x} , they can be mapped in polar coordinates representation (r, θ) , that are respectively computed as

$$\theta_k = \arctan\left(\frac{x_k}{x_{k-1}}\right) \tag{3.13}$$

$$r_k = \sqrt{x_k^2 + x_{k-1}^2} \tag{3.14}$$

Then the DM embedding is applied to the angle θ_k obtaining the quantized value θ_k^Q , which is given by

$$\theta_k^Q = 2\Delta_\theta \left\lfloor \frac{\theta_k - m_k \Delta_\theta / 2}{2\Delta_\theta} \right\rfloor + m_k \frac{\Delta_\theta}{2} \tag{3.15}$$

with $\Delta_\theta = \pi/M$. The factor M has to be chosen as a trade-off between the embedding distortion and the robustness to additive noise. Since the radius coordinate remains

unchanged in the embedding process, the watermarked samples in the host domain are obtained by inverting eqs. (3.13) and (3.14), so that $y_k = r_k \sin(\theta_k^Q)$ and $y_{k-1} = r_k \cos(\theta_k^Q)$.

When the whole watermarked signal \mathbf{y} is scaled by the same gain ρ , the 2-dimensional angular coordinates does not change at all, so that DM embedding is proven to be theoretically invariant to a pure gain attack. In fact at the decoder, the hidden information is retrieved through DM decoding from

$$\begin{aligned} \theta'_k &= \arctan\left(\frac{z_k}{z_{k-1}}\right) = \\ &= \arctan\left(\frac{\rho(y_k + n_k)}{\rho(y_{k-1} + n_{k-1})}\right) = \\ &= \arctan\left(\frac{y_k + n_k}{y_{k-1} + n_{k-1}}\right) \end{aligned} \tag{3.16}$$

which is perfectly equal to θ_k^Q in case of no additive noise on the channel.

In [106] the AQIM embedding method is described also in the generic L -dimensional case, in which $L - 1$ bits are embedded in the angles formed by a length- L host samples vector $(x_k, x_{k-1}, \dots, x_{k-L+1})$. Moreover in [125] the decoding error probability for AQIM under AWGN has been analytically derived, while in [126] the data to watermark ratio has been defined as a function of the angle quantization step-size

$$DWR \approx \frac{\Delta_\theta}{2(L-1)(\Delta_\theta - \sin(\Delta_\theta))} \tag{3.17}$$

In Fig. 3.5 the empirical BERs of AQIM evaluated using different values of M are presented for \mathbf{X} Gaussian distributed and additive white Gaussian noise as attack on the channel.

3.2.2 DM in the logarithmic domain

A brief overview of DM in the logarithmic domain [107] is here given. This is also motivated by the fact that the comparison of DM in the logarithmic with hyperbolic RDM, which will be presented in Section 4.3, will be very useful in the next chapter to clearly understand the achievable performance of hyperbolic RDM against additive noise. For sake of brevity, heretofore DM in the logarithmic domain will be referred as logarithmic DM.

In [107], the embedding is realized by quantizing the host signal in the logarithmic domain, which means that the embedding rule is given by

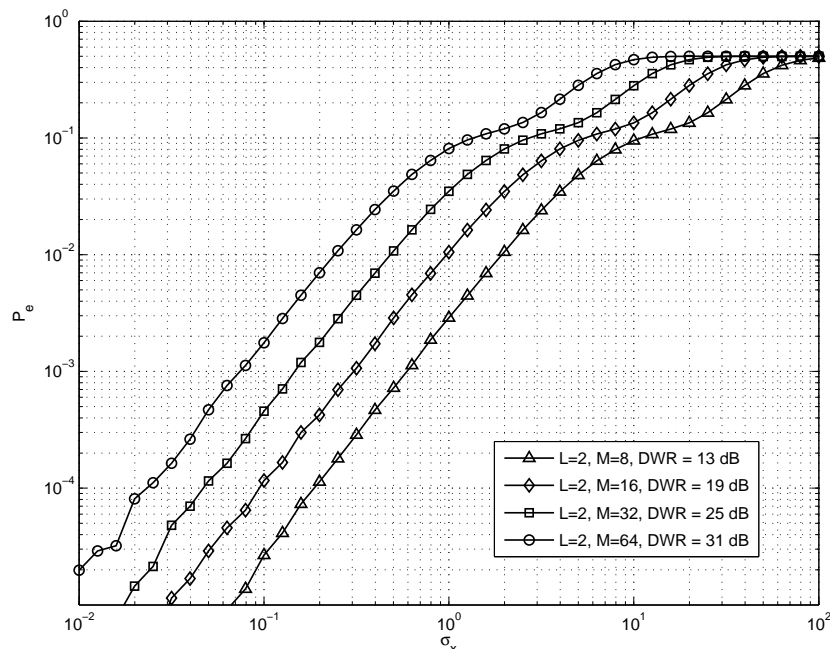


Figure 3.5: Empirical values of the error probability of AQIM as a function of σ_n for different values of M .

$$\log(|y_k|) = Q_{m_k}(\log(|x_k|)) \quad (3.18)$$

Exploiting the properties of logarithmic function, a gain invariant scheme is obtained quantizing the difference between the current host sample and the previous marked one, so that we have

$$\log(|y_k|) = Q_{m_k}(\log(|x_k|) - \log(|y_{k-1}|)) \quad (3.19)$$

which will be named differential logarithmic DM. In both cases the k th watermarked sample has to be mapped back to the original host domain; the inverse mapping function is defined as

$$y_k = \text{sign}(x_k) \exp(\log(|y_k|)) \quad (3.20)$$

The gain invariance property for the differential scheme is guaranteed by the properties of logarithm. In fact, at the decoder side, the DM decoding is applied to the difference

3.2 Gain attack and countermeasures

between the absolute value of two successive received samples, so that the possible gain scaling is canceled out by the logarithm:

$$\begin{aligned}
 & \log(|z_k|) - \log(|z_{k-1}|) = \\
 & = \log(|\rho(y_k + n_k)|) - \log(|\rho(y_{k-1} + n_{k-1})|) = \\
 & = \log(|(y_k + n_k)|) - \log(|(y_{k-1} + n_{k-1})|)
 \end{aligned} \tag{3.21}$$

As a consequence the decoding results intrinsically invariant to gain scaling on the channel.

A quite interesting property of both the differential and the non-differential logarithmic DM is that the embedding distortion is approximately equal to $D_w \approx \sigma_X^2 \Delta^2 / 12$, for any distribution of the original host and for $\Delta \ll 1$, which are reasonable values due to imperceptibility constraints. Thus, the data-to-watermark ratio is independent of the host signal power and is given by $DWR = 12/\Delta^2$.

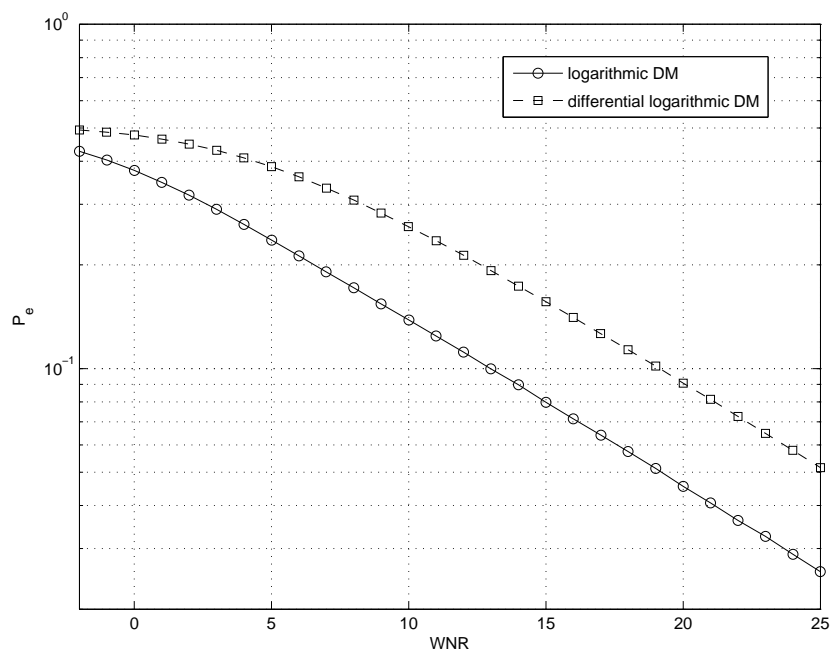


Figure 3.6: Empirical values of the error probability for differential and non-differential logarithmic DM (DWR=25 dB).

Also, in [107] the probability of decoding error has been analytically derived. In Fig. 3.6 the empirical BERs for logarithmic DM methods are shown for \mathbf{X} Gaussian distributed with $\sigma_x = 100$, DWR=25 dB and additive white Gaussian noise as attack on the channel.

Eventually, logarithmic DM has been proven to shape the watermark to be compliant with the Human Visual System (HVS). Due to logarithmic function, the equivalent quantization step-size in the original domain is scaled according to the magnitude of the host sample to be quantized, so that a larger watermark sample is accommodated in a larger host sample. In fact, denoting by \mathbf{v} the watermark signal in the logarithmic domain, we have that the magnitude of the k th watermark sample in the original domain is given by

$$\begin{aligned} w_k &= y_k - x_k = \\ &= \text{sign}(x_k) \exp(\log(|x_k| + v_k)) - x_k = \\ &= x_k (1 - \exp(v_k)) \end{aligned} \tag{3.22}$$

for large DWRs and then for $\Delta \ll 1$ we can approximate $1 - \exp(v_k) \approx -v_k$, so that we have

$$w_k = x_k (1 - \exp(v_k)) \approx x_k (-v_k) \tag{3.23}$$

From the perceptual point of view, this behavior is very appreciable since it reflects the contrast masking effect of HVS [36], which has been largely exploited in multiplicative spread spectrum watermarking. For more details on the perceptual masking property of logarithmic DM as well as some experimental results, refer to [107].

3.3 Rational Dither Modulation

Rational Dither Modulation (RDM) is based on the use of a gain-invariant adaptive quantization step size which is computed at the embedder and which can be blindly estimated at the decoder from the received, and possibly attacked, host signal. The key idea is to compute the adaptive step size from the L previous watermarked samples, which are available at decoding side, even if corrupted. The memory L of such a function is a crucial design parameter: small memory is useful when the target is to cope with varying gains, while large memory allows to reach high performance against additive noise, due to the more robust step-size estimation.

In the L th-order RDM, the k th information bit is embedded with a scalar quantizer into the k th host sample using the L previously watermarked samples; the step-size of

3.3 Rational Dither Modulation

the DM quantizer is scaled by $g(\mathbf{y}_{k-1})$, being the vector $\mathbf{y}_{k-1} = (y_{k-1}, y_{k-2}, \dots, y_{k-L})$, so that the k th watermarked sample is given by

$$y_k = g(\mathbf{y}_{k-1}) Q_{m_k} \left(\frac{x_k}{g(\mathbf{y}_{k-1})} \right) \quad (3.24)$$

where the function $g : \mathbb{R}^L \rightarrow \mathbb{R}$, $L \geq 1$, belongs to the set \mathcal{G} , whose elements have the following property:

$$g(\rho \mathbf{y}) = \rho g(\mathbf{y}) \quad \forall \rho > 0, \mathbf{y} \in \mathbb{R}^L \quad (3.25)$$

At the decoder side, the hidden bit is retrieved from z_k by applying the standard DM decoding rule to the rational function $z_k/g(\mathbf{z}_{k-1})$, where $\mathbf{z}_{k-1} = (z_{k-1}, z_{k-2}, \dots, z_{k-L})$. Ideally we should divide the received sample by $g(\mathbf{y}_{k-1})$ to recover the same quantized quantity at the encoder, but, due to the unavailability of \mathbf{y}_{k-1} , we use \mathbf{z}_{k-1} as its estimate. Hence the hidden information bit is estimated according to

$$\hat{m}_k = \arg \min_{m_k \in \{-1, 1\}} \left| \frac{z_k}{g(\mathbf{z}_{k-1})} - Q_{m_k} \left(\frac{z_k}{g(\mathbf{z}_{k-1})} \right) \right| \quad (3.26)$$

The embedding/decoding scheme is fully depicted in Fig. 3.7. It is worth noting that the system needs to be initialized to a state shared by the embedder and the decoder.

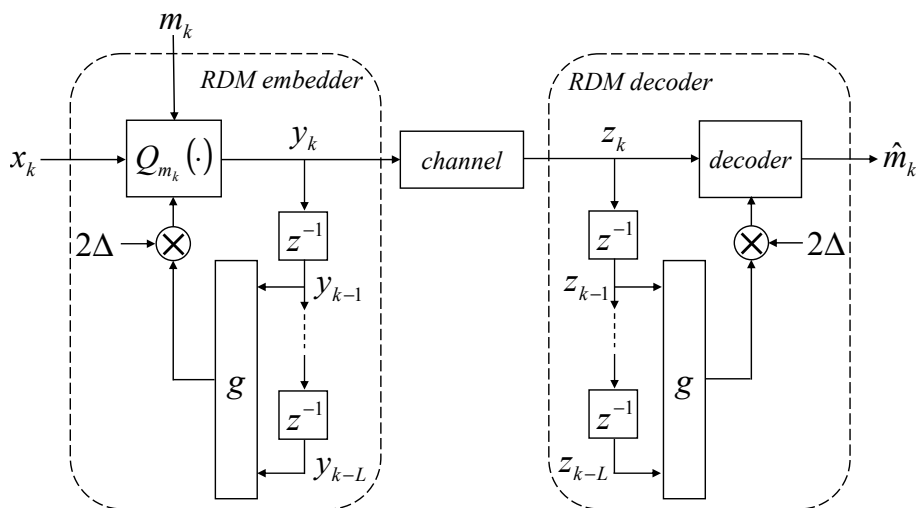


Figure 3.7: Block diagram of L th-order RDM.

The decoder output is intrinsically invariant to scaling applied to the watermarked signal since the gain ρ is canceled out by the ratio $z_k/g(\mathbf{z}_{k-1})$ in the decoding rule; in fact

3.3 Rational Dither Modulation

if $\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N})$, recalling the properties of the function $g \in \mathcal{G}$, we have

$$\frac{z_k}{g(\mathbf{z}_{k-1})} = \frac{\rho(y_k + n_k)}{g(\rho(\mathbf{y}_{k-1} + \mathbf{n}_{k-1}))} = \frac{y_k + n_k}{g(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})} \quad (3.27)$$

It is worth noting that the main element in RDM is the function $g \in \mathcal{G}$ and a significant parameter is the system memory L . In [104] the l_p vector-norms is adopted as function g , given by:

$$g(\mathbf{y}_{k-1}) = \left(\frac{1}{L} \sum_{i=1}^L |y_{k-i}|^p \right)^{1/p} \quad (3.28)$$

In [104] the selection of the parameter p that determines the l_p vector-norm is related with the distribution of X_k in order to minimize the discrepancy between $g(\mathbf{z}_{k-1})$ and $g(\mathbf{y}_{k-1})$. When X_k follows a generalized Gaussian distribution with shape parameter $c \in [1, \infty)$ and variance σ_x^2 , it is suggested to choose $p = c$, while for $c \leq 1$ a reasonable election is $p = 1$.

Increasing L the influence of the attacking noise on the decoding quantization step-size decreases [104]; in fact for large L , $g(\mathbf{z}_{k-1}) \approx g(\mathbf{y}_{k-1})$ and consequently the embedding and decoding step-size become very close in spite of the additive noise source in the channel.

The most interesting properties of RDM are noticeable when the memory size L is large enough to approximate the function $g(\mathbf{Y}_{k-1})$ as a constant value and $g(\mathbf{Z}_{k-1}) \approx g(\mathbf{Y}_{k-1})$.

For large L , we have $\{g(\mathbf{Y}_{k-1})\} \rightarrow g(\tilde{\mathbf{Y}})$ and its pdf has been computed in [104] as

$$f_{g(\tilde{\mathbf{Y}})}(s) \approx \frac{p s^{p-1}}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(s^p - M_{yp})^2}{2\sigma_r^2}\right) \approx \delta\left(s - M_{yp}^{1/p}\right) \quad (3.29)$$

being $M_{yp} \triangleq E[|Y|^p]$ and $\sigma_r^2 \triangleq (1/L)E[|Y|^{2p}] - (1/L)E^2[|Y|^p]$. Moreover, for DWR sufficiently large, it can be reasonably assumed $M_{yp} \approx M_{xp}$ and $\sigma_r^2 = (1/L)E[|X|^{2p}] - (1/L)E^2[|X|^p]$. Hence, recalling the RDM embedding function, it is clear that for large L it behaves as a DM embedding function with step-size $2\Delta M_{xp}^{1/p}$ and, if DWR is large enough to assume X having an almost flat pdf within each quantization bin, according to eq. (3.4) the embedding distortion can be written as

$$D_w = \frac{\Delta^2 M_{xp}^{2/p}}{3} \quad (3.30)$$

Moreover, increasing L the influence of the attacking noise on the decoding quantization step-size decreases [104]; in fact for large L , $g(\mathbf{Z}_{k-1}) \approx g(\mathbf{Y}_{k-1})$ and $f_{g(\tilde{\mathbf{Z}})}(s) \approx f_{g(\tilde{\mathbf{Y}})}(s)$, so that the embedding and decoding step-sizes become very close in spite of

the additive noise source in the channel. Under these hypothesis, the analytical error probability of RDM has been computed as

$$P_e = \int_0^\infty f_{g(\tilde{\mathbf{Z}})}(s) P_{DM}(2\Delta s) ds \quad (3.31)$$

where $P_{DM}(2\Delta s)$ is the error probability of DM as it has been written in eq. 3.7. Hence the analytical error probability of RDM can be obtained by averaging that of DM for all the possible values of the step-size. However, for L large enough to approximate $f_{g(\tilde{\mathbf{Z}})}(s) \approx \delta\left(s - M_{yp}^{1/p}\right)$, $P_e \approx P_{DM}(2\Delta M_{yp}^{1/p})$ so that the same performance of DM are expected.

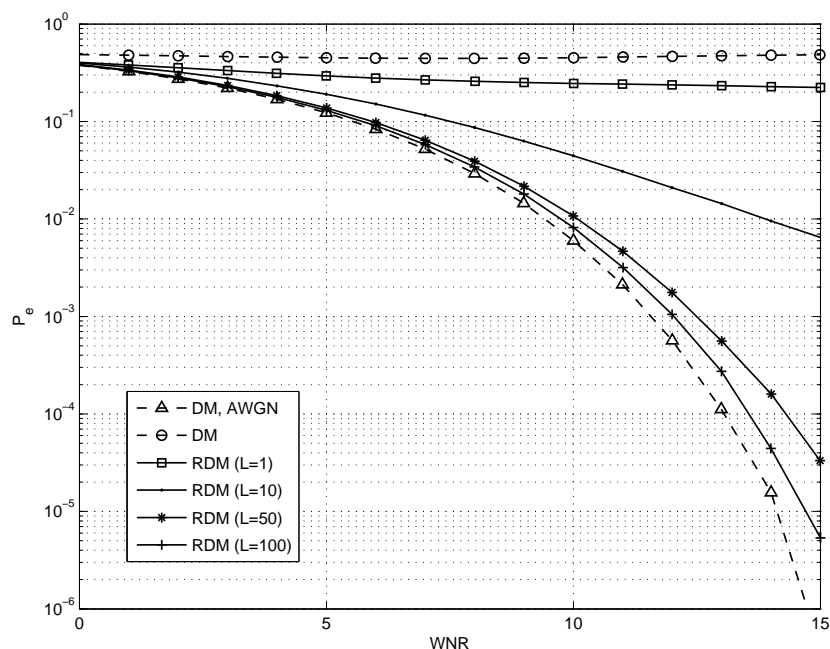


Figure 3.8: Empirical values of error probability of RDM against FGA ($\rho = 1.05$) for different values of L , Gaussian host and DWR=25 dB.

In Fig. 3.8, the error probabilities of RDM for different values of L are shown considering the host samples generated as i.i.d. Gaussian random variables, l_p -norm with $p = 2$ as $g(\cdot)$ function and DWR=25 dB. The error probability of DM with DWR=25 dB against AWGN channel have been also plotted as reference. Here it can be seen that it is possible to reduce the performance loss of RDM w.r.t. DM for the same DWR by increasing the memory size. On the other hand, the error probability of RDM is proven

3.3 Rational Dither Modulation

to be invariant to constant gain while, as it is shown in Fig. 3.8, the error probability of DM is approximately equal to 0.5 already for gain scaling $\rho = 1.05$.

However, a great variety of data hiding methods have been proposed to achieve robustness to gain scaling, as it has been discussed in Section 3.2. Even if, as it is stated in [104], the comparison of RDM with watermarking methods that provide robustness to gain scaling through other approaches is quite difficult, RDM has undeniable advantages w.r.t. techniques based on spherical codewords and those based on pilot insertion. First of all, RDM may work in a scalar fashion, so that its complexity is much less than algorithms based on spherical codewords. On the other hand, pilot-based algorithms waste some payload for the pilot embedding, while RDM does not incur in this penalty.

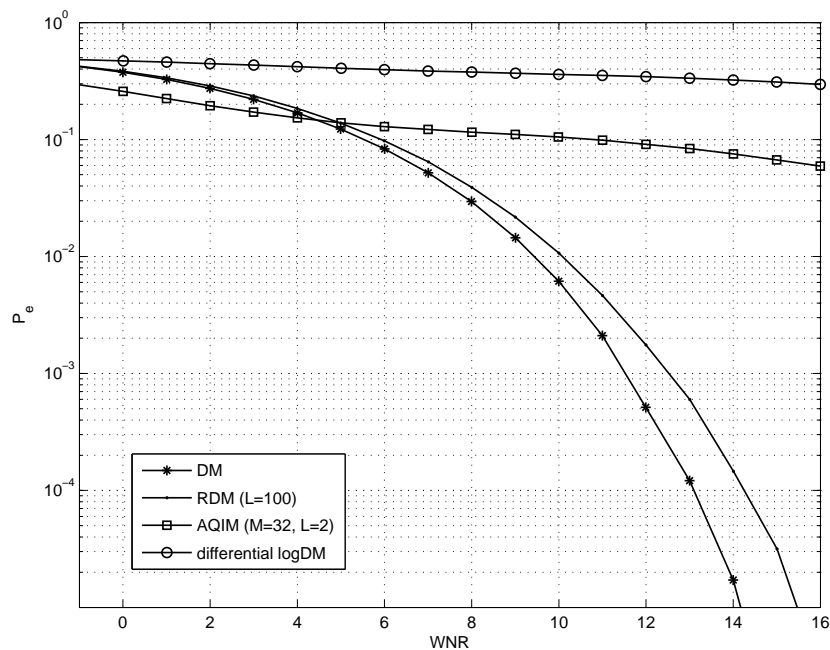


Figure 3.9: Empirical values of error probability for different gain invariant quantization-based techniques, for Gaussian host and DWR=25 dB.

A fair comparison can be conducted with respect to other scalar quantization-based methods able to cope with gain scaling on the channel. The outstanding performance of RDM in case of AWGN channel and Gaussian host can be noticed in Fig. 3.9, where the BER of DM has been depicted as reference. Here we have compared RDM with AQIM [106] and differential logarithmic DM [107], fixing the DWR to 25 dB in all the

simulations. We want to remark that the payload of AQIM is half the payload of RDM and differential logarithmic DM. From these results, it can be seen how RDM combines the properties of basic scalar quantization-based embedding with gain scaling invariance, containing the loss of robustness to additive noise.

3.4 Concluding remarks

The most interesting result introduced by RDM is without any doubt that, for large L , a gain invariant quantization-based method is able to asymptotically achieve the performance of DM against additive white Gaussian noise. The considerations drawn in the previous Section clarify why, since RDM was proposed, it has received quite a lot of interest from the data hiding community. Moreover, in [104], RDM was supported by a deep theoretical analysis which proves its performance and provides an useful basis to develop practical applications and improvements. So far, several other subsequent methods have exploited the properties of RDM combining it with channel coding [127] or source coding [128] and applying RDM to image data hiding with Watson-like perceptual distances [114], to audio data hiding [129] and to video data hiding [130]. Recently RDM has been also adopted in a buyer-seller watermarking protocol [131].

However, considering the gain attacks somewhat solved for quantization based schemes by the various algorithms proposed following different approaches, research interest has moved to other desynchronization attacks. In fact gain scaling can be considered the simpler volumetric distortion that the host signal can undergo on the channel. Extending robustness of DM-based methods to more complex volumetric desynchronization attacks is an essential requirement to develop practical data hiding applications based on them. RDM is theoretically proven to be invariant to gain scaling and it retains the fundamental advantage of DM, which can be identified in satisfactory trade-off between robustness to the AWGN channel, capacity capability and host signal fidelity. Hence it can be usefully exploited as basic tool to account for more complex desynchronization attacks. Recently, some data hiding methods based on RDM have been proposed that achieve robustness against nonlinear distortions [55,56] and linear time invariant filtering [58,105]. Moreover, starting from the theoretical framework provided in [104] for RDM, a solid theoretical analysis has been furnished also for the proposed extension of RDM, clarifying their actual achievable performance.

4

Providing invariance to nonlinear valumetric scaling for quantization based data hiding

In this chapter, it is presented a class of data hiding methods to counteract a nonlinear valumetric channel. The theoretical invariance property to nonlinear valumetric attack is achieved developing a proper extension of any gain invariant quantization-based algorithm. Among the wide class of nonlinear distortion, the *power-law attack* is here defined, which consists of a constant exponentiation and a constant gain scaling of the watermarked signal. The interest in this attack relies on the fact that it models common processings for image and video, such as gamma correction.

To provide invariance to power-law attack, the proposed method amounts to perform a mapping of the host samples in a convenient transformed domain where the watermark can be subsequently embedded using any gain invariant QIM-based technique. Even when used in this framework, RDM exhibits better performance than any other gain invariant methods and consequently a deep study has been carried on the proposed scheme using RDM, which has been named hyperbolic RDM. A theoretical analysis has been provided for hyperbolic RDM to identify the lower bound of the decoding error probability and to quantify the embedding distortion. Moreover, the results of several experiments for both Gaussian host and real images have been provided to verify the invariance to power-law attack and to confirm the exactness of the theoretical analysis.

The power-law attack and the proposed framework are defined in Section 4.1. In

Section 4.2 the domain invariant to both gain scaling and exponentiation is presented and the experimental results for some gain invariant QIM-based techniques are shown. Section 4.3 is devoted to the theoretical analysis of hyperbolic RDM and then the empirically evaluated decoding error probabilities are compared with the analytical approximation. Eventually, some final considerations are summarized in Section 4.4.

4.1 The power-law attack

Among the wide class of valumetric distortions, as defined in Section 3.2, one of the basic attacks consists of a constant gain scaling the watermarked signal, namely the fixed gain attack. However, nonlinear functions are a more-widely used distortion class; as an example, the gamma correction $\Gamma_\gamma(p)$ is especially relevant for video watermarking applications in case the signal has passed through DA/AD conversions. Gamma correction function is given by

$$\Gamma_\gamma(p) = p_{max} \left(\frac{p}{p_{max}} \right)^\gamma \quad (4.1)$$

where p_{max} is the maximum value of a pixel.

The power-law attack, which is defined as a constant exponentiation and a constant gain scaling of the amplitudes of the watermarked signal, is a nonlinear attack channel, as illustrated in Fig. 4.1; gamma correction is then a particular case of power-law attack. We assume that on the channel a zero-mean additive white Gaussian noise \mathbf{N} with variance σ_n^2 and independent of the watermarked signal is added too. In this way the context of a watermarked signal affected by a nonlinear distortion can be compared with other QIM-based methods, even those non-robust against valumetric distortions. Let $\rho > 0$ denote the gain parameter and $\gamma > 0$ the exponent; the attacked signal is written as

$$\mathbf{Z} = \rho (\mathbf{Y} + \mathbf{N})^\gamma \quad (4.2)$$

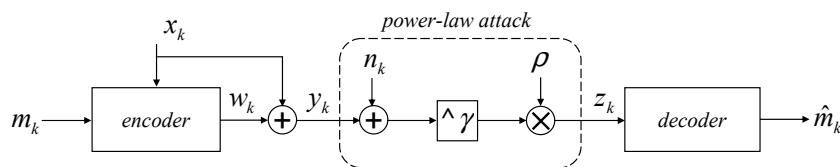


Figure 4.1: Power-law attack.

It is worth recalling here the considerations drawn in Section 3.2 on the inability for MSE-based distortion measures to account for non-additive attacks. For the power-law attack, both the scale factor and the exponentiation may produce very large MSE values, but the resulting perceptual modifications on the attacked signal can be not significant, as it is shown in Fig. 3.2. Hence, also in this framework the valumetric distortion is considered imperceptible and it is reasonable to disregard it when the effects of noise addition are evaluated. In case of power-law attack, the attacking noise distortion D_c is then conveniently defined so that it is independent of ρ and γ :

$$\begin{aligned} D_c &= \frac{1}{M} \sum_{k=1}^M E \left[\left((\rho^{-1} Z_k)^{-\gamma} - Y_k \right)^2 \right] = \\ &= \frac{1}{M} \sum_{k=1}^M E [N_k^2] = \sigma_n^2 \end{aligned} \tag{4.3}$$

In the previous chapter, various approaches and techniques to make quantization-based data hiding robust to gain scaling have been presented. But these techniques, developed all to cope with gain attack, lack of any property to counteract nonlinear distortions, which can severely impair the watermark decoding. Consequently, in case of power-law attack, the error probability for gain invariant quantization-based schemes are expected to be nearly 0.5. hence, methods to preserve the watermark information when the watermarked signal undergoes these kind of distortions are obviously welcome as not much literature exists on this problem. Among these, Guerrini et al. [132] present a method which embeds the watermark samples into the kurtosis of selected image blocks since this parameter remains approximately constant if the image content is not modified significantly. However the proposed decoding method is able only to verify the presence of the watermark (one-bit watermarking). In [120] Bas proposes a method which uses a proper set of quantizers, named floating quantizers; here the quantization step size is a function of the minimum and the maximum of a triplet of pixels. In this way the quantizers use a content-dependent quantization grid, whose variations in case of nonlinear distortions do not impair the hidden message detection.

4.2 Hyperbolic transformation providing invariance to power-law attack

To make quantization-based data hiding robust to power-law attack, a possible approach is to define a convenient domain proven to be invariant to such attack, where the embedding and the decoding are performed to successfully retrieve the hidden information.

Here, a domain invariant to both gain and exponentiation is presented. Mapping the host samples into this convenient domain, the information embedding can be performed according to a rule within the class of gain invariant QIM-based methods. Therefore any gain invariant QIM-based algorithm can be made invariant to the power-law attack.

We recall here how a point (s, t) in the first quadrant of a two dimensional Cartesian plane can be described by its hyperbolic coordinates representation (v, u) . The mapping from Cartesian to hyperbolic plane is a function $\{\mathbf{f} : (\mathbb{R}^+, \mathbb{R}^+) \rightarrow (\mathbb{R}^+, \mathbb{R})\}$ whose coordinates are given by:

$$v = \sqrt{st} \tag{4.4}$$

$$u = -\frac{1}{2} \log \left(\frac{t}{s} \right) \tag{4.5}$$

where u is the hyperbolic angle and v is the geometric mean.

The inverse mapping $(s, t) = \mathbf{f}^{-1}(v, u)$ to the Cartesian coordinates representation gives

$$s = v \exp(u) \tag{4.6}$$

$$t = v \exp(-u) \tag{4.7}$$

For the forthcoming discussion it can be shown that the hyperbolic angle, computed from (s, t) according to (4.5), exhibits an interesting behavior against power-law scaling of both Cartesian samples. In fact given $\mathbf{f}(s, t) = (u, v)$, the hyperbolic angle corresponding to $(\rho s^\gamma, \rho t^\gamma)$ results

$$u' = -\frac{1}{2} \log \left(\frac{\rho t^\gamma}{\rho s^\gamma} \right) = -\frac{1}{2} \gamma \log \left(\frac{t}{s} \right) = \gamma u \tag{4.8}$$

Eq. (4.8) shows that mapping two samples $(\rho s^\gamma, \rho t^\gamma)$ in hyperbolic angle, the gain ρ is canceled out by the ratio and the exponent γ becomes a scaling gain. Hence the basic idea is to apply a gain invariant QIM-based scheme to the transformed variable u , so that the message embedding is performed in a domain intrinsically invariant to both gain scaling

4.2 Hyperbolic transformation providing invariance to power-law attack

and exponentiation. Since only the hyperbolic angle coordinate constitutes the domain invariant to both gain and exponentiation, a formal representation of the mapping as a transformation in a hyperbolic domain seems to be not essential; however throughout the thesis, the name hyperbolic and the aforementioned notation will be used to clearly distinguish the proposed algorithms from the others in literature.

4.2.1 Embedder

Let x_k denote the current sample to be marked and \mathbf{y}_{k-1} denote the vector containing the L_h previously watermarked sample, i.e. $\mathbf{y}_{k-1} = (y_{k-1}, y_{k-2}, \dots, y_{k-L})$. Here, the couple of variables $(s, t) = (x_k, h(\mathbf{y}_{k-1}))$, composed by the k th host signal sample and a proper function $h(\cdot)$ computed on the L_h previous watermarked samples, is mapped into the transformed domain as

$$u_k = -\frac{1}{2} \log \left(\frac{h(\mathbf{y}_{k-1})}{x_k} \right) \quad (4.9)$$

A set \mathcal{H} of functions $h : \mathbb{R}^{L_h} \rightarrow \mathbb{R}$, with $L_h \geq 1$, is needed, which have the property that

$$h(\rho \mathbf{y}^\gamma) = \rho h(\mathbf{y})^\gamma \quad \forall \rho > 0, \forall \gamma > 0, \mathbf{y} \in \mathbb{R}^{L_h} \quad (4.10)$$

The function $h \in \mathcal{H}$ has to be properly chosen since the whole processing depends on its behavior. An example of function belonging to the set \mathcal{H} is the geometric mean:

$$h(\mathbf{y}_{k-1}) = \left(\prod_{i=1}^{L_h} |y_{k-i}| \right)^{1/L_h} \quad (4.11)$$

which will be used in the forthcoming discussion. However other, more or less complicated functions could exist. The main weakness of the geometric mean $h(\mathbf{z}_{k-1})$ is its proneness to drift away from the desired value $h(\mathbf{y}_{k-1})$ in case of additive noise. Choosing $L_h = 1$, would result $h(y_{k-1}) = |y_{k-1}|$, so that the k th sample in the transformed domain u_k depends on two subsequent sample, without the need of the function $h(\mathbf{y}_{k-1})$. But, as it will be verified in the experimental results in Section 4.3.4, for small values of L_h the drift caused by noise is expected to dramatically reduce the system performance.

As sketched in Fig. 4.2, the k th information bit m_k is embedded into k th sample u_k by the chosen QIM-based encoding rule, obtaining the relative quantized sample u_k^Q .

4.2 Hyperbolic transformation providing invariance to power-law attack

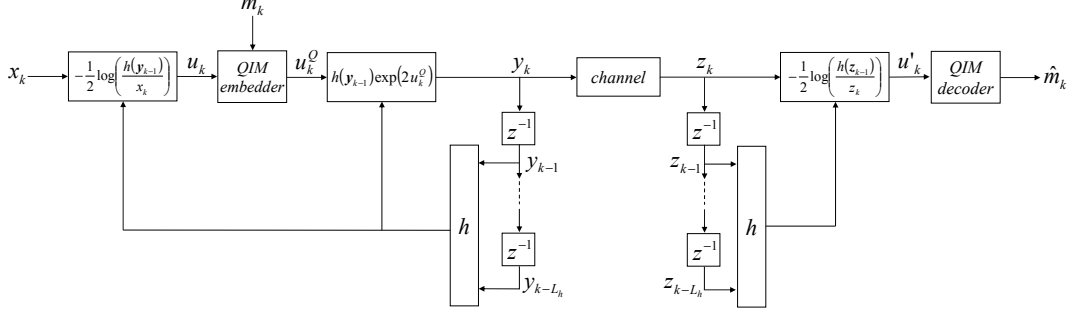


Figure 4.2: Block diagram of the hyperbolic watermarking system.

Then the marked sample u_k^Q has to be converted back to Cartesian coordinates. If the inverse mapping were performed using the equations (4.6) and (4.7), then the embedded watermark sample $(u_k^Q - u_k)$ would spread over both Cartesian coordinates. However the second coordinate $h(\mathbf{y}_{k-1})$ cannot vary since it depends uniquely on the previously watermarked samples. Since we need to map back only the hyperbolic angle coordinate, where the k th bit was embedded, the inverse mapping equation actually used is obtained by inverting eq. (4.9):

$$y_k = h(\mathbf{y}_{k-1}) \exp\left(2 u_k^Q\right) \quad (4.12)$$

where y_k is the k th watermarked sample in the host domain.

4.2.2 Decoder

At the decoder side, the information bits are retrieved applying the chosen gain invariant QIM-based decoder to the estimated marked samples after hyperbolic transformation. Hence, given the k th received and possibly attacked sample z_k , the decoder has to map this sample into the transformed domain and retrieve the embedded information bit using the decoder of the chosen QIM-based method, as shown in Fig. 4.2. To recover exactly the k th sample in the transformed domain, the decoder should ideally compute u'_k from $(z_k, h(\mathbf{y}_{k-1}))$, but the unavailability of $h(\mathbf{y}_{k-1})$ forces the use of $h(\mathbf{z}_{k-1})$ as an estimate of it. Hence the received k th sample in the transformed domain is computed as

$$u'_k = -\frac{1}{2} \log\left(\frac{h(\mathbf{z}_{k-1})}{z_k}\right) \quad (4.13)$$

The whole embedding/decoding scheme is sketched in Fig. 4.2.

4.2 Hyperbolic transformation providing invariance to power-law attack

The decoder output is intrinsically invariant to a power-law attack applied to the watermarked signal. In fact the gain ρ is canceled out by the ratio between the current received sample and $h(\mathbf{z}_{k-1})$, which is performed within the mapping function; the power exponent becomes a gain scaling in the hyperbolic angle representation and a QIM-based algorithm robust to gain scaling can counteract this attack. Thus, recalling the property of $h \in \mathcal{H}$ and (4.5), in case of a power-law attack $\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N})^\gamma$, the k th received sample in the transformed domain can be written as

$$\begin{aligned} u'_k &= -\frac{1}{2} \log \left(\frac{h(\mathbf{z}_{k-1})}{z_k} \right) = -\frac{1}{2} \log \left(\frac{h(\rho(\mathbf{y} + \mathbf{n})_{k-1}^\gamma)}{\rho(y_k + n_k)^\gamma} \right) = \\ &= -\frac{1}{2} \log \left(\frac{\rho h(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})^\gamma}{\rho(y_k + n_k)^\gamma} \right) = -\gamma \frac{1}{2} \log \left(\frac{h(\mathbf{y}_{k-1})}{y_k} \right) \end{aligned} \quad (4.14)$$

Inserting (4.12) in (4.14) and neglecting the noise contribution we have

$$u'_k = \gamma u_k^{\text{Q}} \quad (4.15)$$

As it was expected the received sample in the transformed domain u'_k is a scaled version of the equivalent sample at the embedder side u_k , from which a gain invariant QIM-based method is able to retrieve the correct information bit.

4.2.3 Experimental results

In this section it is exhibited the effectiveness of the proposed transformed domain invariant to the power-law attack and the results obtained for different QIM-based techniques are compared. We have tested the proposed algorithm using RDM, logarithmic DM and AQIM, that have been described in the previous chapter, to embed the hidden information in the transformed samples.

Even if the proposed algorithm can be applied to any host signal, independently of the particular content type, it is well suited to be used in image data hiding to make the watermark robust to gamma correction. Hence the host samples x_k can be considered as generated according to a random variable X which resembles the pixel image distribution, since the valumetric distortions are typically performed in the space domain. X_k are then assumed independent and identically distributed (i.i.d.) Gaussian random variables with mean μ_x and variance σ_x^2 , constrained to have real value within the range $[0, 255]$.

In all the experiments, the DWR was fixed at 25 dB. The strength of the noise addition attack is measured by the WNR, as defined in eq. (4.3).

4.2 Hyperbolic transformation providing invariance to power-law attack

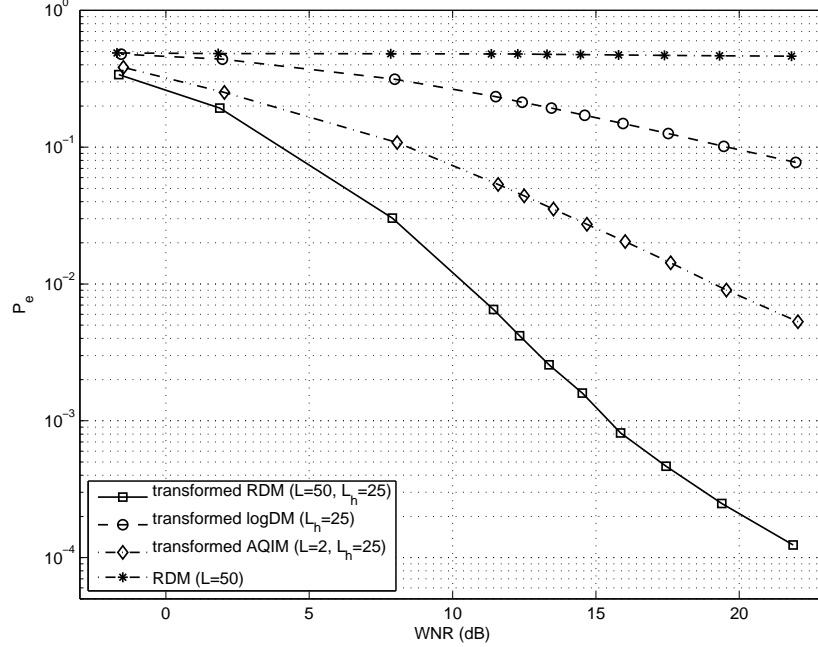


Figure 4.3: Empirical values of the error probability for power-law attack with $\rho = .7$ and $\gamma = 1.2$ (DWR=25 dB).

Fig. 4.3 shows the empirical values of the BER for RDM, logarithmic DM and AQIM applied in the transformed domain in case of power-law attack. Moreover in Fig. 4.3 it is also shown the error probability for RDM directly applied in the host signal domain undergone to the same attack, which is roughly equal to 0.5 as it was expected. The experimental results confirm the invariance to power-law attack of the proposed scheme for every QIM-based embedding since the error probabilities are equal to the ones measured for the Gaussian noise addition alone ($\rho = 1$ and $\gamma = 1$), here not presented.

Moreover from Fig. 4.3 it can be noticed that RDM applied after hyperbolic transformation, that we shall name hyperbolic RDM, outperforms the other schemes to perform the embedding in the transformed domain. Such a behavior is expected since RDM has lower error probabilities against additive noise also in case that it is directly applied to the host, as it has been discussed in Section 3.2. Reasonably, starting from these preliminary results, the research activity was devoted to investigate in deep hyperbolic RDM and to develop a theoretical analysis to prove its achievable performance.

4.3 Hyperbolic RDM

Hyperbolic RDM is an extension of basic RDM to provide invariance to power-law attack by mapping the host samples in the convenient domain described in Section 4.2. The whole embedding/decoding scheme is depicted in Fig. 4.4.

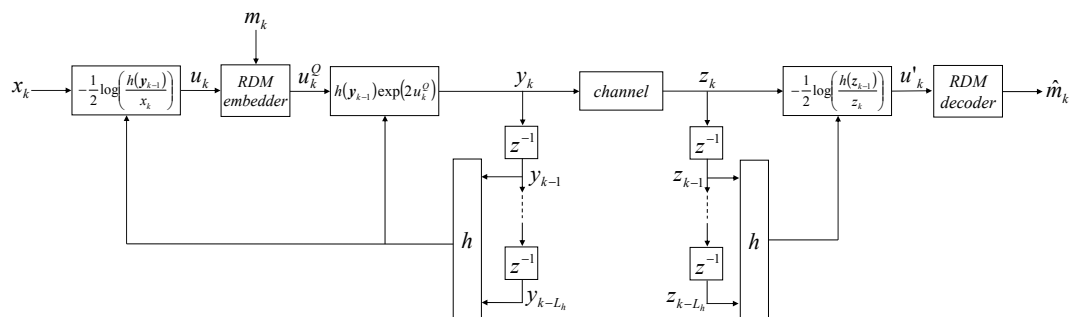


Figure 4.4: Block diagram of the hyperbolic RDM embedding/decoding scheme.

It is worth noting that in hyperbolic RDM we need to initialize both the inner and the outer memory vectors, which feed the functions $g \in \mathcal{G}$ in RDM embedder and $h \in \mathcal{H}$ in hyperbolic mapping, respectively. The whole initial state has to be shared by the embedder and the decoder. If the basic RDM decoder needs to receive L samples to update every element of the memory vector, in the proposed scheme the latency will be greater since, for the decoder memory to be updated, not only the outer memory but also the inner memory vector must be filled and updated with the received samples. Consequently the state of the system depends uniquely on the received samples after that $(L + L_h - 1)$ samples have been processed. It will be shown however that reasonable values of the inner and outer memory lengths do not impair too much on the system performance, since the length of the feature vector is typically much greater than $L + L_h$.

A theoretical analysis of hyperbolic RDM has been developed to compute the embedding distortion as a function of the quantization step-size Δ and to provide a lower bound for the decoding error probability. Being hyperbolic RDM an extension of RDM, the developed theoretical analysis is based on that presented in [104]. For analytical purposes the samples x_k are considered as generated according to a random variable X , so that X_k are assumed independent and identically distributed (i.i.d.) Gaussian random variables with mean μ_x and variance σ_x^2 . In the experiments, the value assumed by the Gaussian

random variables X_k has been constrained within the range $[0, 255]$ to resemble the usual 8-bit pixel image dynamic.

4.3.1 Embedding distortion

To evaluate the embedding distortion, we start from recalling the properties of RDM embedder fed by the host samples in the hyperbolic domain u_k . According to the analysis carried out in [104], for large L and large DWRs, RDM behaves similarly to a classical DM with step-size $2\Delta M_{up}^{1/p}$. M_{up} is the p th absolute moment of the random variable U and then it is defined as $M_{up} \triangleq E[|U|^p]$. Thereby, under these hypothesis, denoting by $\mathbf{v} \triangleq \mathbf{u}^\Omega - \mathbf{u}$ the watermark signal in the hyperbolic domain, V can be assumed independent of the host signal U and its steady-state variance results $\sigma_v^2 \approx (\Delta^2 M_{up}^{2/p} / 3)$. Hence we can write the k th watermarked sample as

$$y_k = h(\mathbf{y}_{k-1}) \exp[2(u_k + v_k)] = x_k \exp(2v_k) \quad (4.16)$$

We are interested in establishing the power of the watermark signal in the host domain $\sigma_w^2 = E[|Y - X|^2]$ ¹ which, for large DWRs and then for $\Delta \ll 1$, can be written as

$$\begin{aligned} \sigma_w^2 &= \int_{-\infty}^{+\infty} \int_{-\Delta M_{up}^{1/p}}^{\Delta M_{up}^{1/p}} [x(1 - \exp(2v))]^2 f_x(x) f_v(v) dx dv \approx \\ &\approx \int_{-\infty}^{+\infty} \int_{-\Delta M_{up}^{1/p}}^{\Delta M_{up}^{1/p}} [x(-2v)]^2 f_x(x) f_v(v) dx dv = \\ &= \int_{-\infty}^{+\infty} x^2 f_x(x) dx \int_{-\Delta M_{up}^{1/p}}^{\Delta M_{up}^{1/p}} (-2v)^2 f_v(v) dv = \\ &= m_x^2 (4\sigma_v^2) = \frac{4 m_x^2 \Delta^2 M_{up}^{2/p}}{3} \end{aligned} \quad (4.17)$$

as a consequence the data-to-watermark ratio is given by

$$DWR \triangleq \frac{m_x^2}{\sigma_w^2} = \frac{3}{4 \Delta^2 M_{up}^{2/p}} \quad (4.18)$$

and it can be noticed that, for any distribution of the host signal, the DWR is independent of the power of the original host, even if it depends on the p th absolute moment of the hyperbolic angle. This behavior, which has been already recognized for logarithmic

¹Note that $D_w \approx \sigma_w^2$ in steady-state, according to the fact that the watermark signal can be assumed stationary.

DM in Section 3.2.2, can be reasonably ascribed to the logarithmic transformation, since hyperbolic mapping is also based on the logarithmic function.

The p th absolute moment of the random variable U in (4.17) can be computed from the knowledge of the distribution of the hyperbolic angle random variable. Under the hypothesis of embedding with large DWRs, the pdf of Y should reasonably resemble the pdf of X . The hyperbolic angle, according to (4.9), can be written as

$$U = \frac{1}{2} \log \left(\frac{X}{h(\mathbf{Y})} \right) \approx \frac{1}{2} \log(X) - \frac{1}{2L_h} \sum_{i=1}^{L_h} \log(X_i) \quad (4.19)$$

Assuming $X \sim \mathcal{N}(\mu_x, \sigma_x)$ and $(3\sigma_x/\mu_x) < 1$ the logarithm can be approximated using a series expansion:

$$\log(1+x) = \sum_1^{\infty} (-1)^{n-1} \frac{x^n}{n}, \quad |x| < 1 \quad (4.20)$$

so that, truncating the series expansion to the second term, we have

$$\begin{aligned} \log(X) &= \log(\mu_x + \mathcal{N}(0, \sigma_x)) = \\ &= \log(\mu_x) + \log(1 + \mathcal{N}(0, \sigma_x)/\mu_x) \approx \\ &\approx \log(\mu_x) + \mathcal{N}(0, \sigma_x)/\mu_x - \frac{\sigma_x^2}{2\mu_x^2} \chi_1^2 \end{aligned} \quad (4.21)$$

where χ_n^2 denotes a chi square variate with n degrees of freedom. Being $(\sigma_x/\mu_x) < 1$, it is a rationale to neglect the chi square term since it is scaled by $(\sigma_x^2/2\mu_x^2) \ll 1$.

Using the above approximation and exploiting the independence of the original host components, the random variable U is now given by

$$U \approx \frac{1}{2} (\mathcal{N}(0, \sigma_x)/\mu_x) - \frac{1}{2L_h} \left(\mathcal{N}(0, \sqrt{L_h} \sigma_x/\mu_x) \right) \quad (4.22)$$

The knowledge of the distribution of the hyperbolic angle random variable allows to numerically evaluate the value of M_{up} and consequently the DWR can be computed as a function of the quantization step-size Δ , according to eq. (4.18).

To verify the exactness of the analysis above, DWRs numerically computed from eq. (4.18) have been compared with experimentally evaluated DWRs. In Figs. 4.5 and 4.6 are shown the DWRs evaluated for different values of Δ and for different values of L_h and L , respectively. As it was expected, the DWR decreases linearly with the logarithm of Δ ; moreover the DWR increases if bigger L_h and L are used, even if this behavior saturates very quickly.

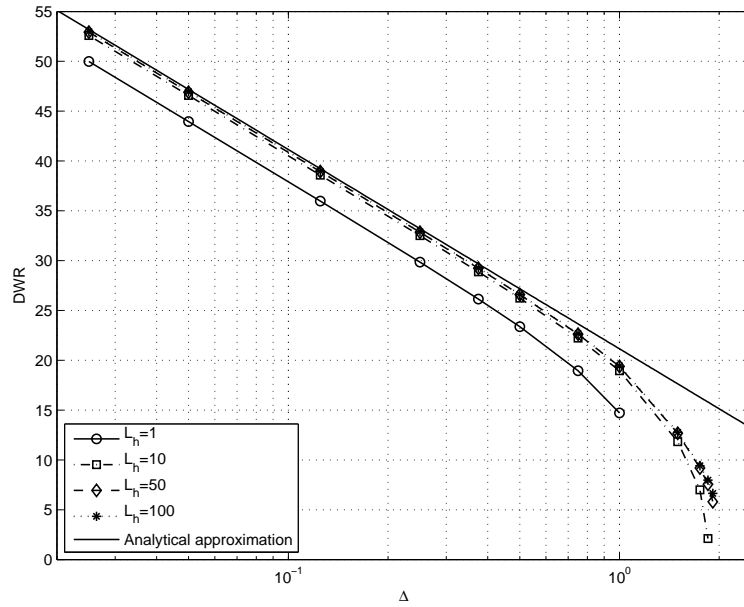


Figure 4.5: Comparison of the empirical values of DWR for different values of L_h with the analytical approximation.

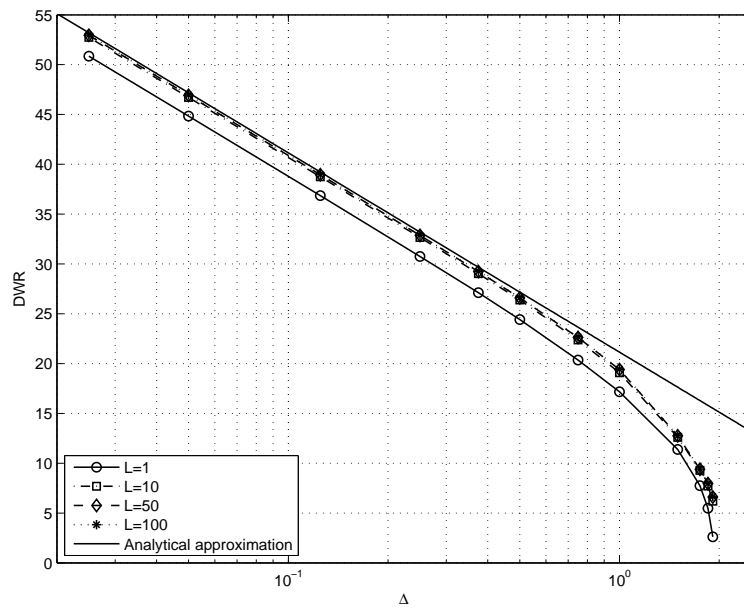


Figure 4.6: Comparison of the empirical values of DWR for different values of L with the analytical approximation.

It has to be noticed that embedding techniques that use a varying step size, like RDM, may suffer considerable peak embedding distortion, since if the step-size is momentarily large, the watermark sample will result also large. Hence, since the DWR is not able to assess this effect, another distortion measure could be adopted. In [104] the so called peak-to-average ratio (PAR) is used to evaluate peak embedding distortion:

$$PAR = \frac{w_+^2}{\sigma_w^2} \quad (4.23)$$

where w_+ is the magnitude exceeded by 1% of the watermark samples. For DM, since the watermarks are uniformly distributed in $[-\Delta, \Delta)$, the PAR is equal to 4.68 dB, while for logarithmic DM we have empirically measured the PAR and it resulted about 6.8 dB. The increased PAR of DM in logarithmic domain with respect to DM is motivated by the fact that in the former method the step size reported in the original domain increases with the magnitude of the host sample. On the other hand this behavior causes a perceptual shaping of the watermark, as it has been discussed in Section 3.2.2, since a larger watermark amplitude is introduced when the host sample takes a larger value. Obviously this effect cannot be taken in account by the PAR index.

Fig. 4.7 shows the embedding PAR for hyperbolic RDM as a function of L , for Gaussian host, DWR of 25 dB and $L_h = 100$. The empirical results in Fig. 4.7 demonstrate as, increasing L , the PAR for hyperbolic RDM approaches the PAR for logarithmic DM. Eventually, it can be inferred that for large L the gap in peak embedding distortion between hyperbolic RDM and DM is mainly due to the hyperbolic transformation; on the other hand the watermark samples embedded by hyperbolic RDM result perceptually shaped too.

4.3.2 Analytical derivation of the bit error rate

In this Section a statistical analysis of achievable BER for hyperbolic RDM against AWGN channel is outlined. Since hyperbolic RDM is intrinsically invariant to gain and exponentiation attacks, we assume $\rho = 1$ and $\gamma = 1$ in the BER analysis, so that $\mathbf{Z} = \mathbf{Y} + \mathbf{N}$. It is worth recalling that the host signal samples are assumed i.i.d. Gaussian random variables with mean μ_x and variance σ_x^2 .

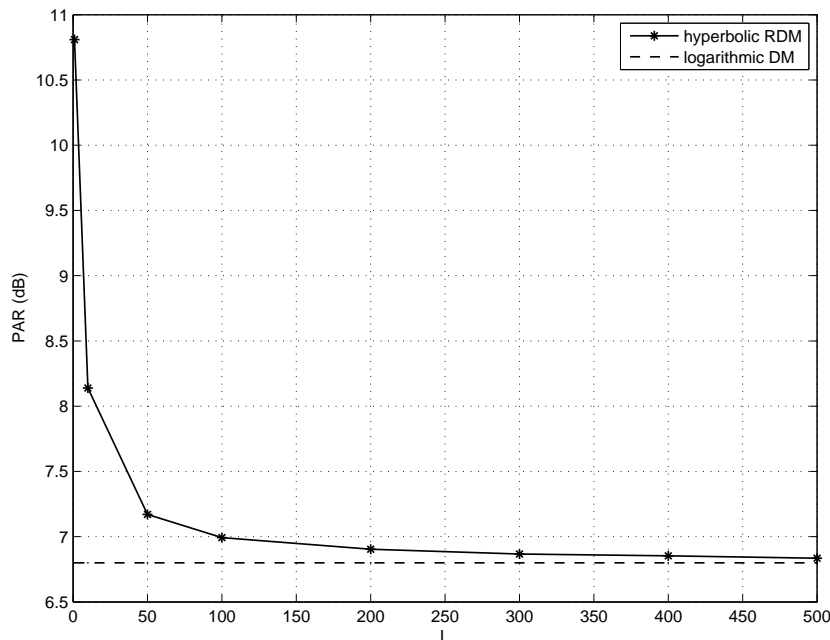


Figure 4.7: Empirical embedding PAR, as a function of memory size L , for a Gaussian host and $L_h = 100$.

In [104] the error probability of RDM for large L is analytically derived and, as it has been shown in Section 3.3, it is given by

$$P_e = \int_0^\infty f_{g(\tilde{\mathbf{Z}})}(s) P_{DM}(2\Delta s) ds \quad (4.24)$$

where $f_{g(\tilde{\mathbf{Z}})}(s)$ denotes the pdf of $g(\tilde{\mathbf{Z}})$, being $\tilde{\mathbf{Z}} = \lim_{k \rightarrow \infty} \{(Z_k, \dots, Z_{k-L+1})^T\}$ the random vector modeling the signal which feeds the RDM decoder. $P_{DM}(2\Delta s)$ denotes the error probability of DM for step-size $2\Delta s$, which is defined in eq. (3.7).

Our purpose is to exploit the decoding error probability in (4.24) within the proposed method. However eq. (3.7) and eq. (4.24) allow to compute the decoding error probability only in case of additive noise at the decoder input. In the considered framework, the noise is added to the watermarked signal and then the noisy samples are mapped to hyperbolic angle prior to RDM decoding, so that the noise at the RDM decoder input is no longer additive.

Thus, we need to define an equivalent channel where the additive noise source is moved just ahead of the RDM decoder, as it is sketched in Fig. 4.8(b). Because of the logarithm

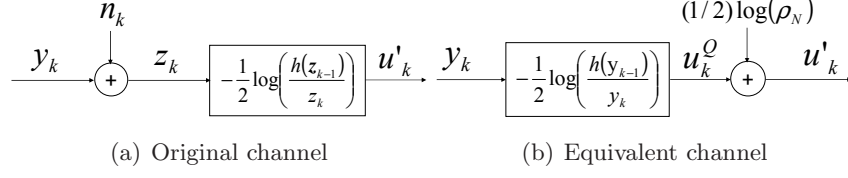


Figure 4.8: Illustration of original and equivalent channel at the decoder.

properties, it is convenient to represent the difference between the encoded and the received samples with a multiplicative factor:

$$\frac{z_k}{h(\mathbf{z}_{k-1})} = \frac{y_k + n_k}{h(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})} = \rho_N \cdot \frac{y_k}{h(\mathbf{y}_{k-1})} \quad (4.25)$$

where ρ_N is defined by

$$\rho_N = \frac{y_k + n_k}{y_k} \cdot \frac{h(\mathbf{y}_{k-1})}{h(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})} \quad (4.26)$$

According to eq. (4.25), the hyperbolic angle samples at the decoder can be written as

$$\begin{aligned} u'_k &= -\frac{1}{2} \log \left(\frac{h(\mathbf{z}_{k-1})}{z_k} \right) = \\ &= -\frac{1}{2} \log \left(\frac{h(\mathbf{y}_{k-1})}{y_k} \right) + \frac{1}{2} \log(\rho_N) = \\ &= u_k^Q + \frac{1}{2} \log \left(1 + \frac{n_k}{y_k} \right) + \frac{1}{2} \log \left(\frac{h(\mathbf{y}_{k-1})}{h(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})} \right) = \\ &= u_k^Q + \frac{1}{2} \log \left(1 + \frac{n_k}{y_k} \right) - \frac{1}{L_h} \sum_{i=1}^{L_h} \frac{1}{2} \log \left(1 + \frac{n_{k-i}}{y_{k-i}} \right) \end{aligned} \quad (4.27)$$

In eq. (4.27), it is exhibited a straight relation between the RDM watermarked samples and the received ones with two additive noise terms in the hyperbolic domain.

The noise terms $\log(1 + n_k/y_k)$ can be simplified using the series expansion of logarithm function in (4.20), being reasonably $(n_k/y_k) < 1$. Hence we can approximate the logarithm truncating the series at the first term $(1/2) \log(1 + n_k/y_k) \approx n_k/(2y_k)$; the noise term $r_k = n_k/(2y_k)$ is the ratio of two independent Gaussian random variables, whose pdf is equal to

$$\begin{aligned} f_R(r) &= \int_{-\infty}^{+\infty} |w| f_{2y}(w) f_n(wr) dw \approx \\ &\approx \int_{-\infty}^{+\infty} |w| f_{2x}(w) f_n(wr) dw = \\ &= \int_{-\infty}^{+\infty} \frac{|w|}{4\pi\sigma_x\sigma_n} \exp \left(-\frac{(w - 2\mu_x)^2}{8\sigma_x^2} - \frac{(wr)^2}{2\sigma_n^2} \right) dw \end{aligned} \quad (4.28)$$

Here the approximation is valid under the hypothesis of large DWRs. If the DWR is large enough and the original host components X_k are independent, the noise samples R_k are independent too.

Then the k th received sample in the hyperbolic angle domain can be written as

$$u'_k \approx u_k^{\text{O}} + r_k - \frac{1}{L_h} \sum_{i=1}^{L_h} r_{k-i} \approx u_k^{\text{O}} + r_k \quad (4.29)$$

where the further approximation is motivated by the fact that for large L_h we can invoke the Central Limit Theorem for the sum of i.i.d. components so that this noise source can be assumed approximately Gaussian with zero mean and variance $(\sigma_r^2/L_h) \ll \sigma_r^2$.

Rearranging properly eq. (4.24), the decoding error probability for hyperbolic RDM is given by

$$P_e = \int_0^\infty f_{g(\tilde{\mathbf{U}})}(s) P_{DM}(2\Delta s) ds \quad (4.30)$$

where $P_{DM}(2\Delta s)$ is computed for the equivalent additive noise pdf $f_R(r)$ in the hyperbolic angle domain.

Recalling the considerations in [104] and under the large DWRs hypothesis, $f_{g(\tilde{\mathbf{U}})}(s)$ can be approximated for large L as

$$f_{g(\tilde{\mathbf{U}})}(s) \approx f_{g(\tilde{\mathbf{U}})}(s) \approx \frac{ps^{p-1}}{\sqrt{2\pi}\sigma_t} \exp\left(-\frac{(s^p - M_{up})^2}{2\sigma_t^2}\right), \quad s \geq 0 \quad (4.31)$$

being $\tilde{\mathbf{U}} = \lim_{k \rightarrow \infty} \{(U_k, \dots, U_{k-L+1})^T\}$, p is the parameter depending on the chosen l_p vector-norm, $M_{yp} \triangleq E[|U|^p]$ and $\sigma_t^2 \triangleq (1/L) E[|U|^{2p}] - (1/L) E^2[|U|^p]$.

Since the pdf of the random variable U is known (see section 4.3.1), the analytical approximation for large L and L_h of the error probability of hyperbolic RDM against AWGN can be numerically computed from (4.30). In Fig. 4.9 the analytical approximation of the error probability has been compared with experimentally evaluated decoding error probabilities for DWR=25 dB. In the experiments, the memory vectors have length $L = L_h = 500$, so that they are large enough to approach the analytical approximation. The excellent matching of the empirical and analytical BERs is here noticeable and it proves the exactness of the developed theoretical analysis.

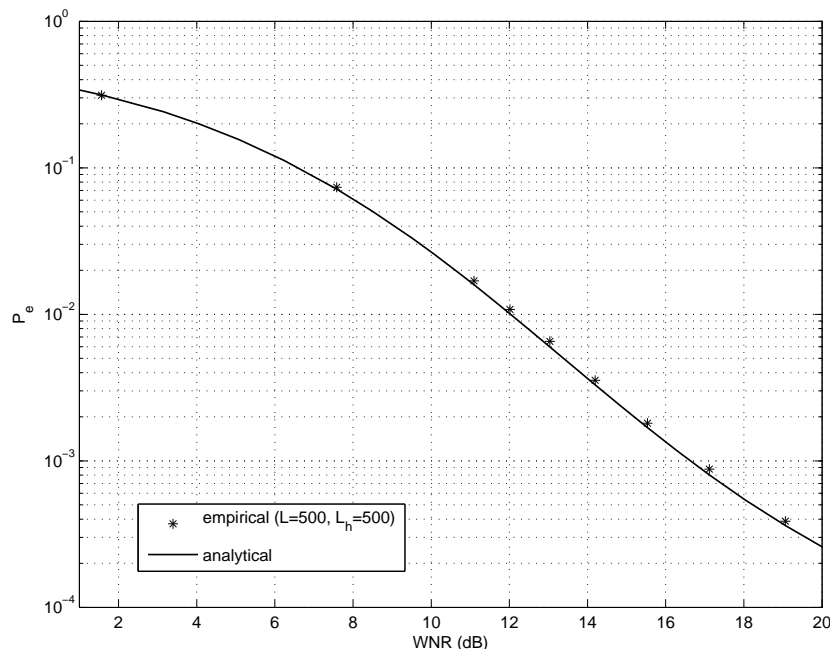


Figure 4.9: Empirical and analytical decoding error probabilities for hyperbolic RDM (DWR=25 dB).

4.3.3 Alternative scheme for memory reduction

It is possible to simplify the proposed scheme for hyperbolic RDM removing the inner memory vector, so that the overall memory size is reduced. In the RDM encoder and decoder, instead of using the function $g(\cdot)$ computed on the previously marked samples in the hyperbolic angle domain $\mathbf{u}_{k-1}^{\text{O}}$, a function g_h can be used, that is fed by the same previously marked samples \mathbf{y}_{k-1} that feed the function $h(\cdot)$.

The function $g_h : \mathbb{R}^{L_h} \rightarrow \mathbb{R}$, $L_h \geq 1$, must obey to the following property

$$g_h(\rho \mathbf{y}^\gamma) = \gamma g_h(\mathbf{y}) \quad \forall \rho > 0, \forall \gamma > 0, \mathbf{y} \in \mathbb{R}^{L_h} \quad (4.32)$$

Thus, in the RDM encoder and decoder, the sample to be marked is divided by the function $g_h(\mathbf{y}_{k-1})$, computed on the previously marked samples \mathbf{y}_{k-1} collected in the outer memory vector. The above proposed encoding scheme is sketched in Fig. 4.10. The advantage of this alternative scheme is that only the outer memory vector is needed. As an example, a

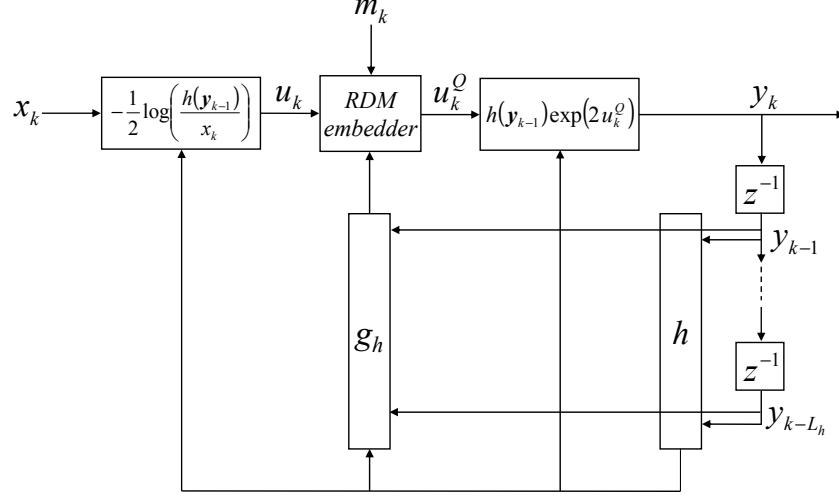


Figure 4.10: Alternative scheme for memory reduction at the encoding side.

function $g_h(\cdot)$ having the previously explained property is

$$g_h(\mathbf{y}_{k-1}) = \left(\frac{1}{L_h} \sum_{i=1}^{L_h} |\log(y_{k-i}) - \log(h(\mathbf{y}_{k-1}))|^{p_h} \right)^{1/p_h} \quad (4.33)$$

where $p_h \in \mathbb{N}$.

4.3.4 Experimental results

In this section it is exposed the effectiveness of the proposed method against power-law attack, for both Gaussian host and real images. Firstly, we illustrate how the memory sizes L and L_h affect the system performance against noise addition. Afterwards, we outline the robustness against additive noise, comparing the empirically measured error probabilities for hyperbolic RDM with the analytical approximation of the achievable BER for memory sizes going to infinity and large DWRs. The decoding error probabilities for classical DM, RDM and logarithmic DM have been considered as reference. Eventually the results for some experiments carried on real images are presented and some considerations are drawn

To evaluate the fidelity of the watermarked signal, the document-to-watermark ratio is used and, unless otherwise specified, it is fixed at 25 dB in all the experiments.

A crucial task in the analysis of hyperbolic RDM is to understand the relation between the size of memory vectors and the system performance. For this purpose the error probability of the whole encoding/decoding scheme was measured under additive noise attack

using a fixed quantization step size. We did not perform this test under power-law attack since the proposed scheme is intrinsically invariant to the corresponding modifications. The error probability obtained for different WNRs and different values of L and L_h are exhibited in Figs. 4.11 and 4.12, respectively. Also, the analytical approximation of BER is plotted as reference.

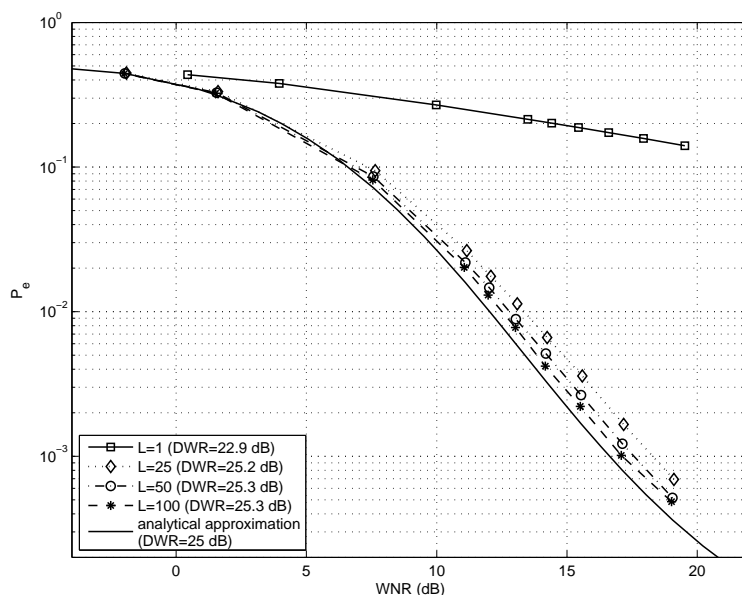


Figure 4.11: Analytical approximation and empirical values of the error probability for different values of L (here $L_h = 50$).

From inspection of Fig. 4.11, it can be noticed that increasing the RDM memory size L for values larger than $L = 25$, the performance slightly increase. On the other hand, in Fig. 4.12 it is shown that the BER is affected by unimportant changes for values bigger than $L_h = 25$. For the forthcoming experiments we have then chosen empirically the values $L = 50$ and $L_h = 25$ as a trade-off between memories sizes and robustness to noise.

Afterwards we have compared the empirical and the analytical BER for hyperbolic RDM w.r.t. DM and RDM ($L = 50$) under noise addition. In Fig. 4.13 the results are depicted. As it is expected, DM and RDM have a similar behavior, since RDM approaches DM for memory size going to infinity [104], and both of them outperform hyperbolic RDM. This is motivated by the fact that DM and RDM have been designed to cope with AWGN channel, as it has been discussed in Section 3.3. The performance loss of hyperbolic RDM can be firstly ascribed to the hyperbolic transformation itself. The white Gaussian noise

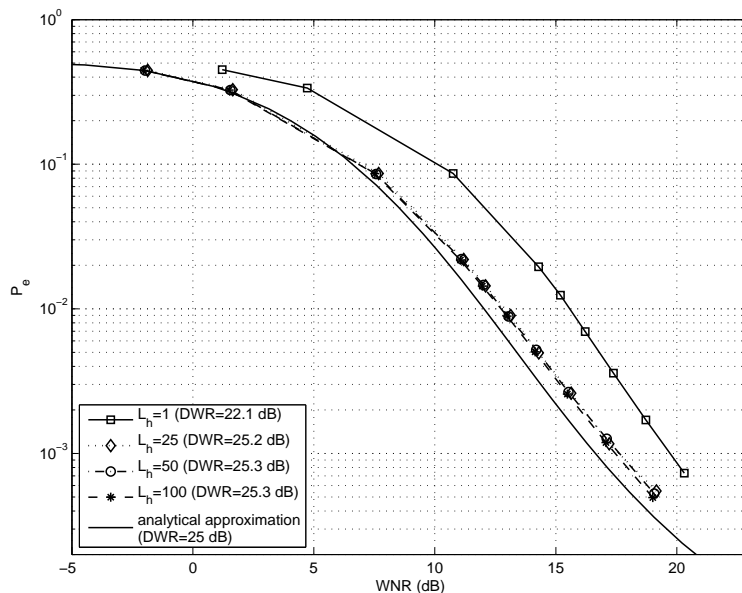


Figure 4.12: Analytical approximation and empirical values of the error probability for different values of L_h (here $L = 50$).

is indeed added to the watermarked signal and then the noisy samples are mapped to hyperbolic angle prior to RDM decoding; in this way the noise at the RDM decoder input is no longer Gaussian and, above all, no longer additive (see Section 4.3.2). The lower performance can be ascribed, in minor part, also to the drift of the geometric mean $h(\mathbf{z}_{k-1})$ at the receiver from the corresponding value $h(\mathbf{y}_{k-1})$ at the encoder caused by the noise when memory sizes are not very high.

A further confirmation of the above explanations about the system performance against additive noise can be inferred by the comparison provided in Fig. 4.13 of the experimental BERs of hyperbolic RDM and logarithmic DM with the analytical approximation of hyperbolic RDM. As shown, hyperbolic RDM approaches logarithmic DM for memory sizes going to infinity and this behavior is expected since RDM approaches DM for memory size going to infinity. However, the achievable BER for both hyperbolic RDM and logarithmic DM is lower than that of classical DM. Then it is possible to conclude that noise sensitivity can be ascribed mainly to the logarithmic transformation.

Finally, since the hyperbolic RDM is intrinsically invariant to power-law distortion, we expect superior performance w.r.t. DM, RDM and logarithmic DM in case the watermarked signal undergoes such an attack. Experimental results sustain this hypothesis.

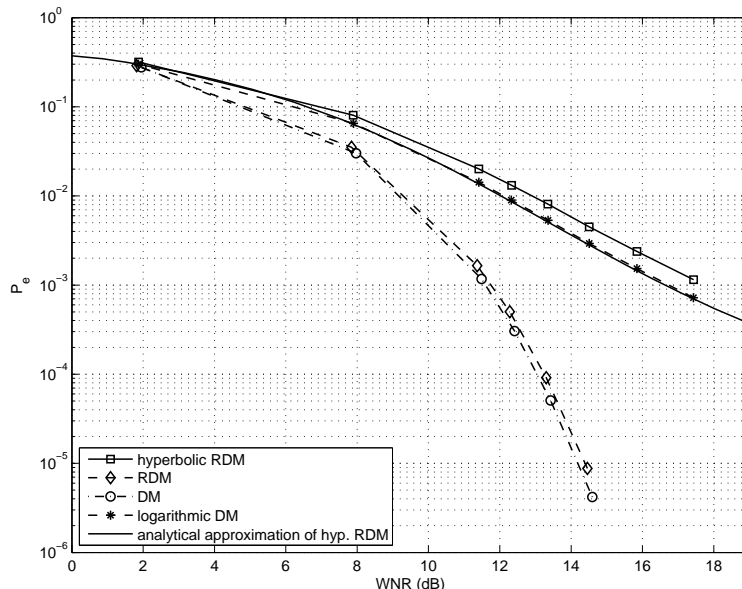


Figure 4.13: Analytical approximation and empirical values of the error probability for DM, RDM, logarithmic DM and hyperbolic RDM (DWR=25 dB).

In Fig. 4.14 are shown the empirical error probabilities for different watermarked signals against power-law attack $\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N})^\gamma$ with $\rho = .7$ and $\gamma = 1.2$. Under this attack, the other DM-based decoders fail and the relative error probabilities are roughly a half because these schemes totally lack in properties to cope with nonlinearity.

The alternative scheme for memory reduction, described in Section 4.3.3, was tested too. The measured error probabilities of the whole encoding/decoding method for different memory length and DWR equal to 25 dB are depicted in Fig. 4.15; here it is shown that the BER suffers unimportant changes for L_h bigger than 25. Moreover by comparison of the results in Fig. 4.11 and in Fig. 4.15, it is noticeable that the performance of the alternative scheme for the selected values of L_h are approximately equivalent to the ones of the basic scheme with two memory vectors.

Hyperbolic RDM has been tested also using real images as host signal. In order to apply the proposed watermarking method to images, embedding is performed in the space domain, inserting one bit per pixel. Prior to perform the embedding, the 2-D host has to be arranged in a 1-D vector \mathbf{x} . If the arrangement order is lexicographical, subsequent x_k are highly correlated and both the relative geometric mean and l_p norm will follow the signal trend. The resulting low variability of the hyperbolic angle reduces the system

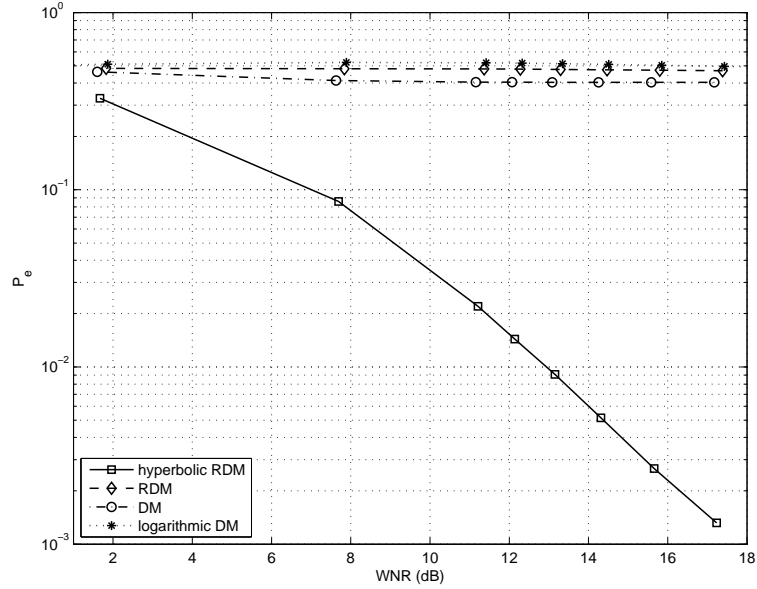


Figure 4.14: Empirical values of the error probability for power-law attack with $\rho = .7$ and $\gamma = 1.2$ (DWR=25 dB).

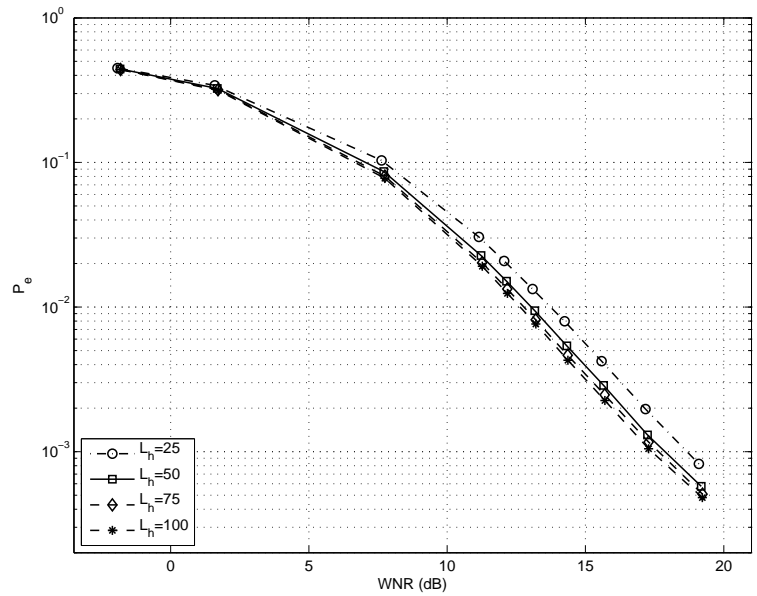


Figure 4.15: Empirical values of the error probability for the alternative scheme and different values of L_h .

performance. Then, before the embedding is performed, the host vector is scrambled changing the sample order. Now subsequent x_k will result highly uncorrelated and both the geometric mean and l_p norm will have small variations around the respective global mean values. By embedding the message into the scrambled image vector, the system performance are approximately similar to the ones obtained in the ideal case of Gaussian host vector.

Furthermore this approach is motivated by the fact that, designing a watermarking method for real images in the space domain, it is a customary to select a feature vector whose samples are a subset of the image pixels taken according to a proper law. On the other hand this is equivalent to perform the embedding into the whole scrambled host vector, since in both cases subsequent samples are not highly correlated.

The measured error probabilities have been computed averaging the results obtained using 10 simulations and the standard images Lena and Baboon as host signal. In all experiments one bit per sample has been embedded.

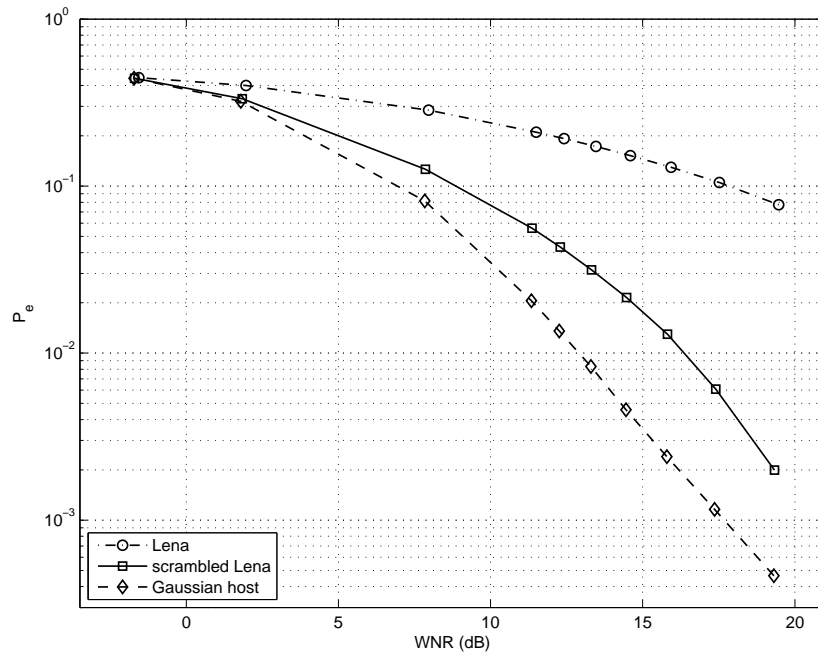


Figure 4.16: Empirical values of the error probability for different ordering of Lena host vector (DWR=25 dB).

In Fig. 4.16 are depicted the system performance in case the image samples are ar-

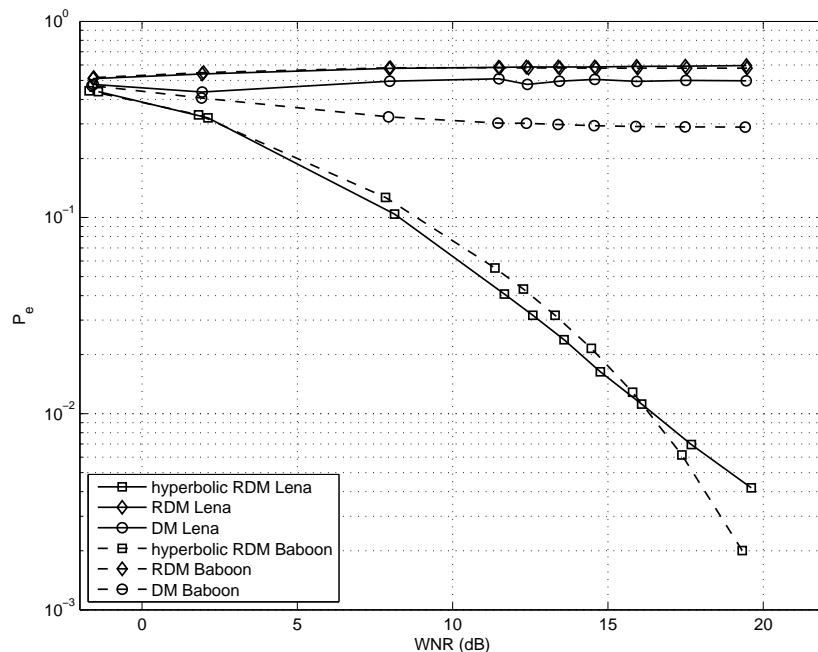


Figure 4.17: Empirical values of the error probability for scrambled Lena host vector and scrambled Baboon host vector under gamma compression with $\gamma = 1.3$ (DWR=25 dB).

ranged both in lexicographical order and in random order. It can be noticed that, embedding the watermark in the scrambled image vector, the measured error probability under noise addition is considerably lower.

The robustness of the proposed method under gamma correction, as defined in (4.1), was tested in case of real image host vector. Since gamma correction resembles a power-law attack, the empirical error probability is expected to be independent of the nonlinear distortion. Into the scrambled Lena vector and into the scrambled Baboon vector the watermark has been embedded using DM, RDM and hyperbolic RDM; the error probabilities evaluated from these watermarked signals, after gamma compression with $\gamma = 1.3$, are depicted in Fig. 4.17. Only hyperbolic RDM exhibits robustness against this attack, as it is expected.

Even if perceptual considerations go beyond the topic of the proposed study, whose aim is to develop a power-law invariant algorithm for a generic host signal, the standard image Lena and Baboon marked by hyperbolic RDM are here shown. Figs. 4.18 and 4.19 expose the comparison of the original images Baboon and Lena, respectively, and the

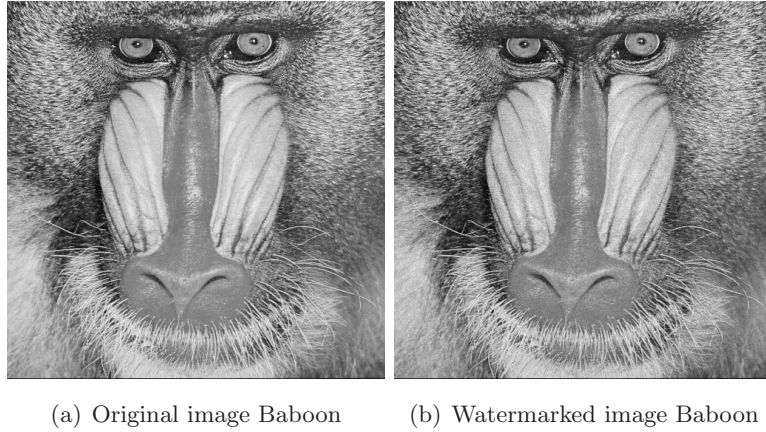


Figure 4.18: Comparison of original image Baboon and hyperbolic RDM watermarked image Baboon (DWR = 25 dB and Watson distance of 17).

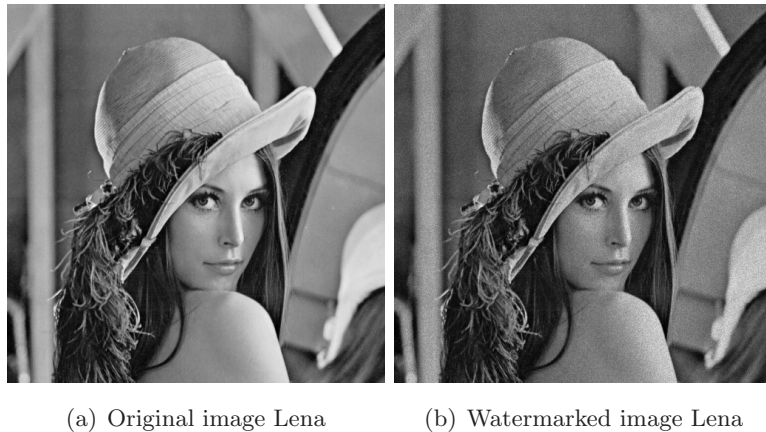


Figure 4.19: Comparison of original image Lena and hyperbolic RDM watermarked image Lena (DWR = 25 dB and Watson distance of 24).

corresponding watermarked image using hyperbolic RDM as it has been described above with DWR=25 dB. It is worth noting that the watermark has been embedded in the pixel domain and that no perceptual masking techniques [36] have been used. The watermarked image Baboon in Fig 4.18(b) exhibits high fidelity in comparison with the original image Baboon and this is confirmed by the Watson distance, which is approximately equal to 17. On the other hand, the distortion introduced by the watermark embedding is noticeable for the watermarked image Lena in Fig 4.19(b), whose Watson distance w.r.t. the original image Lena is approximately equal to 24. From Fig. 4.19 it can be seen that the watermark

is more visible in flat areas, where the marked image appears noisy. This effect is not appreciable in the standard image Baboon since it has not large flat areas. This noisy appearance of the watermark is reasonably due to the scrambling operation that changes the sample order and consequently neighboring pixels can be assumed RDM-quantized with independent quantization step-sizes, since it depends on the L previous watermarked samples. Remarking that the embedded watermark has not been perceptually shaped and that the scrambling operation is not mandatory, the development of practical data hiding applications for images based on hyperbolic RDM is still an open issue.

Regarding the comparison of hyperbolic RDM with other methods that are also designed to cope with nonlinear distortions, it must be said that it is quite difficult to find a set of fair conditions to establish such a comparison. The method proposed by Guerrini et al. [132] belongs to the class of one-bit watermarking, since it detects the presence of a known watermark sequence within the received host signal. Hence, it is not possible to compare fairly the method in [132] with hyperbolic RDM, whose decoder retrieves an estimate of each symbol hidden by the embedder. On the other hand the method proposed by Bas in [120] has the disadvantage that just one watermark symbol is embedded in a triplet of pixels, hence its payload is a third of that of hyperbolic RDM. Moreover, using a fractal set of quantizers as in [120], the watermark signal is not theoretically invariant to power-law attack and the strength of the distortions caused by the channel is constrained to be in a given range, according to some hypothesis in [120].

4.4 Concluding remarks

An extension of any gain invariant QIM-based algorithm to achieve invariance against power-law attack has been here proposed. This study is motivated by the fact that nonlinear valumetric distortions cause a dramatic increase of the decoding error probability for quantization-based data hiding schemes, due to the desynchronization introduced on the decoding lattice volumes. The proposed extension is proven to be intrinsically invariant to both gain scaling and exponentiation and it can be applied to every existing gain invariant quantization-based method.

Also, hyperbolic RDM has been proposed and it has been provided a theoretical analysis to have a lower bound of the achievable decoding error probabilities. It has to be noticed that a performance loss w.r.t. DM and RDM against additive noise has been

evaluated and a reason for this behavior has been found to be due essentially to the logarithmic transformation. However, even if hyperbolic RDM exhibits lower BERs against additive noise, it guarantees invariance to power-law attack. Moreover some experiments have been conducted for hyperbolic RDM applied to standard images, even if the application of the proposed method to real-world signals constitutes an open research issue and a considerable amount of work is necessary to tune hyperbolic RDM to the requirements of practical systems.

Although power-law attack is a well-defined attack and reasonably hyperbolic RDM could be vulnerable against other valumetric distortions, RDM has been proven to be easily extended to provide robustness to attacks different from gain scaling. Moreover, as it can be seen in Sections 4.3.1 and 4.3.2, the theoretical analysis for hyperbolic RDM has been developed starting from that provided for RDM in [104]. Thus, the thorough analytical assessment of the performance of RDM has been proven to be an useful tool to analytically validate an extension of RDM.

5

High-rate data hiding robust to linear filtering

The vulnerability of quantization-based data hiding methods to desynchronization attacks, such as filtering, has been discussed in Chapter 2. Considering the wide use of linear filtering in processing for every kind of media, the development of practical data hiding applications based on QIM requires to guarantee robustness against filtering.

To this purpose the discrete Fourier transform - rational dither modulation (DFT-RDM) has been proposed in [105]. There, the robustness against linear-time-invariant (LTI) filters is achieved combining the properties of discrete Fourier domain, which converts a filtering in time domain into an approximate multiplication on each frequency channel, and the gain invariance of RDM. Through experiments and theoretical analysis, DFT-RDM has been proven to give high data rate for white Gaussian host, but those rates considerably decrease for non-white hosts.

To understand the rationale behind this behavior, the theoretical analysis of DFT-RDM is generalized to colored Gaussian hosts and a frequency-domain approach is adopted to separately evaluate the performance on each RDM channel. Departing from this generalized analysis, an extension of DFT-RDM is presented to improve its performance for colored hosts without assuming any knowledge on the attack filter. Moreover, the expected performance improvement is also verified for real audio signals, that are nonstationary, non-Gaussian and colored hosts.

The chapter is organized as follows. The LTI filtering attack is defined in Section 5.1 and an overview of the discrete Fourier transform - rational dither modulation is given in

Section 5.2. In Section 5.3 the performance analysis of DFT-RDM is extended to non-white Gaussian hosts, while the proposed extension is described in Section 5.4. Finally, some concluding remarks are given in Section 5.5.

5.1 LTI filtering attack

In multimedia processing, probably the most common operation is the linear-time-invariant (LTI) filtering. Within the wide class of LTI filtering, we can recall as some examples equalization for audio signals, denoising, deblurring and interpolation for images. In the context of data hiding, LTI filtering can be considered an unintentional attack to robustness which impairs the watermark retrieving.

LTI filtering attack is here defined as $z_l = y_l * h_l$, where $*$ denotes convolution and h_l is the impulse response of a real-valued LTI filter. As customary in the thesis, y_l denotes the l th watermarked host sample and z_l denotes the l th attacked host sample. For sake of simplicity, the cursor of the attack filter (i.e., the largest magnitude coefficient) is assumed to be located at the origin. No other attacks such as noise addition are taken into account in the LTI filtering attack as it is defined in Fig. 5.1, due to the fact that the host-dependent noise produced by filtering is considered the dominant impairment. However, additive noise on the channel could be easily incorporated in a subsequent improved model of the attack.

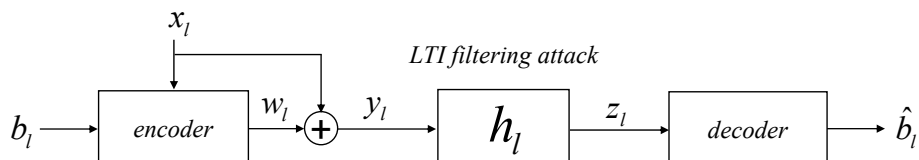


Figure 5.1: LTI filtering attack channel.

In spite of the simplicity and of the pervasiveness of the LTI filtering attack, existing quantization-based data hiding methods have not been designed to survive such attacks. To give an idea of the effect of filtering on decoding error probability, consider the outcome of the following experiment on binary time-domain DM when the host signal is white, DWR is approximately equal to 25 dB and the watermarked signal is attacked by an ideal lowpass filter of variable cutoff frequency. To have a decoding error probability of 0.5 it is sufficient a cutoff frequency equal to 0.99π rad, while, increasing the cutoff frequency to

0.999 π rad, the BER is still 0.11 [105]. The dramatic effect of this attack is as much evident as the distortion introduced on the signal is expected to be perceptually unnoticeable, due to the tiny fraction of bandwidth that is thrown away.

Also LTI filtering attack can be considered belonging to the class of desynchronization attacks, since the distortion error due to filtering moves the watermarked signal away from the correct quantization centroids and a dramatic increase of the decoding error probability is produced. Recalling that spread spectrum-based methods are quite easily made robust to filtering attack, as it has been discussed in Section 2.5.1, they are quite straightforwardly used in real applications. Consequently, in order to develop practical data hiding applications based on QIM, achieving robustness against filtering is an essential requirement.

In contrast to the performance of quantization-based methods against AWGN attacks, where it has been proven that the host signal power is virtually irrelevant for most cases of practical interest (see Section 3.1), the distortion introduced by filtering attack is as more dangerous as host power is larger. In fact distortion error due to filtering increases with the host signal power, so that, for a fixed watermark power, higher BER are expected for larger host power. In [105] the distortion error due to the filtering is referred as filtered-host interference (FHI) and the attack strength can be measured by the watermark-to-interference ratio (WIR), which is defined as the power ratio of the watermark signal and of the FHI. The filtered-host interference can be easily identified by writing the l th received sample as

$$z_l = y_l * h_l = h_0 y_l + \sum_{i \neq 0} h_i y_{l-i} \approx h_0 y_l + \sum_{i \neq 0} h_i x_{l-i} \quad (5.1)$$

where the approximation is valid under the assumption of large DWR. The FHI corresponds to the term $\sum_{i \neq 0} h_i x_{l-i}$. Assuming for simplicity $h_0 = 1$, the WIR is then given by

$$WIR \triangleq \frac{\sum_{l=1}^M E [W_l]^2}{\sum_{l=1}^M \left[\sum_{i \neq 0} h_i X_{l-i} \right]^2} \quad (5.2)$$

under the hypothesis of i.i.d. host samples, we have that the Watermark to Interference Ratio (WIR) is inversely proportional to the DWR. The main consequence of this dependence is that, if the DWR is increased to make the watermark less perceptible, the WIR and the robustness to FHI will be correspondingly decreased.

As it is proposed in [94], a possible approach to counteract the filtering attack is to estimate the filter, so that it can be equalized prior to the hidden information decoding. The estimation capability requires either transmitting pilot signals or dealing with a very large number of samples at the decoder to blindly estimate the filter. As a consequence, this approach seems only feasible for relatively simple filters and for a decoder with some prior knowledge about the attack filter. In fact, the data hiding method proposed in [94] is able to cope with a two-band filter in which the high-frequency band is amplitude-scaled, but the decoder is assumed to know the filter coefficients except for a gain factor. Then, at the decoder, the gain in high-frequency band is estimated using the maximum-likelihood criterion. The drawbacks of this approach are the large number of samples that are needed to accurately estimate the high-frequency gain and the hypothesis on the knowledge of the filter, which may be not realistic in some practical scenarios.

5.2 Discrete Fourier transform RDM

In [105] a new algorithm is proposed for constructing a quantization-based data hiding scheme robust to LTI filtering without assuming any prior knowledge about the attack filter. The key idea behind this method is to combine the gain-invariance property of RDM and the convolution theorem [133], that allows writing in the Fourier domain the output of the filter as the multiplication between the input signal and the filter response. As a consequence, at least in principle, linear filtering can be addressed by applying RDM in the Fourier domain. The method in [105] is named discrete Fourier transform - rational dither modulation (DFT-RDM), due to the fact that in a real application DFT-RDM uses the discrete Fourier transform in a block-by-block basis instead of the full-sequence Fourier transform.

As in the previous chapters, the host is assumed to be arranged in a 1-D vector \mathbf{x} , however the analysis of DFT-RDM can be easily extended to 2-D host even if it is not pursued here. Heretofore, to denote a variable in the DFT domain, the tilde will be used, so that the random variable $\tilde{X}_{m,k}$ is the k th coefficient of the DFT computed on the m th block of the host signal. Similarly, if h_l is the impulse response of a real-valued LTI filter, $H(e^{j\omega})$ denotes its Fourier transform so that we have $\tilde{h}_k \triangleq H(e^{j2\pi k/N})$.

As it is stated above, DFT-RDM uses discrete Fourier transform in a block-by-block basis instead of the full-sequence Fourier transform. A former reason for this choice is that

working on full-sequence is impractical due to computational complexity, the latter is that RDM requires a vector of past samples to feed the function $g(\cdot)$, that cannot be available in a full-sequence fashion. The practical solution is then to operate in a block-by-block basis using the DFT as an approximation of the Fourier transform. But in this practical framework the exact multiplication in the DFT domain would only be achieved with a circular convolution between the watermarked signal and the filter, whereas the filtered signal is obtained through an ordinary convolution, by definition. As a consequence, the effect of filtering on each DFT channel cannot be modeled by a pure scaling, but a host-dependent error has to be considered too. It is worth noting that well-known techniques, such as zero padding and cyclic prefixing [62], which are usually adopted to convert ordinary convolution into circular one, cannot be used for the analogous problem in data hiding context, because they would heavily distort the host signal.

Assuming non-overlapping DFT blocks of length N , let \mathbf{x}_m be the m th block of the host signal and $\tilde{x}_{m,k}$ the k th coefficient of the DFT of such block, which is computed as

$$\tilde{x}_{m,k} = \sum_{l=0}^{N-1} x_{m,l} \exp\left(-j\frac{2\pi k l}{N}\right) \quad (5.3)$$

The m th block of information bits $b_{m,k}$ is embedded into the absolute value of the DFT coefficients, taking care in preserving the symmetry of the DFT for real signals. Essentially, on each of the first $N/2 + 1$ discrete frequencies an RDM-like channel is constructed so that the absolute value of the watermarked signal is

$$|\tilde{y}_{m,k}| = g(\tilde{\mathbf{y}}_{m-1,k}) Q_{b_{m,k}} \left(\frac{|\tilde{x}_{m,k}|}{g(\tilde{\mathbf{y}}_{m-1,k})} \right) \quad (5.4)$$

where $\tilde{\mathbf{y}}_{m-1,k} \triangleq (\tilde{y}_{m-1,k}, \tilde{y}_{m-2,k}, \dots, \tilde{y}_{m-L,k})^T$ and $0 \leq k \leq N/2$. The phase of $\tilde{y}_{m,k}$ is set equal to the phase of $\tilde{x}_{m,k}$ so that the embedding distortion is minimized; in order to preserve symmetry, the remaining DFT coefficients are updated according to the rule $\tilde{y}_{m,k} = \tilde{y}_{m,N-k}^*$ for $N/2 + 1 < k < N$, where the superscript $*$ denotes the complex conjugate. The watermarked signal is then mapped back into the original domain through a non-overlapping block-by-block inverse DFT of the marked coefficients:

$$y_{m,l} = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{y}_{m,k} \exp\left(j\frac{2\pi l k}{N}\right) \quad (5.5)$$

5.2 Discrete Fourier transform RDM

We want to remark here that DFT-RDM transmits $N/2 + 1$ bits per DFT block of size N , so that an approximate rate of 1/2 bits per host sample is achieved, that is half of that attainable with time-domain binary RDM. This payload loss is due to choice of embed the watermark only in the magnitude in the DFT domain, which is motivated in [105].

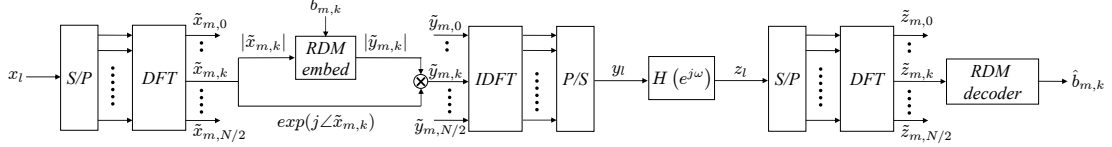


Figure 5.2: Block scheme of the whole embedding/decoding chain for DFT-RDM.

At the decoder, with \mathbf{z}_m denoting the m th block of the received signal, the relative DFT coefficients are computed and the standard RDM decoder is then applied to each discrete frequency to estimate the embedded information bits. The whole embedding/decoding block scheme is shown in Fig. 5.2.

Assuming $z_l = y_l * h_l$, under the hypothesis of N sufficiently large to approximate an ordinary convolution, we have $\tilde{z}_{m,k} \approx \tilde{h}_k \tilde{y}_{m,k}$, from which the RDM decoder is able to recover the correct information bits. However this approximation is more accurate for larger value of N and smoother frequency response of the attack filter.

Due to the orthogonality of the DFT, the DWR in the DFT domain is expected to be identical to that in the time domain. Hence DFT-RDM inherits from the standard RDM the relations between quantization step-size, power of the watermark signal and DWR.

As it has been widely discussed in Section 3.3, for memory size sufficiently large and DWR large enough to have an approximately flat host pdf in each quantization bin, time-domain RDM has data-to-watermark ratio equal to

$$DWR = \frac{3\sigma_x^2}{\Delta^2 M_{xp}^{2/p}} \quad (5.6)$$

that for $p = 2$ in the $g(\cdot)$ function can be written as $DWR = 3/\Delta^2$. In DFT-RDM, due to the orthogonality of the DFT, it can be observed that the relation between DWR and Δ is not changed by the DFT itself, so that the same DWR is obtained by applying RDM with the same quantization step-size in time domain or in DFT domain, which is computed inverting eq. (5.6), given the target DWR.

The above relations hold in the time domain, but we are also interested in deriving the per-DFT channel DWR. In DFT domain the variance of the random variable modeling the host samples on the k th channel easily follows from (5.3):

$$\sigma_{\tilde{X}}^2(k) = N\sigma_x^2 \quad (5.7)$$

Recalling the formula of the embedding distortion for RDM in eq. (3.30) for $p = 2$ in the g function, the per-channel watermark signal power is

$$\sigma_{\tilde{W}}^2(k) = \frac{\Delta^2}{3}\sigma_{\tilde{X}}^2(k) \quad (5.8)$$

where the quantization step-size Δ is set to have a watermarked signal with the desired DWR according to (5.6). In this way, it can be noticed that in each DFT frequency the per-channel DWR is equal to $\sigma_{\tilde{X}}^2(k)/\sigma_{\tilde{W}}^2(k) = 3/\Delta^2$, which corresponds to the overall DWR in time-domain.

5.2.1 Analytical derivation of the bit error probability

In this Section the analytical derivation of the per-channel bit error probability given in [105] is presented. The upcoming analysis is developed for i.i.d. zero-mean white Gaussian host samples with variance σ_x^2 . In DFT-RDM, a different error probability is expected to be measured on each DFT channel depending strictly on the filter h_l and, to this aim, the effects of the filtered-host interference have to be modeled in the DFT domain. Due to the effects of the circular convolution, the random variable representing the k th received DFT coefficient can be written as

$$\tilde{Z}_{m,k} = \tilde{h}_k \left(\tilde{X}_{m,k} + \tilde{W}_{m,k} + \tilde{N}_{m,k} \right) \quad (5.9)$$

where $\tilde{N}_{m,k}$ models the deviation from a pure multiplication (which would correspond to full-length DFTs) and so it will be referred to as *per-channel multiplication error*. According to the analysis in [105], when $H(z)$ is an FIR filter with no zeros on the unit circle, $\tilde{N}_{m,k}$ is a zero-mean complex Gaussian random variable asymptotically independent of $\tilde{X}_{m,k}$ and $\tilde{W}_{m,k}$.

Under the hypothesis of large DWR and using the filter-bank interpretation of the DFT [133], this term can be expressed as

$$\tilde{N}_{m,k} = (X_l + W_l) * f_{l,k}|_{l=mN+N-1} \approx X_l * f_{l,k}|_{l=mN+N-1} \quad (5.10)$$

where $f_{l,k}$ is given by

$$f_{l,k} \triangleq \left(\frac{h_l}{\tilde{h}_k} - \delta_l \right) * \phi_{l,k}^*, \quad k = 0, 1, \dots, N-1 \quad (5.11)$$

with δ_l denoting the Kronecker's delta. By definition, $\phi_{l,k} \triangleq v_l \exp(-j2\pi lk/N)$ for $l, k = 0, \dots, N-1$ and zero otherwise and $\phi_{l,k}$ represents the impulse response of the k th DFT basis function multiplied by a window $\mathbf{v} = (v_0, v_1, \dots, v_{N-1})^T$ whose purpose will be made clear shortly. Hence, from (5.10) and (5.11) it can be seen that the per-channel multiplication error is strictly dependent on both the filter coefficients h_l and the host signal.

Let $\{X_l\}$ be a zero-mean white process with variance σ_x^2 , then the process $\{\tilde{X}_l * f_{l,k}\}$ can be assumed stationary as discussed in [105], so that $\tilde{N}_{m,k}$ on the k th DFT channel will approximately have zero mean and variance

$$\sigma_{\tilde{N}}^2(k) = \frac{\sigma_x^2}{2\pi} \int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| 1 - \frac{H(e^{j(\omega+2\pi k/N)})}{H(e^{j2\pi k/N})} \right|^2 d\omega \quad (5.12)$$

where $\Phi(e^{j\omega})$ is the Fourier transform of the window \mathbf{v} .

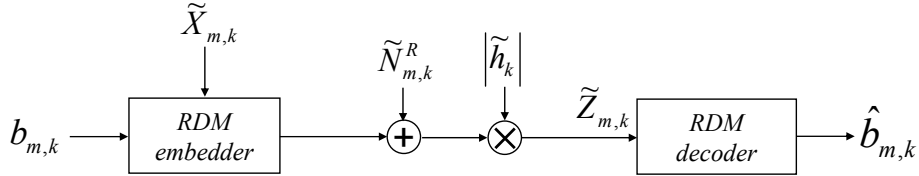


Figure 5.3: Equivalent model for the k th RDM-like channel in DFT-RDM.

From the knowledge of the distribution of $\tilde{N}_{m,k}$, its effect on the RDM-like channel constructed on the k th discrete frequency can be evaluated. Observing that the decisions at the decoder are based on the magnitude of $\tilde{Z}_{m,k}$ while $\tilde{N}_{m,k}$ is a complex random variable, the per-channel multiplication error is modeled by a white-noise real process $\{\tilde{N}_{m,k}^R\}$. The equivalent RDM-like channel is depicted in Fig. 5.3. Since the marginal pdf $f_{\tilde{N}_{m,k}^R}(\tilde{n}_{m,k}^R)$ cannot be obtained in closed form, it has been defined using the auxiliary random variable $\Theta \sim \mathcal{U}[0, 2\pi)$. Hence, we have

$$f_{\tilde{N}_{m,k}^R|\Theta}(\tilde{n}_{m,k}^R|\theta) = \mathcal{N}\left(0, \sigma_{\tilde{N}}^2(k, \theta)\right) \quad (5.13)$$

with $\sigma_{\tilde{N}}^2(k, \theta)$ computed as described in [105]. In essence, for each $\Theta = \theta$ the basic RDM channel is obtained, where the decoder is invariant to $|\tilde{h}_k|$ and the noise is zero-mean Gaussian, so that the RDM error probability in (3.31) can be used to predict the per-channel BER for the particular value assumed by Θ . As a consequence, assuming large L in $g(\cdot)$ function, the decoding error probability in the k th DFT channel for a given $\Theta = \theta$ results in

$$P_e(k, \theta) = P_{e,RDM} \left(L, \frac{\Delta \sigma_{\tilde{X}}(k)}{\sigma_{\tilde{N}}(k, \theta)} \right) \quad (5.14)$$

where $P_{e,RDM}(L, s)$ denotes the bit-error probability of classical RDM defined in (3.31), with s the effective signal-to-noise ratio. To obtain the overall error probability $P_e(k)$ for DFT-RDM in the k th DFT channel, the average of $P_e(k, \theta)$ with respect to θ has to be computed, so that we have

$$P_e(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{e,RDM} \left(L, \frac{\Delta \sigma_{\tilde{X}}(k)}{\sigma_{\tilde{N}}(k, \theta)} \right) d\theta \quad (5.15)$$

As it will be proven by experimental results, eq. (5.15) provides an accurate prediction of the per-channel error probability for DFT-RDM and LTI filtering attack.

An upper bound for the bit-error probability was also provided in [105]. Since the bound $\sigma_{\tilde{N}}(k, \theta) \leq \sigma_{\tilde{N}}(k)$ is always verified for every θ , the upper bound can be computed by substituting $\sigma_{\tilde{N}}(k, \theta)$ in (5.15) by the standard deviation of the per-channel multiplication error $\sigma_{\tilde{N}}(k)$ defined in 5.12, so that we have

$$\begin{aligned} P_e(k) &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{e,RDM} \left(L, \frac{\Delta \sqrt{N} \sigma_x}{\sigma_{\tilde{N}}(k)} \right) d\theta = \\ &= P_{e,RDM} \left(L, \frac{\Delta \sqrt{N} \sigma_x}{\sigma_{\tilde{N}}(k)} \right) \end{aligned} \quad (5.16)$$

Refer to [105] for more details on the analysis.

The upper bound formula allows to link directly the per-channel WNR and the per-channel bit-error probability. It is very important to remark that while the WNR is usually defined as the ratio between the power of the watermark signal and the power of the attack noise, since in our framework the only impairment is the filtering, we will define the the per-channel WNR as the ratio between the power of the watermark signal and that of the multiplication-error for each frequency channel

$$\text{WNR}(k) \triangleq \frac{E \left[|\tilde{Y}_{m,k} - \tilde{X}_{m,k}|^2 \right]}{\sigma_{\tilde{N}}^2(k)} = \frac{\sigma_{\tilde{W}}^2(k)}{\sigma_{\tilde{N}}^2(k)} \quad (5.17)$$

where $E[\cdot]$ denotes the statistical expectation.

According to (5.8), we can substitute $\sigma_{\tilde{X}}(k) = (\sqrt{3}/\Delta)\sigma_{\tilde{W}}(k)$ into (5.16) so that the upper bound results in

$$\begin{aligned} P_e(k) &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{e,RDM} \left(L, \frac{\sqrt{3}\sigma_{\tilde{W}}(k)}{\sigma_{\tilde{N}}(k)} \right) d\theta = \\ &= P_{e,RDM} \left(L, \sqrt{3 \text{WNR}(k)} \right) \end{aligned} \quad (5.18)$$

as it is observed in [58].

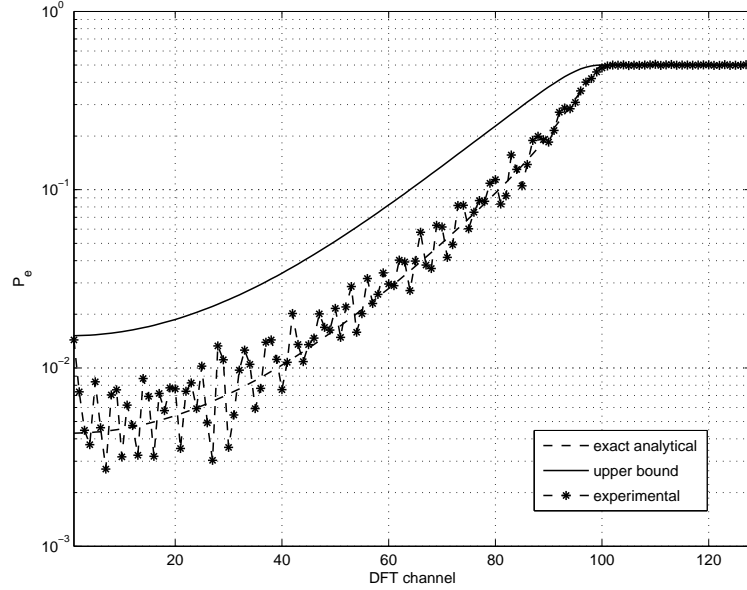


Figure 5.4: BER versus DFT channel for an ideal low-pass filter with $\omega_c = 0.8\pi$ rad, $N = 256$ and DWR = 25 dB.

As an example to exhibit the accurateness of the performance analysis developed above, in Fig. 5.4 the experimentally evaluated decoding error probabilities versus DFT channel are compared with the analytically computed ones using eqs. (5.15) and (5.16). The results shown here are obtained for white host, DWR was set to 25 dB and the memory of the $g(\cdot)$ function is $L = 100$. The length of the block used in DFT-RDM is $N = 256$

and the attack filter is an ideal low-pass filter with cutoff frequency $\omega_c = 0.8\pi$ rad. From inspection of Fig. 5.4 it can be noticed the excellent matching between the experimental BERs and the analytical error probability computed according to eq. (5.15).

5.2.2 Improvements

To reduce the error probability, in [105] two improvements have been proposed: windowing and spreading. The former entails multiplying the block \mathbf{x}_m by a properly designed window \mathbf{v} while the latter amounts to adding M length- N block and then computing the size- N DFT.

Not all the windows $\mathbf{v} = (v_0, v_1, \dots, v_{N-1})^T$, where N is the DFT size, can be used to improve DFT-RDM, since \mathbf{v} is constrained to the class of windows for which $(v_l)^{-1}$ exists for all $l = 0, 1, \dots, N - 1$. This property is required to guarantee that the watermarked signal is perfectly equal to the original host signal in case of the RDM quantization step-size is $\Delta = 0$, which is equivalent to not insert the watermark signal. As a consequence of the previous considerations, eqs. (5.3) and (5.5) are updated becoming respectively

$$\tilde{x}_{m,k} = \sum_{l=0}^{N-1} v_{l-mN} x_{m,l} \exp\left(-j \frac{2\pi k l}{N}\right) \quad (5.19)$$

and

$$y_{m,l} = v_{l-mN}^{-1} \frac{1}{N} \sum_{k=0}^{N-1} \tilde{y}_{m,k} \exp\left(j \frac{2\pi l k}{N}\right) \quad (5.20)$$

According to eq. (5.19) and eq. (5.7) the variance of the k th coefficient in the DFT domain is equal to $\sigma_{\tilde{X}}^2(k) = \|\mathbf{v}\|^2 \sigma_x^2$. In [105] the resulting DWR has been computed as

$$DWR = \frac{3N^2}{\Delta^2 \|\mathbf{v}\|^2 \|\mathbf{v}^{-1}\|^2} \quad (5.21)$$

It can be noticed that all the relations that hold when no window is applied are particular cases of those presented in this section for $\mathbf{v} = 1$. Moreover, being $\|\mathbf{v}\|^2 \|\mathbf{v}^{-1}\|^2 \geq N^2$ as a consequence of Cauchy-Schwarz inequality, for a given quantization step-size Δ the DWR is decreased using a window different from $\mathbf{v} = 1$.

In [105] the influence of \mathbf{v} has been included in the formulas of the per-channel decoding probability, allowing the comparison of the performance of different windows. Moreover, departing from the knowledge of the probability of error for each DFT channel and from

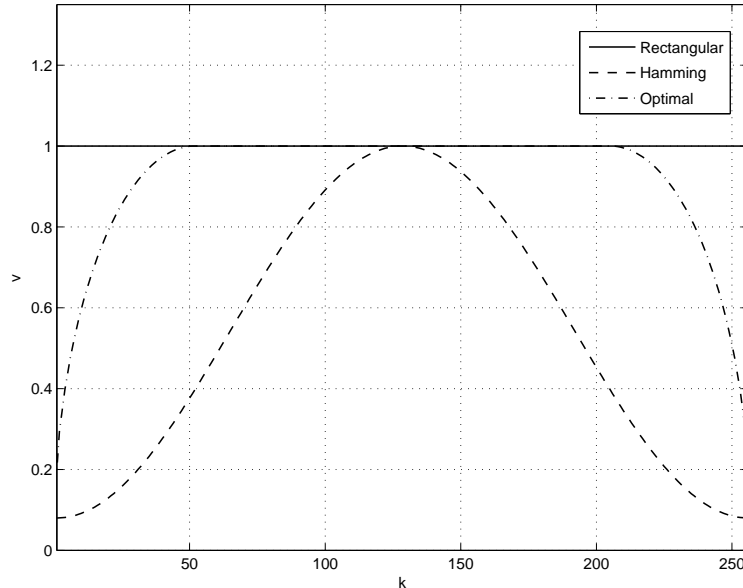


Figure 5.5: Comparison of optimal window with rectangular and hamming window.

eq. (5.21), it has been designed the optimal window to minimize the variance $\sigma_N^2(k)$ of the per-channel multiplication error. The optimal window, computed according to [105], is depicted in Fig. 5.5, where it is compared with rectangular and hamming window.

The robustness improvement due to windowing is clearly appreciable in Fig. 5.6, where are compared the analytical per-channel error probabilities computed for DFT-RDM using rectangular, Hamming or optimal window, as it is defined in [105], in case of an ideal low-pass filter as attack filter. We want to remark here that the drawback of the windowing improvement is an increased peak-to-average distortion, which can be controlled in the window design process.

The spreading improvement lies on summing M length- N blocks \mathbf{x}_m and then applying the DFT-RDM embedding on the so composed array. Spreading can be intuitively seen as a practical way to increase the DFT effect on the signal without suffering the increase of computational complexity. As a consequence, by spreading, the robustness against filtering is increased at the price of payload reduction by a factor of M .

For sake of simplicity we assume here that rectangular window is applied, so that the kM th coefficient of the p th block of length- N DFT of the host signal results in

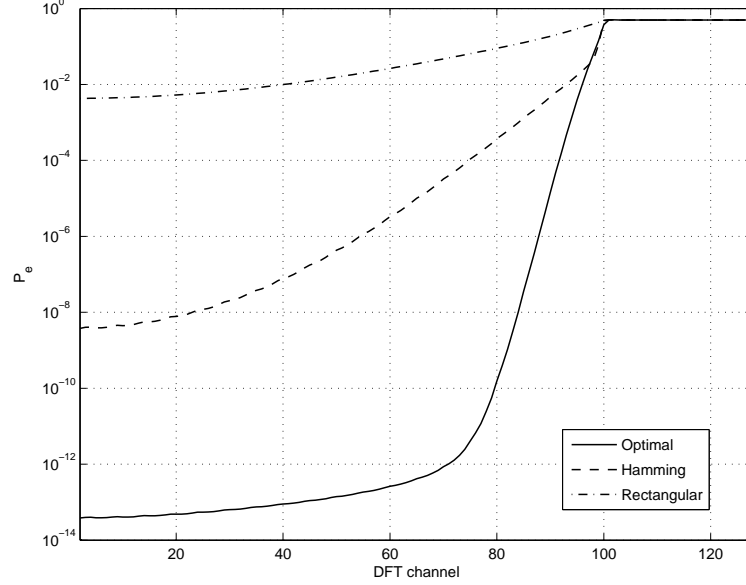


Figure 5.6: BER vs. DFT channel for DFT-RDM with $N = 256$ and $DWR = 25$ dB using different windows against an ideal low-pass filter with $\omega_c = 0.8\pi$ rad.

$$\tilde{x}_{p,kM} = \sum_{l=0}^{N-1} \left(\sum_{m=pM}^{pM+M-1} x_{m,l} \right) \exp \left(-j \frac{2\pi k}{N} l \right) \quad k = 0, 1, \dots, N-1 \quad (5.22)$$

Then RDM is applied to each $\tilde{x}_{p,kM}$ to obtain the corresponding watermarked coefficient $\tilde{y}_{p,kM}$.

Obviously the sum of M length- N blocks is not an operation that can be losslessly reverted, but it is sufficient to compute the additive watermark signal in time domain to embed the desired information into the host signal. Hence, given the watermark samples in DFT domain $\tilde{w}_{p,kM} = \tilde{y}_{p,kM} - \tilde{x}_{p,kM}$, the watermark samples in time domain are obtained through IDFT

$$w_l = \frac{1}{MN} \sum_{k=0}^{N-1} \tilde{w}_{p,kM} \exp \left(j \frac{2\pi l}{N} k \right) \quad pMN \leq l \leq pMN + MN - 1 \quad (5.23)$$

where the total time-domain watermark is equally shared among the M blocks and the watermarked signal is simply given by $y_l = x_l + w_l$.

5.2 Discrete Fourier transform RDM

In [105] the analytical results in Section 5.2 have been extended to the case when spreading is applied considering a length- MN DFT and taking into account the DFT coefficients indexed by kM , for $k = 0, 1, \dots, N/2 + 1$. Thus the variance of the k th coefficient in the DFT domain is $\sigma_{\tilde{X}}^2(k) = MN\sigma_x^2$ and the DWR is updated accordingly. The influence of spreading on decoding error probability has been also evaluated and it has been verified that the upper bound of per-channel bit-error rate is given by

$$P_e(k) \leq P_{e,RDM} \left(L, M\sqrt{3\text{WNR}(k)} \right) \quad (5.24)$$

from which it can be inferred that spreading provides a gain of $10 \log_{10}(M^2)$ dB in the effective WNR.

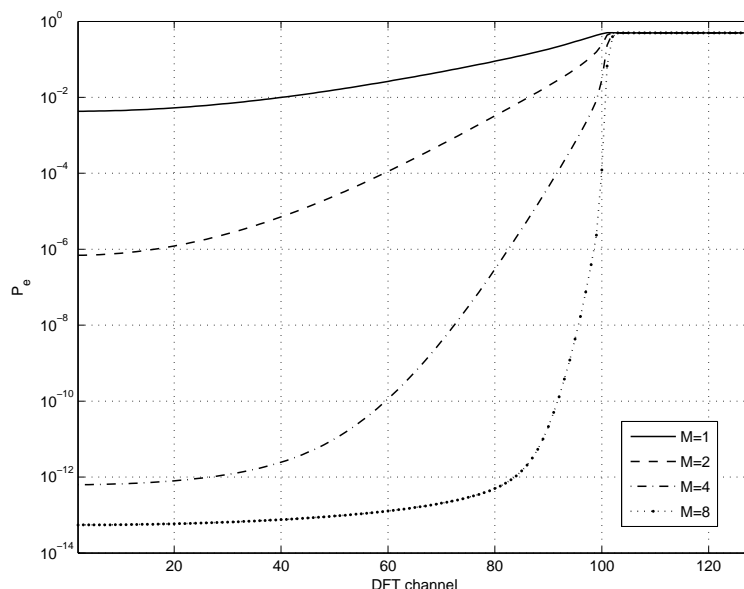


Figure 5.7: BER vs. DFT channel for DFT-RDM with $N = 256$ and $\text{DWR} = 25$ dB using different spreading factors against an ideal low-pass filter with $\omega_c = 0.8\pi$ rad.

In Fig. 5.7 are shown the analytical per-channel error probabilities computed for DFT-RDM using different spreading factors in case of an ideal low-pass filter as attack filter. Here the decrease of the bit-error rate given by spreading is evident. Eventually, spreading can be considered as a recommended technique for increasing the WNR at the price of reducing the data rate. As final remark, we observe that spreading can be used in combination with windowing to achieve better performance.

5.3 Performance analysis for colored Gaussian hosts

Analytical and experimental results in [105] demonstrate the effectiveness of DFT-RDM when the attack consists of LTI filtering and that a high data rate can be achieved for white Gaussian hosts. In fact the developed analysis, which has been presented in Section 5.2, is focused uniquely on white Gaussian hosts, providing an accurate prediction of the behavior of DFT-RDM for this class of host signals. However in [105] DFT-RDM has been tested for real audio tracks as host signals and in the experimental results it is shown a severe loss of performance when DFT-RDM is applied to these nonstationary, non-Gaussian and colored hosts. As an example in Fig. 5.8 the analytically computed per-channel bit-error probabilities of DFT-RDM for white Gaussian host signal are compared with those experimentally evaluated for a real audio track with $N = 256$, $DWR \approx 25$ dB, spreading factor $M = 4$ and optimal window in case of an audio equalizer is used as attack filter. Here it is evident that, for the same system parameters, DFT-RDM gives unsatisfactory BERs for a nonstationary, non-Gaussian and colored host; in fact the measured overall error probability is $P_e \approx 0.15$ while for a white Gaussian host it is $P_e \approx 0.017$.

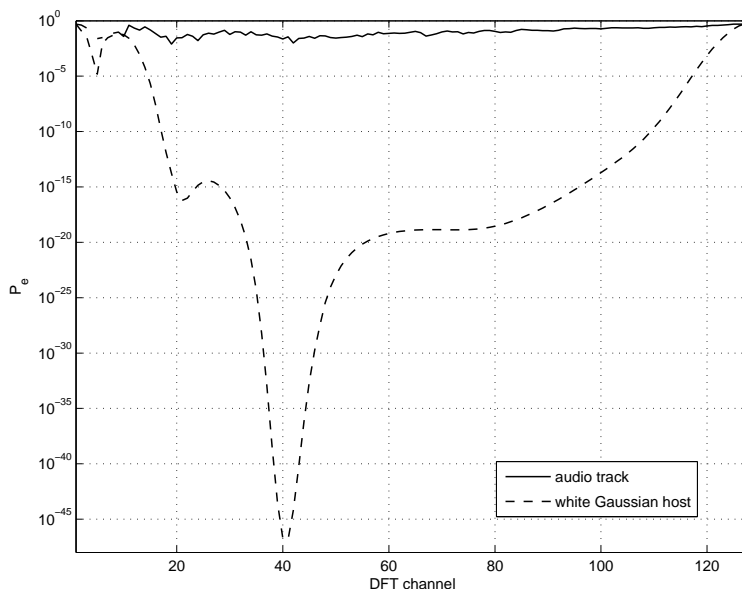


Figure 5.8: BER vs. DFT channel for DFT-RDM applied to different hosts with $N = 256$, $DWR \approx 25$ dB, spreading factor $M = 4$ and optimal window against an audio equalizer.

Unfortunately the analysis provided in [105] cannot give hints to understand the ex-

5.3 Performance analysis for colored Gaussian hosts

perimental results obtained for real audio tracks, since there the host is assumed to be a white Gaussian signal. To justify the experimental results obtained for audio signals, the performance analysis of DFT-RDM has been generalized to Gaussian non-white hosts modeling the colored signal with an autoregressive (AR) [133] random process. Hence, given a zero-mean white Gaussian host \mathbf{x}_0 with power spectral density (psd) $\sigma_{x_0}^2$, the colored host \mathbf{x} can be regarded to as the output of an all-pole filter $H_{AR}(z) = 1/A(z)$ excited by \mathbf{x}_0 . The host power spectral density can then be written as:

$$S_x(e^{j\omega}) = \frac{\sigma_{x_0}^2}{|A(e^{j\omega})|^2} \quad (5.25)$$

The idea is to work with a colored host whose psd resembles that of a generic audio signal, which typically has most of its power concentrated at lower frequencies. Heretofore for colored hosts we will assume an AR signal which models the spectral contents of this generic audio signal.

While in [105] per-channel multiplication error was characterized in the time domain, we pursue here a frequency-domain approach, which is needed to separate each RDM-like channel and that will lead to give a reasonable explanation of the behavior of DFT-RDM for non-white hosts. In fact the rationale for this behavior can be found in the inner working of DFT-RDM, which is essentially an RDM-like scheme for every DFT channel, and in the influence of a non-flat psd on the per-channel multiplication error.

Moreover, we will introduce the per-channel watermark-to-noise ratio, which has been defined in eq. (5.17), as a simple and intuitive measure to evaluate the reliability of each RDM-like channel, since WNR is directly related to the BER as described in eq. (5.18).

As a first step towards obtaining the per-channel WNR, the per-channel host power in the DFT domain has to be derived. To this aim, the filter-bank interpretation of the DFT [133] can be adopted, according to which it is possible to get

$$\tilde{X}_{m,k} = X_l * \phi_{l,k}^* |_{l=mN+N-1} \quad (5.26)$$

The variance of the zero-mean process $\tilde{X}_{m,k}$ is given by $\sigma_{\tilde{X}}^2(k) = E \left\{ \left| X_l * \phi_{l,k}^* \right|^2 \right\}$ and it can be computed by applying Parseval's relation, so that we have

$$\sigma_{\tilde{X}}^2(k) = \frac{\sigma_{x_0}^2}{2\pi} \int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| \frac{1}{A(e^{j(\omega+2\pi k/N)})} \right|^2 d\omega \quad (5.27)$$

5.3 Performance analysis for colored Gaussian hosts

Applying RDM to the k th DFT channel for a colored host and assuming $p = 2$ in the $g(\cdot)$ function, the per-channel watermark signal power defined in eq. (5.8) becomes

$$\sigma_{\tilde{W}}^2(k) = \frac{\Delta^2}{3} \sigma_{\tilde{X}}^2(k) = \frac{\Delta^2}{3} \frac{\sigma_{x_0}^2}{2\pi} \int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| \frac{1}{A(e^{j(\omega+2\pi k/N)})} \right|^2 d\omega \quad (5.28)$$

so that the per-channel watermark signal power is proportional to the per-channel host power due to the properties of RDM. As a consequence a larger watermark signal originates from those host DFT channels having stronger spectral contents. As an example, in the lower frequencies of an audio-like colored host, a larger per-channel watermark signal than the corresponding to higher frequencies is reasonably expected. This shaping of the per-channel watermark power alters the behavior of DFT-RDM with respect to that for white Gaussian hosts, where the per-channel watermark power is uniform as analyzed in [105].

On the other hand, the spectral shaping of the host influences also the per-channel multiplication error. Recalling (5.12) and assuming reasonably the stationarity of $\tilde{N}_{m,k}$, its variance can be now written as

$$\sigma_{\tilde{N}}^2(k) = \frac{\sigma_{x_0}^2}{2\pi} \int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| \frac{1}{A(e^{j(\omega+2\pi k/N)})} \right|^2 \left| 1 - \frac{H(e^{j(\omega+2\pi k/N)})}{H(e^{j2\pi k/N})} \right|^2 d\omega \quad (5.29)$$

Departing from the knowledge of the per-channel watermark signal power and of the variance of the per-channel multiplication error, the WNR can be computed on each DFT channel. The watermark-to-noise ratio can be useful to infer whether the RDM channel is able to correctly convey the information bits, because the decoding error probability for a DM-based embedding technique approaches 0.5 when the power of the watermark signal is approximately equal to that of the additive noise (see Section 3.1). Thus, the per-channel WNR is given by

$$\text{WNR}(k) = \frac{\frac{\Delta^2}{3} \int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| \frac{1}{A(e^{j(\omega+2\pi k/N)})} \right|^2 d\omega}{\int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| \frac{1}{A(e^{j(\omega+2\pi k/N)})} \right|^2 \left| 1 - \frac{H(e^{j(\omega+2\pi k/N)})}{H(e^{j2\pi k/N})} \right|^2 d\omega} \quad (5.30)$$

To intuitively understand the influence of the spectral shaping of the host on the WNR, it is useful to approximate the per-channel host power as

5.3 Performance analysis for colored Gaussian hosts

$$\sigma_{\hat{X}}^2(k) \approx N \frac{\sigma_{x_0}^2}{|A(e^{j2\pi k/N})|^2} \quad (5.31)$$

By this approximation, which is valid only in the case of a rectangular window, the effects of computing the DFT on finite-length blocks are neglected. Consequently, the WNR can be approximated as follows

$$WNR(k) \approx \frac{N\Delta^2/3}{\frac{1}{2\pi} \int_{-\pi}^{\pi} |\Phi(e^{j\omega})|^2 \left| \frac{A(e^{j2\pi k/N})}{A(e^{j(\omega+2\pi k/N)})} \right|^2 \left| 1 - \frac{H(e^{j(\omega+2\pi k/N)})}{H(e^{j2\pi k/N})} \right|^2 d\omega} \quad (5.32)$$

If the host signal is white, then the ratio $R_A(\omega, k) \triangleq |A(e^{j2\pi k/N}) / A(e^{j(\omega+2\pi k/N)})|$ is equal to 1 for every k and consequently $WNR(k)$ depends only on the attack filter; if the host is colored, this ratio is a function which has great variations for different channels k , thus affecting heavily $WNR(k)$.

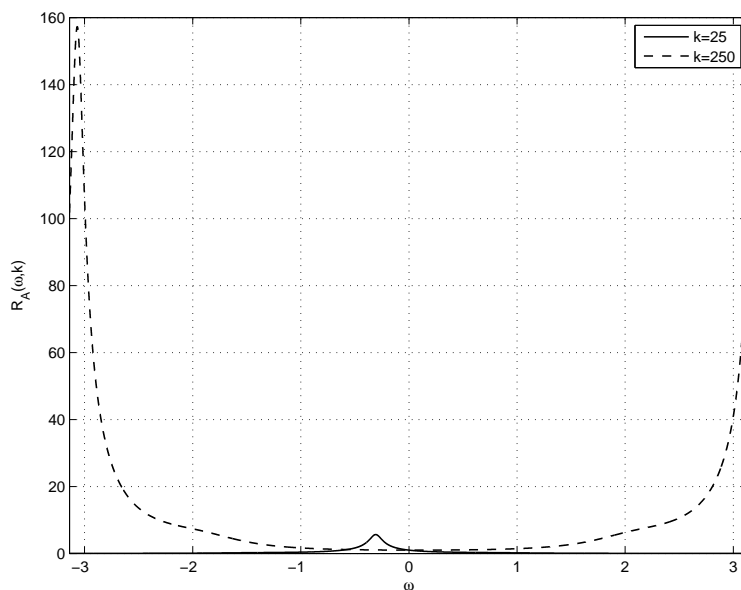


Figure 5.9: $R_A(\omega, k)$ versus discrete frequency for $k = 25$ and $k = 250$ ($N = 512$).

As shown in Fig. 5.9, because of the high-pass behavior of $A(z)$, for k corresponding to the high frequency channels, the function $R_A(\omega, k)$ takes values much larger than those corresponding to low frequencies. Therefore, the spectral shaping of the host yields less robustness in high frequency channels compared to low frequency channels; however,

5.3 Performance analysis for colored Gaussian hosts

strictly speaking, the per-channel WNR also depends on the attack filter, as is evident from (5.32).

As stated above, the per-channel WNR is a simple measure that can be used to have an idea of the reliability of the k th frequency channel in DFT-RDM. However we are also interested in predicting the performance of DFT-RDM in the sense of per-channel bit-error rate when the host is colored. In Section 5.2 it has been recalled the performance analysis developed for DFT-RDM in [105] relying on the results in [104], where the bit-error probability of an RDM channel is derived for i.i.d. host samples and additive noise independent of the host signal. If this analytical model is applied in case of colored hosts, the predicted error probabilities will be only an approximation of the actual BERs. The inaccuracy of the analytical model is expected to be noticeable for those DFT channels whose $\tilde{X}_{m,k}$ is more correlated with the neighboring channels; in this case, the per-channel multiplication error will increase due to the leakage from those host DFT coefficients at adjacent channels. To evaluate the correlation between the k th channel and the t th channel, the correlation coefficient $\rho_{k,t}$ can be employed. Using the approximate expression of the per-channel host power we can write

$$\begin{aligned} \rho_{k,t} &\triangleq \frac{E [\tilde{X}_k \tilde{X}_t^*]}{\left(E [|\tilde{X}_k|^2] E [|\tilde{X}_t|^2] \right)^{1/2}} \approx \\ &\approx \frac{|A(e^{j2\pi k/N})| |A(e^{j2\pi t/N})|}{N \sigma_{x_0}^2} E [\tilde{X}_k \tilde{X}_t^*] \end{aligned} \quad (5.33)$$

where the influence of the frequency response of the AR filter on the correlation coefficient is highlighted.

We want to remark that the analysis carried out here for DFT-RDM and colored hosts gives a first explanation of the experimental results that were given in [105] for DFT-RDM applied to audio signals.

5.3.1 Experimental results

Some experiments are here presented to validate the generalized analysis for DFT-RDM developed in the previous section. The presented experimental results are devoted to the comparison of the analytically predicted per-channel WNR with the experimentally evaluated measures. Moreover the comparison of the WNR per-DFT channel with the

5.3 Performance analysis for colored Gaussian hosts

experimentally and analytically computed bit-error rates allows to verify the suitability of the per-channel WNR as a measure of the reliability of each RDM-like channel.

In all the experiments the DWR was set to 25 dB, in the $g(\cdot)$ function the memory L was set to 100 and p was set to 2. An AR model with order $Q = 10$ is assumed in all the experiments. Unless otherwise specified, we assume that the DFT length is $N = 512$ and that neither spreading nor windowing are used.

The colored host signal is the output of an all-pole filter $1/A_{av}(z)$ whose coefficients have been obtained by AR modeling of several audio tracks in order to resemble the power spectral density of a generic (average) audio signal; Fig. 5.10 represents the magnitude of the frequency response of the filter $A_{av}(e^{j\omega})$ that has been used in the subsequent simulations.

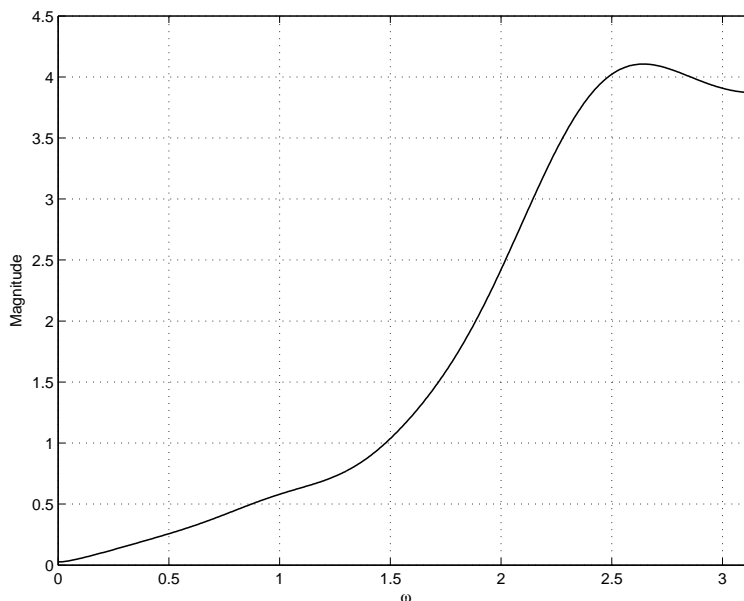


Figure 5.10: Magnitude of the frequency response of the filter $A_{av}(e^{j\omega})$ with order $Q = 10$.

As a preliminary test, the exactness of the analytically predicted per-channel watermark signal power has been verified. In Fig. 5.11 it is compared the experimental per-channel watermark signal power with the analytically computed values using the exact and approximated formulas. While the matching between the experimental results and the exact analytical values obtained substituting (5.27) in (5.8) is excellent, a mismatch in the high frequency channels is apparent when using in (5.8) the approximate formula

5.3 Performance analysis for colored Gaussian hosts

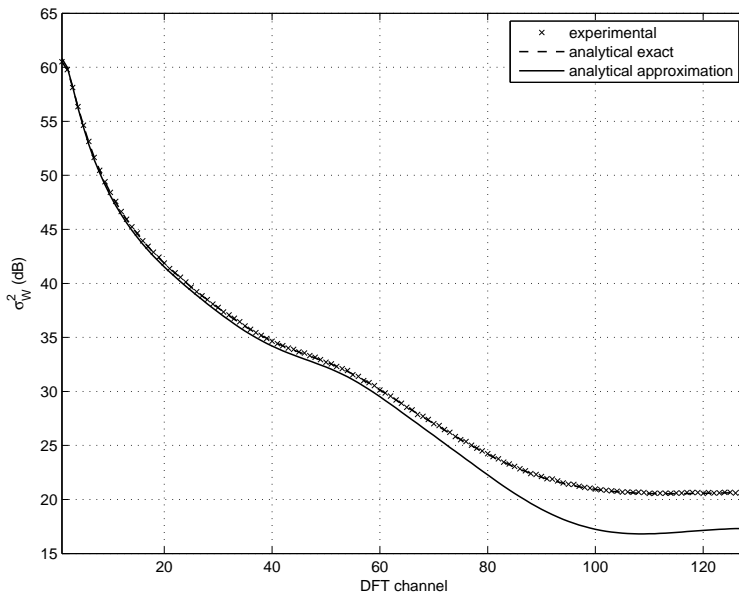


Figure 5.11: Per-channel watermark signal power in dB ($\sigma_{x_0}^2 = 1000$ and $N = 256$).

(5.31) for the per-channel host power.

According to the considerations drawn above, the analytical model developed in [105] to compute the per-channel decoding error probability is expected to be less accurate for those DFT channels whose $\tilde{X}_{m,k}$ is more correlated with the neighboring channels. Since the per-channel multiplication is essentially due to the leakage from those host samples at adjacent channels, the error will reasonably increase for more correlated channels. In order to verify the existing correlation between channels when the host signal is non-white, the magnitude of the correlation coefficient $|\rho_{k,t}|$ has been evaluated for some channels of the watermarked signal $\tilde{Y}_{m,k}$ according to eq. (5.33).

First, we plot in Fig. 5.12(a) the magnitude of the correlation coefficient for several DFT channels when the watermarked signal is white Gaussian. As it can be verified, the correlation between neighboring channels is very small for all k, t , with $k \neq t$. Obviously, for $k = t$ we have $\rho_{k,t} = 1$, since the correlation coefficient corresponds to the normalized autocorrelation.

In contrast, for a colored host the correlation coefficient is strictly dependent on the selected channels, as it is evident in Fig. 5.12(b). As expected from (5.33) for a high-pass filter $A(z)$, the correlation between two low frequency neighboring channels is quite small,

5.3 Performance analysis for colored Gaussian hosts

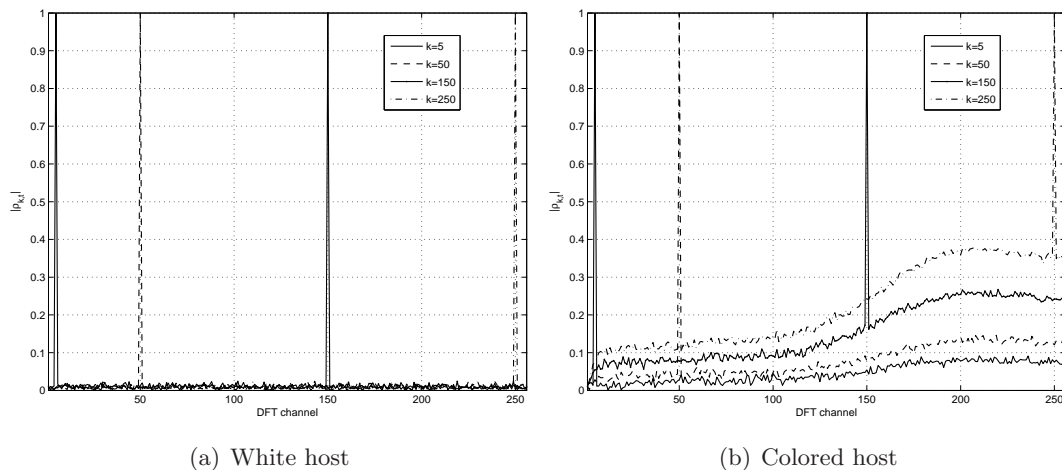


Figure 5.12: Magnitude of the correlation coefficient $|\rho_{k,t}|$ for white and colored watermarked signals evaluated at channels $k = 5$, $k = 50$, $k = 150$ and $k = 250$.

while it noticeably increases when neighboring higher frequency pairs are considered. Since the analytical results are less accurate when DFT channels become more correlated, we should expect worse predictions for high frequency channels in the subsequent experiments where always this AR model is considered.

After these preliminary experiments, the exactness of the extended analysis for DFT-RDM and colored Gaussian hosts has been verified attacking the watermarked host with different LTI filters.

The watermarking system has been firstly tested against an ideal low-pass filter with cut-off frequency $\omega_c = 0.8\pi$ rad, whose frequency response is shown in Fig. 5.13, while the results are presented in Fig. 5.14. Fig. 5.14(a) compares the experimentally evaluated WNR with the analytical WNR computed according to (5.30) and the analytical approximation of the WNR obtained from (5.32). It is worth noting that the WNR is much larger for the low frequency channels where the host power is also larger and the filter response is flat. In Fig. 5.14(b) the experimental BER is compared with the analytically derived BER and its upper bound, according to the formulas in [105] (here and in the following, the analytical BER is computed using the exact formula of the per-channel signal power); here we show only the range of channels having an experimental BER larger than 10^{-5} . From the comparison of Figs. 5.14(a) and 5.14(b) it can be verified that the error probability is approximately 0.5 for those DFT channels whose WNR is lower than 0 dB, as we have

5.3 Performance analysis for colored Gaussian hosts

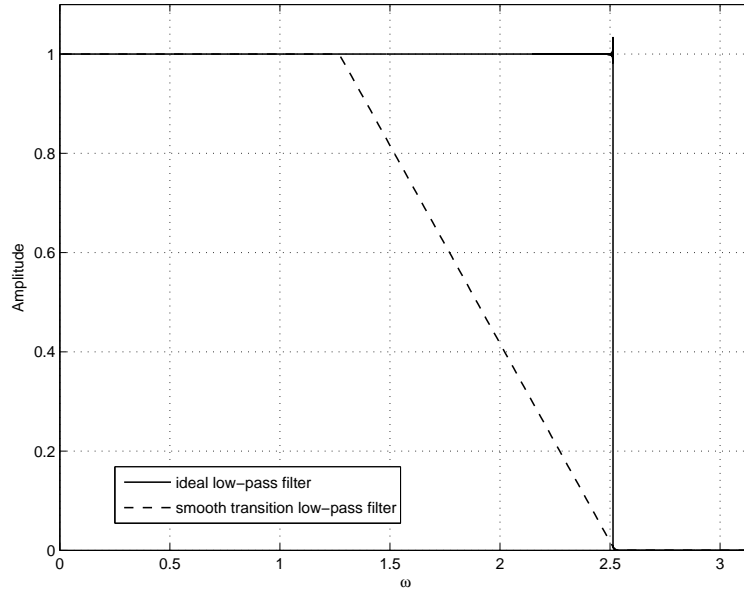


Figure 5.13: Magnitude of the frequency responses of the low-pass filters used in the experiments.

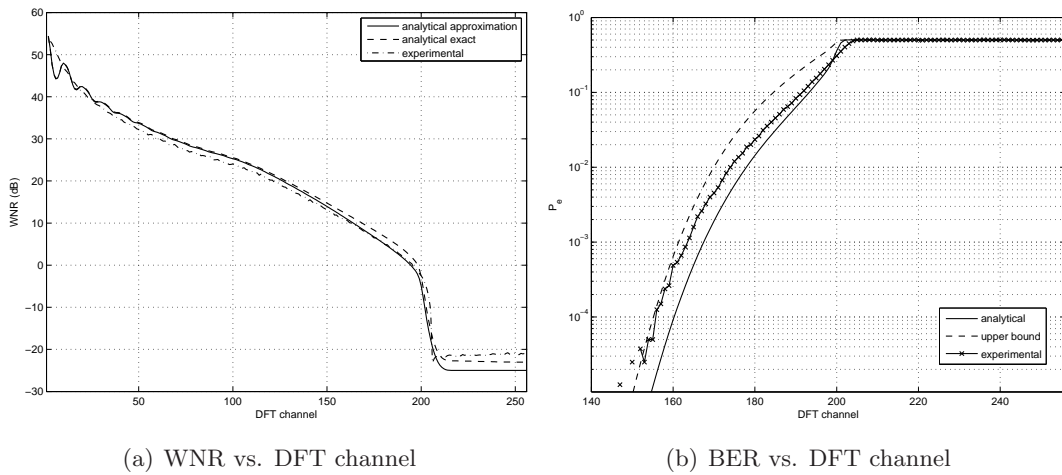


Figure 5.14: Analytical and experimental results for colored host and low-pass filter with $\omega_c = 0.8\pi$.

already discussed.

To understand how different AR models influence the WNR, the analytical WNR for the low-pass filter has been computed using (5.30) for different orders Q of the AR model.

5.3 Performance analysis for colored Gaussian hosts

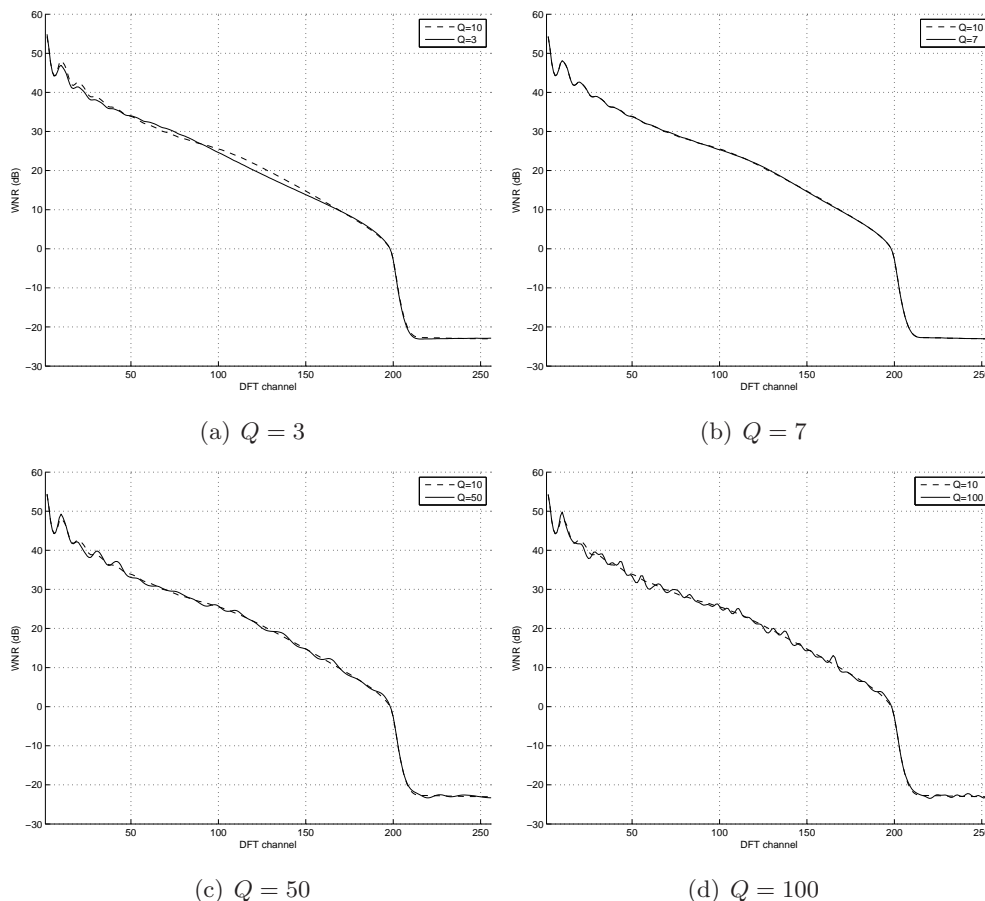


Figure 5.15: Analytical WNR versus DFT channel for different order AR filters.

In Fig. 5.15 the WNR for AR(10) is compared with the analytical WNR computed for AR(3), AR(7), AR(50) and AR(100). For $Q = 3$ the WNR is slightly lower than that of $Q = 10$, while for $Q = 7$ approximately the same WNR of $Q = 10$ is obtained. As the order of AR model increases, the WNR has more ripples but it has the same average trend of that for AR(10), as it is shown in Figs. 5.15(c) and 5.15(d). We conclude that the order of the AR model has little impact on the final results.

Then we have tested the watermarking system with a low-pass filter having passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad, with a smooth transition in the middle as it is shown in Fig. 5.13. The measured WNR and BER are presented in Fig. 5.16. In Fig. 5.16(a), where the comparison of the experimental WNR and the analytical one is shown, it can be noticed that they differ in the frequency range where the inter-channel correlation

5.3 Performance analysis for colored Gaussian hosts

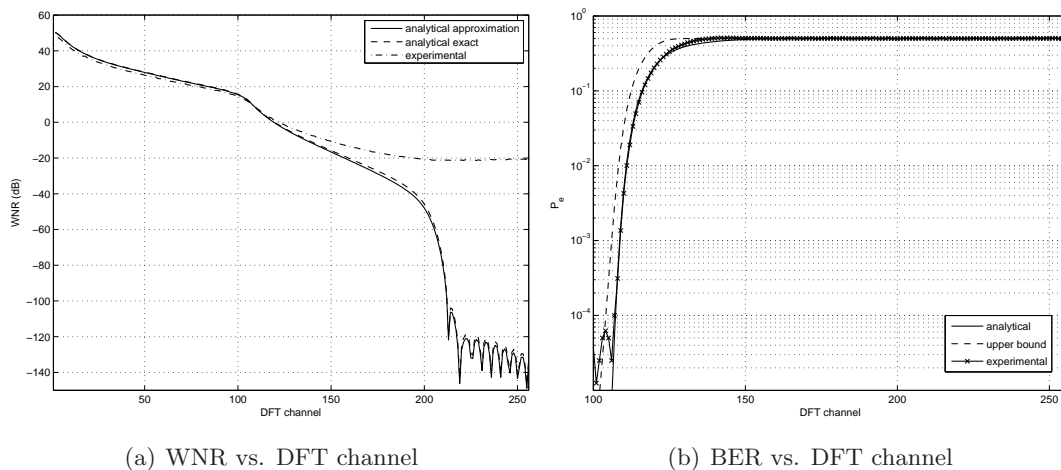


Figure 5.16: Analytical and experimental results for colored host and low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad.

is larger. In Fig. 5.16(b) the experimental BER is compared with the analytically derived BER and its upper bound (again only the range of channels having an experimental BER larger than 10^{-5} is shown). Here it can be seen that the analytical error probability matches the experimental one since all the channels with $P_e < 0.5$ are not in the range of high correlation. It is worth noting that the error probability is approximately 0.5 in the majority of channels belonging to the transition band, which is approximately between channels 102 and 204.

The performance DFT-RDM applied to a non-white host dramatically decrease when a different attack filter is used. In Fig. 5.18 the decoding error probabilities and the per-channel WNR are shown in the case of a band-pass attack filter, which has $|H(e^{j\omega})| = 1$ for $0.1\pi \leq \omega \leq 0.9\pi$ and $|H(e^{j\omega})| = 0.5$ for $0 \leq \omega \leq 0.05\pi$ and $0.95\pi \leq \omega < \pi$, with smooth connections in the transition bands. The magnitude of the frequency response of the considered band-pass filter is depicted in Fig. 5.17. We want to remark that attenuating or cutting away a little portion of the spectrum around the zero frequency is not unusual for audio signals, e.g. the analog phone signal is obtained by band-pass filtering between 300 Hz and 3400 Hz. In Fig. 5.18(a) the experimental per-channel WNR is compared with the analytical predictions. Here it can be noticed the excellent matching between the experimental measures and the analytically computed WNR in the middle frequencies, while the prediction is inaccurate at lower and higher channels where the

5.3 Performance analysis for colored Gaussian hosts

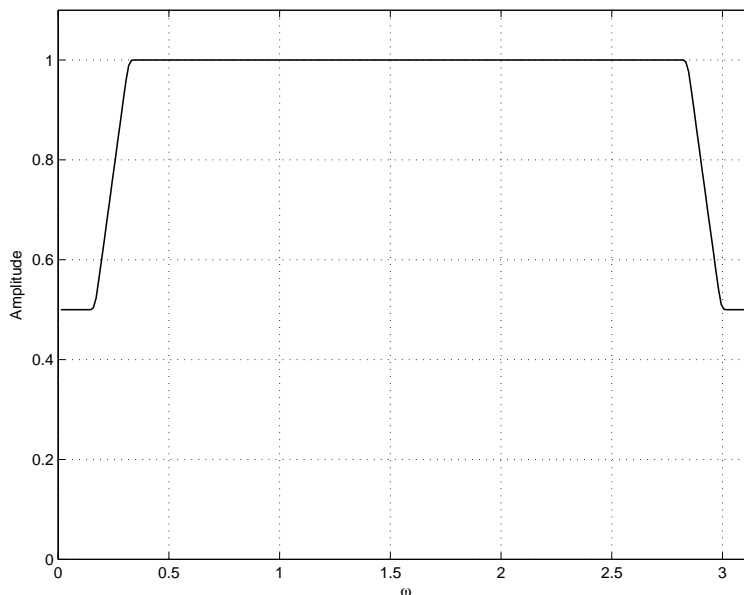


Figure 5.17: Magnitude of the frequency response of the band-pass filter used in the experiments.

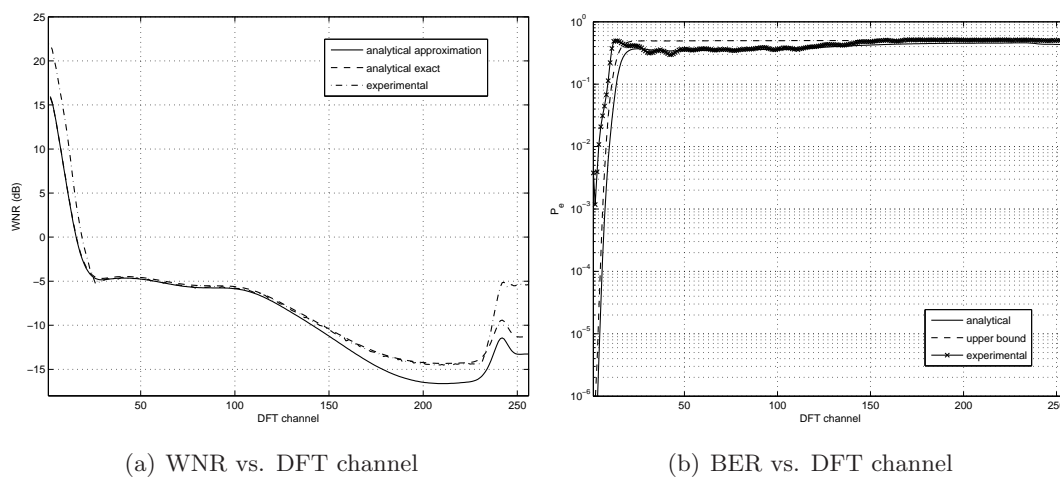


Figure 5.18: Analytical and experimental results for colored host and band-pass filter.

filter has large spectral variation. As it is shown in Fig. 5.18(b), DFT-RDM exhibits low error probabilities for the first frequency channels, while it is approximately equal to 0.5 for all the channels after the first transition band. The dramatic effect of this attack is evident noting that, though the filter amplitude is different from zero for each frequency

5.3 Performance analysis for colored Gaussian hosts

channel, the overall error probability which has been measured after DFT-RDM decoding is $P_e \approx 0.451$.

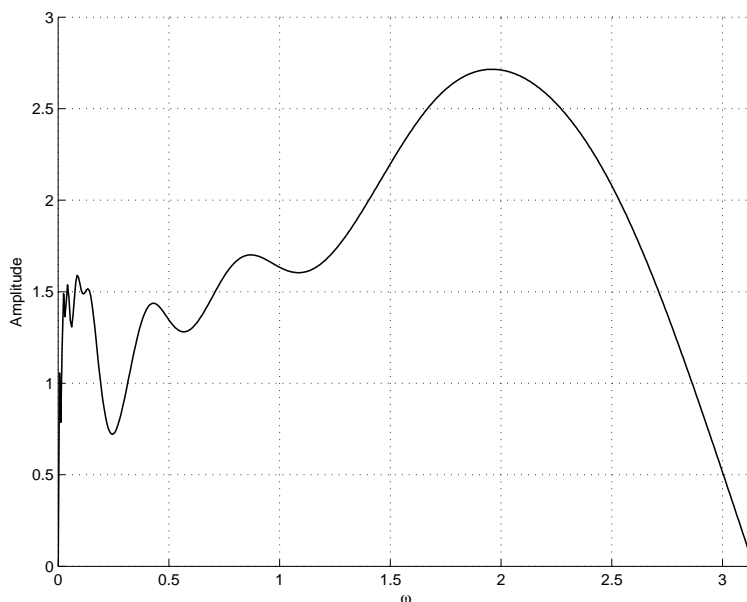


Figure 5.19: Magnitude of the ten-band audio equalizer used in the experiments.

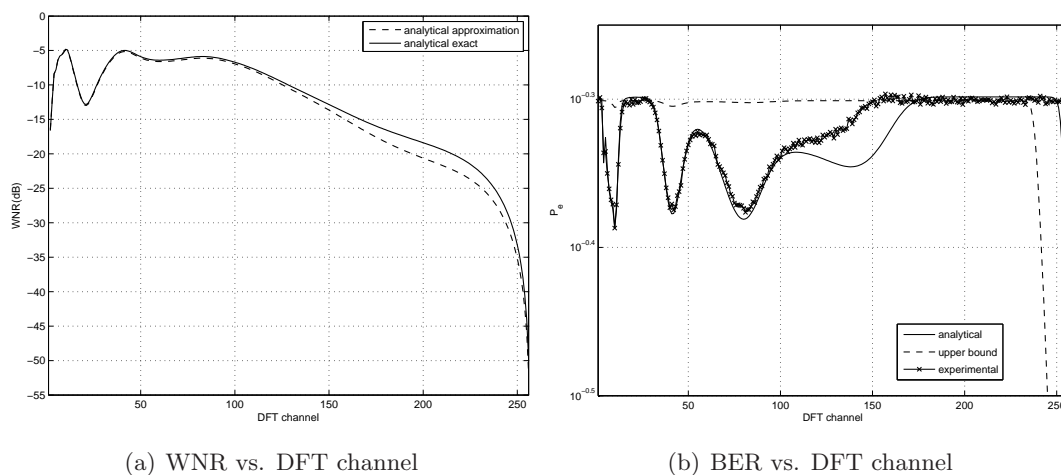


Figure 5.20: Analytical and experimental results for colored host and ten-band equalizer attack.

Finally, a ten-band graphic audio equalizer has been used as an attack filter. In the following experiments we have set the equalizer subband filters so that they produce

the overall frequency response depicted in Fig. 5.19, which is the same that was used in the experiments presented in [105]. Fig. 5.20 illustrates the analytical WNRs and the comparison of the experimental BERs with the analytical ones, respectively. From the per-channel WNR shown in Fig. 5.20(a) it can be inferred that the expected error probability will be very high, especially for the high frequency channels, and this behavior is confirmed by the experimental BERs shown in Fig. 5.20(b). One can also notice that the analytical error probabilities provide a minimal prediction only for the low frequency channels. We conjecture that the observed inaccuracies are due to the correlation among neighboring channels of the colored host, which could be increased even further by the equalizer. Above all, this experiment reveals that DFT-RDM applied to a colored host does not guarantee at all the robustness of the watermark against an equalizer attack, especially as neither windowing nor spreading are here used, since the overall BER is approximately 0.48. It is worth noting that by embedding the watermark with the same system parameters into a white Gaussian host, the overall BER is approximately 0.21.

5.4 Whitened DFT-RDM

From the analysis of DFT-RDM for colored hosts developed in Section 5.3, it can be inferred that any colored host will have unavoidably different watermark signal powers for different DFT channels; consequently, there will be some DFT channels more exposed than others to the per-channel multiplication error, as it has been explained above. Assuming that neither the embedder nor the decoder have any prior knowledge about the attack filter, it is reasonable to perform the RDM embedding in every DFT channel with the same watermark power. Clearly, this choice does not assure the best BER for every attack filter but it is a trade-off to have a good BER even if the attack filter is totally unknown. The optimum would be to shape the per-channel watermark power so that it is larger in those DFT channels which are less modified by the attack filter, but this assumes prior knowledge, so we have decided not to follow this path.

On the other hand, according to [105], if the host signal is white, the per-channel multiplication error is approximately independent of both the host and the watermark signal, so the correlation between neighboring channels, which usually leads to higher per-channel error probabilities, becomes small.

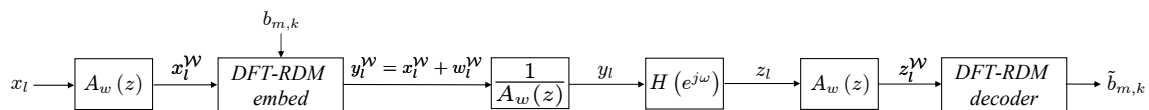


Figure 5.21: Block scheme of the whole embedding/decoding chain for W-DFT-RDM.

These considerations lead to whiten the host signal and use the same embedding power on every DFT channel. The idea is then to perform the DFT-RDM embedding in the host signal $\mathbf{x}^{\mathcal{W}}$ obtained as the output of a whitening filter $A_w(z)$ excited by the colored host \mathbf{x} . Heretofore the superscript \mathcal{W} is used to denote signals which are obtained by whitening. After the embedding, the watermarked signal $\mathbf{y}^{\mathcal{W}}$ is filtered by the inverse of the whitening filter to reshape the signal, as shown in Fig. 5.21, where the whole block scheme for Whitened DFT-RDM (W-DFT-RDM) is depicted. At the decoder side the received host signal \mathbf{z} feeds the whitening filter $A_w(z)$ and from the obtained signal $\mathbf{z}^{\mathcal{W}}$ the DFT-RDM decoder recovers the estimated hidden message.

In this section we will assume that the host is an AR random signal which is generated as described in section 5.3 by the all-pole filter $1/A(z)$. If the whitening filter $A_w(z)$ is equal to $A(z)$, then we have $\mathbf{x}^{\mathcal{W}} = \mathbf{x}_0$, which is a white Gaussian host with power spectral density $\sigma_{x_0}^2$ by construction of the colored signal. After DFT-RDM embedding, the watermarked signal can be expressed as $\mathbf{y}^{\mathcal{W}} = \mathbf{x}^{\mathcal{W}} + \mathbf{w}^{\mathcal{W}}$. Since DFT-RDM embedding is performed on the white signal \mathbf{x}_0 , the resulting watermark signal $\mathbf{w}^{\mathcal{W}}$ can be also assumed to be white and uncorrelated with the host signal from the properties of DFT-RDM. Consequently, the reconstruction filter shapes both the host and watermark signal in the same way, so that their power spectral densities have approximately the same trend, as it is shown in Fig. 5.22.

Moreover, given the whiteness of the watermark signal and the superposition principle, the overall DWR is not changed by the reconstruction filter:

$$\text{DWR} = \frac{\sigma_x^2}{\sigma_w^2} = \frac{\int_{-\pi}^{\pi} \sigma_{x_0}^2 / |A(e^{j\omega})|^2 d\omega}{\int_{-\pi}^{\pi} \sigma_{w^{\mathcal{W}}}^2 / |A(e^{j\omega})|^2 d\omega} = \frac{\sigma_{x_0}^2}{\sigma_{w^{\mathcal{W}}}^2} \quad (5.34)$$

and it is approximately equal to the DWR measured on each DFT-RDM channel, as expected according to (5.8). Thus, even if DFT-RDM is applied to the host signal after whitening, the relation between the overall DWR and Δ is the same as in DFT-RDM, as described in [105]. From this it can be inferred that DFT-RDM with whitening does not

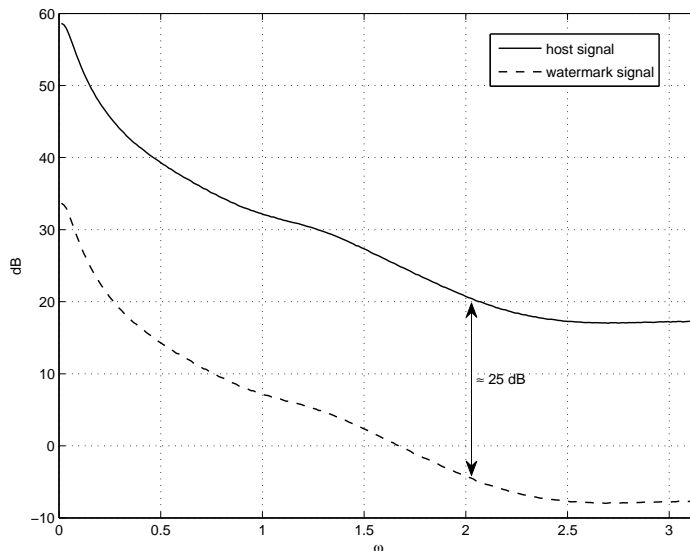


Figure 5.22: Experimental power spectral densities of host and watermark signal after reconstruction filtering for $DWR = 25$ dB.

incur in any penalty in terms of embedding distortion with respect to DFT-RDM, which is a desirable property of the proposed extension.

At the decoder side, after the whitening filter $A_w(z)$, we have $z_l^W = y_l^W * h_l$, hence the white watermarked signal \mathbf{y}^W goes through an equivalent channel where there is only the attack filter. Consequently, even if the host is colored, using the above proposed scheme we expect the same performance as for DFT-RDM applied to a white host for the same attack filter and the same system parameters.

A remarkable property of W-DFT-RDM is that it does not incur in any penalty in terms of embedding distortion and payload with respect to DFT-RDM. Moreover, as DFT-RDM embedding is performed on a white Gaussian signal, it is possible to combine the whitening with the improvements to the basic DFT-RDM that have been described in Section 5.2.2.

In the next section the performance of W-DFT-RDM applied to real audio tracks will be shown. Since the whitening filter $A_w(z)$ is the inverse of an AR filter which resembles the spectral contents of a generic audio signal, we can no longer expect \mathbf{x}^W to be really a white signal. Moreover, since real audio signals are not stationary, the matching between the whitening filter and the psd of the signal will be time-varying. However, \mathbf{x}^W will

usually have a per-channel host power more evenly distributed than the original host.

5.4.1 Experimental results for colored Gaussian host

Some experiments were conducted to verify the effectiveness of the extension of DFT-RDM proposed in Section 5.4; in the following, the host will be assumed to be colored by $1/A_{av}(z)$ with order $Q = 10$, whereas perfect whitening is assumed, i.e., $A_w(z) = A_{av}(z)$. In all the experiments the DWR was set to 25 dB, in the $g(\cdot)$ function the memory L was set to 100 and p was set to 2, and, unless otherwise specified, we assume that the DFT length is $N = 512$ and that neither spreading nor windowing are used.

First of all, we have compared the performance of W-DFT-RDM with that of DFT-RDM applied to both white and colored hosts. The experimental BERs measured for different attack filters are presented in Figs. 5.23, 5.24, 5.25 and 5.26, where it is verified that the BER of DFT-RDM applied to a white host matches always that of W-DFT-RDM applied to a colored host, as it was expected.

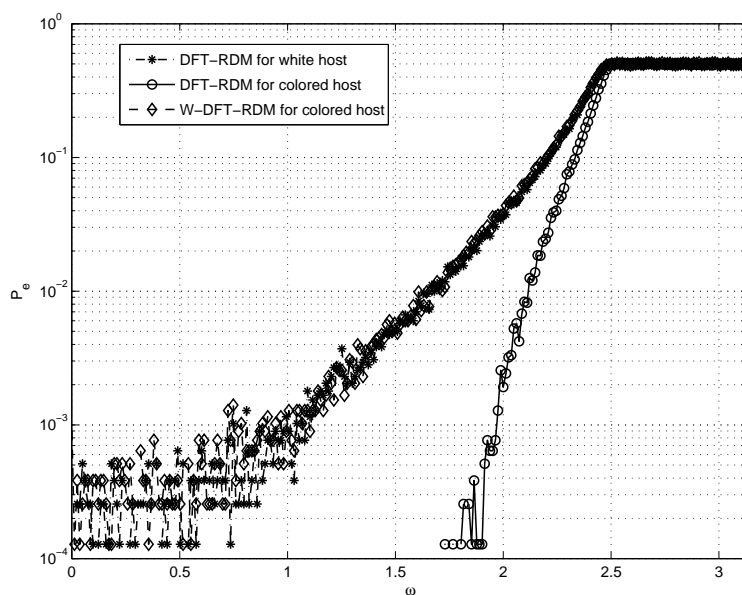


Figure 5.23: BERs versus discrete frequency for low-pass filter with $\omega_c = 0.8\pi$ rad.

In Fig. 5.23 are shown the experimental BERs measured for the ideal low-pass attack filter with cut-off frequency $\omega_c = 0.8\pi$ rad. It can be noticed that for the given attack filter, the overall error probability of DFT-RDM applied directly to the colored host is

$P_e \approx 0.12$, which is less than the overall error probability of DFT-RDM for a white host ($P_e \approx 0.134$). This behavior can be easily explained by the fact that the per-channel watermark signal power is larger at low frequency channels which are not modified at all by the attack filter. This result confirms the conclusion that whitening does not always assure the best BER for every attack filter.

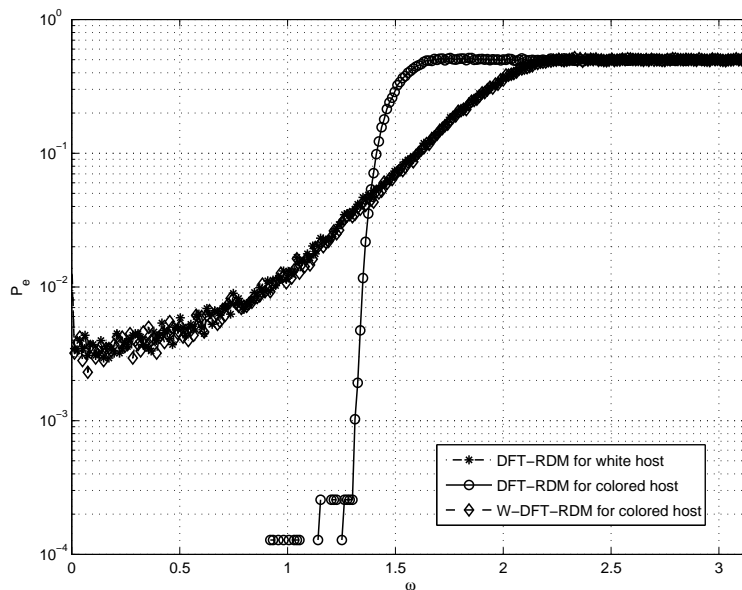


Figure 5.24: BERs versus discrete frequency for low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad.

Then we have tested the watermarking methods with the low-pass filter having passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad, with a smooth transition in the middle. The experimental BERs are shown in Fig. 5.24. In this case, the error probability of W-DFT-RDM is approximately 0.5 only in the stopband, while for DFT-RDM applied to a colored host it is 0.5 in the transition band too. This yields the overall error probability of DFT-RDM ($P_e \approx 0.26$), which is larger than that of W-DFT-RDM ($P_e \approx 0.21$).

In Section 5.3.1 the effect of the considered band-pass filter, whose frequency response is shown in Fig. 5.17, on a DFT-RDM watermarked colored host has been shown, highlighting that the overall error probability is $P_e \approx 0.451$. If the colored host is marked using W-DFT-RDM the decoder is able to correctly retrieve the data embedded in the middle channels, while the error probability increases in correspondence of the transition bands, as it is shown in Fig. 5.25. As a consequence W-DFT-RDM outperforms DFT-RDM for

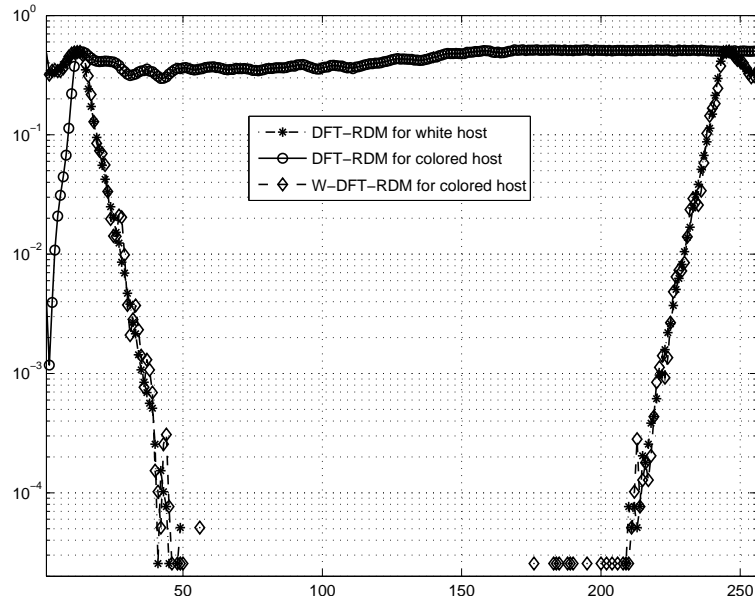


Figure 5.25: BERs versus discrete frequency for the band-pass filter.

colored host signal since the measured overall error probability is $P_e \approx 0.055$, which is approximately ten times less than DFT-RDM.

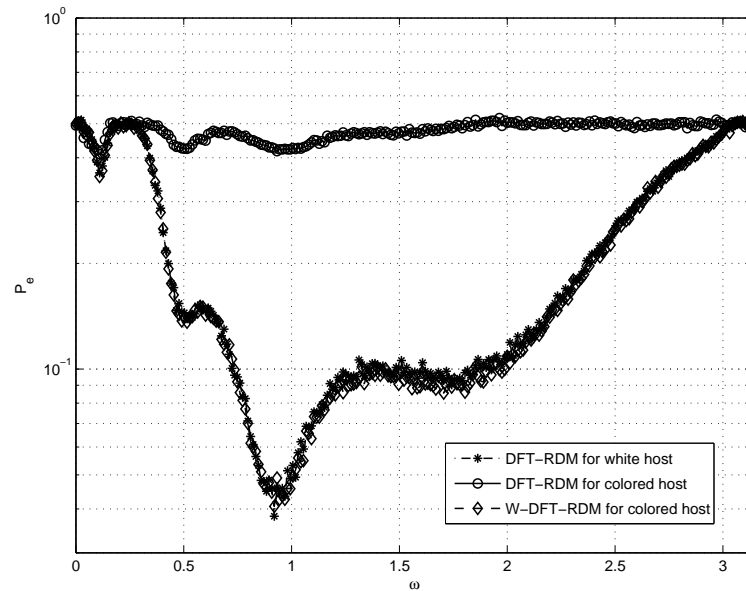


Figure 5.26: BERs versus discrete frequency for the ten-band equalizer attack.

In Fig. 5.26 are shown the BERs for the ten-band graphic audio equalizer. With this attack filter, since the filtering effect is spread over all frequencies, W-DFT-RDM outperforms DFT-RDM for colored hosts (the overall error probabilities are respectively $P_e \approx 0.21$ and $P_e \approx 0.48$).

5.4.2 Experimental results for audio tracks

We have also compared the behavior of W-DFT-RDM and of DFT-RDM using real audio tracks sampled at 44.1 kHz with 16 bits as host signal. These experiments have been conducted using for all the audio tracks a fixed whitening filter, which is again $A_w(z) = A_{av}(z)$. We remark here that perfect whitening does not occur with audio tracks since the whitening filter $A_w(z)$ is the inverse of an AR filter which resembles the spectral contents of a generic audio signal.

The measured DWRs have been obtained fixing the target DWR at 25 dB; we remark here that with nonstationary, non-Gaussian and non-white hosts the analytical derivation of the DWR for DFT-RDM is only an approximation. For sake of completeness, we recall that in the experiments the DFT length is $N = 512$, in the $g(\cdot)$ function the memory L was set to 100 and p was set to 2.

In Tables 5.1, 5.2, 5.3 and 5.4 the overall error probabilities evaluated for a spreading factor $M = 1$ (i.e., no spreading) and a rectangular window are given. Notice that in all the experiments, for the same audio track, the DWRs produced by the two embedding techniques are approximately equal.

Table 5.1 shows the experimental results for the low-pass filter with cut-off frequency $\omega_c = 0.8\pi$ rad. As it was to be expected from the results presented in Section 5.4.1 for Gaussian colored hosts, DFT-RDM has lower bit error probabilities than W-DFT-RDM also for audio signals. Similar results have been obtained attacking the watermarked host with the low-pass filter having passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad. As it is shown in Table 5.2, the overall error probabilities for DFT-RDM are mostly lower than the respective ones for W-DFT-RDM; however, the behavior depends on the particular audio track, as it can be noticed from the results obtained for the tracks "Spff" and "Spfg".

In contrast, for the band-pass filter and the ten-band equalizer attack, W-DFT-RDM yields an improved overall BER for all the audio tracks. From the inspection of table 5.3, it can be noticed that W-DFT-RDM outperforms DFT-RDM for all the tested tracks. Apart from the track "Trpt", for which DFT-RDM and W-DFT-RDM have approximately the

Table 5.1: Overall error probabilities for the low-pass filter with $\omega_c = 0.8\pi$ rad ($M = 1$ and rectangular window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	24.82	0.110	24.72	0.131
Jarre	25.01	0.125	24.98	0.185
REM	24.96	0.102	24.75	0.129
Sopr	24.97	0.116	24.79	0.131
Spff	24.81	0.108	24.97	0.114
Spfg	24.66	0.106	24.53	0.115
Trpt	25.05	0.100	24.79	0.114
Vioo	25.23	0.105	25.23	0.162

Table 5.2: Overall error probabilities for the low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi]$ rad ($M = 1$ and rectangular window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	24.82	0.237	24.74	0.249
Jarre	24.97	0.274	24.96	0.299
REM	24.96	0.219	24.73	0.263
Sopr	24.98	0.235	24.79	0.264
Spff	24.79	0.247	24.94	0.179
Spfg	24.64	0.237	24.52	0.172
Trpt	25.03	0.177	24.76	0.264
Vioo	25.24	0.246	25.22	0.283

same performance, for the remaining tracks the improvement with respect to DFT-RDM goes from a factor of 1.5 to 5.6 in terms of error probability.

In table 5.4 the measured bit-error rates for the ten-band equalizer are presented. Also for this attack filter, W-DFT-RDM gives better overall probabilities than DFT-RDM for all the tested audio tracks. Moreover, it is worth noting that for the tracks "Spff" and "Spfg" the measured BERs are lower than the BER obtained for white Gaussian host signal marked with DFT-RDM against the equalizer attack filter, which is $P_{\approx}0.21$.

Table 5.3: Overall error probabilities for the band-pass filter ($M = 1$ and rectangular window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	24.86	0.421	24.83	0.250
Jarre	25.04	0.414	24.98	0.088
REM	24.96	0.467	24.79	0.314
Sopr	25.00	0.435	24.72	0.292
Spff	24.77	0.352	24.92	0.063
Spfg	24.67	0.333	24.53	0.076
Trpt	24.95	0.465	24.78	0.453
Vioo	25.34	0.459	25.19	0.205

Table 5.4: Overall error probabilities for the ten-band equalizer attack ($M = 1$ and rectangular window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	24.81	0.463	24.70	0.359
Jarre	24.99	0.471	24.98	0.308
REM	24.95	0.481	24.76	0.392
Sopr	24.96	0.457	24.78	0.344
Spff	24.79	0.370	24.93	0.166
Spfg	24.69	0.364	24.55	0.149
Trpt	25.01	0.488	24.74	0.481
Vioo	25.21	0.493	25.19	0.360

The dependence of the performance on the particular track is reasonably due to the matching between the fixed whitening filter and the power spectral density of the track. To verify this hypothesis, assuming that the audio signals are modeled according to an AR model with order $Q = 10$, the coefficients of the all-pole filter $1/A_{av}(z)$ have been computed for the tracks "Spff" and "Trpt", which in the previous experiments exhibit highest and lowest performance for W-DFT-RDM, respectively. In Fig. 5.27 we have compared the magnitude of the frequency response of the filters $A_{av}(e^{j\omega})$ evaluated for

these two tracks and that of the whitening filter, which has been computed to resemble the power spectral density of a generic audio signal. In Fig. 5.27 it can be noticed that the whitening filter exhibits a quite good matching with the filter computed from the track "Spff" in the range $[0, 2.5]$ rad. In contrast, as it was expected, the magnitude of the filter computed from the track "Trpt" is far away from that of the whitening filter $A_w(z)$. We want to remark that modeling the audio track with an AR model relies on the false assumption that the audio signal is stationary, hence the computed all-pole filters only represent an average AR model of the audio track.

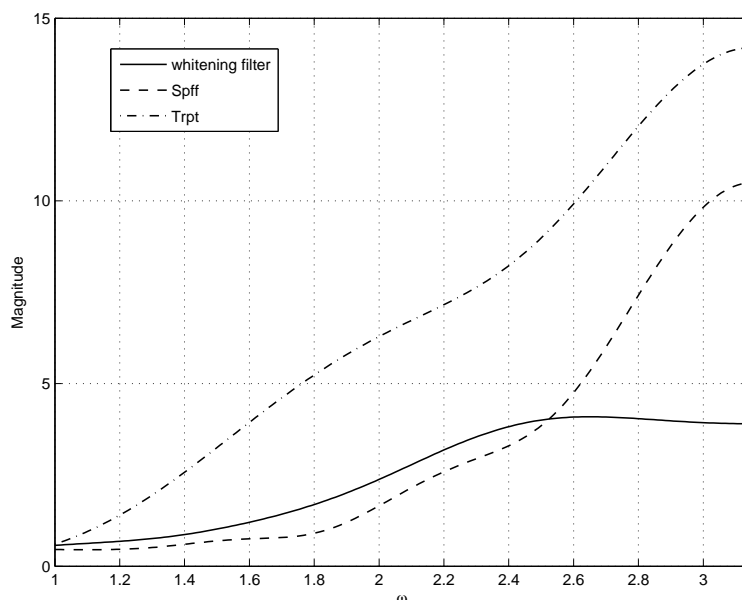


Figure 5.27: Magnitude of the frequency response of the whitening filter and the filters $A_{av}(e^{j\omega})$ evaluated for the tracks "Spff" and "Trpt".

We must remark that the BERs given above for both DFT-RDM-based schemes would be unacceptable in a real watermarking application, thus the experiments have been repeated using a spreading factor $M = 8$ and the optimal window, which has been computed according to [105]. We remind that spreading grants a robustness improvement at the expense of a reduction of the data rate, which becomes 1/16 bits/sample for $M = 8$. From the inspection of the DWRs listed in Tables 5.5, 5.6, 5.7 and 5.8 it can be noticed that in all the experiments, for the same audio track, the DWRs produced by the two embedding techniques are approximately equal.

Table 5.5: Overall error probabilities for the low-pass filter with $\omega_c = 0.8\pi$ rad ($M = 8$ and optimal window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	22.96	0.112	22.93	0.100
Jarre	22.99	0.100	23.11	0.101
REM	26.60	0.100	26.33	0.102
Sopr	23.78	0.110	23.74	0.100
Spff	23.33	0.101	23.45	0.099
Spfg	25.37	0.100	25.25	0.100
Trpt	31.01	0.101	30.73	0.101
Vioo	25.29	0.101	25.27	0.099

Table 5.5 shows the results for the low-pass filter with cut-off frequency $\omega_c = 0.8\pi$ rad. Here, for every audio track, both DFT-RDM-based schemes reach the minimum error probability, which corresponds to the correct detection of all those watermark bits embedded in DFT channels within the passband and is approximately 0.1. In fact in the frequency range $[0.8\pi, \pi[$ rad the filter cuts away all the spectral content and no hidden information can be retrieved. Since the stopband corresponds to $1/5$ of the DFT channels and assuming the correct detection in the remaining channels, the minimum error probability achievable for the ideal low-pass filter attack is approximately 0.1.

From the comparison of the results for the low-pass attacking filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad, that are listed in Table 5.6, we can notice that whitening yields a minimum error probability, that is again approximately 0.1, in almost all the experiments. Moreover, DFT-RDM has always an overall error probability higher than W-DFT-RDM and away from the minimum error probability.

In Table 5.7 it is shown the comparison of the experimental bit-error probabilities evaluated for DFT-RDM and W-DFT-RDM using spreading factor $M = 8$ and optimal windowing for the band-pass filter. As it was expected, the performance of W-DFT-RDM are clearly improved with respect to those listed in Table 5.3. Moreover the DFT-RDM is outperformed for every audio track; in particular it is interesting to notice that for the track "Trpt", even if the fixed whitening filter does not match the psd of the audio track as it is shown in Fig. 5.27, the BER obtained with DFT-RDM is still high while W-DFT-

Table 5.6: Overall error probabilities for the low-pass filter with passband $[0, 0.4\pi]$ rad and stopband $[0.8\pi, \pi)$ rad ($M = 8$ and optimal window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	22.95	0.235	22.94	0.113
Jarre	22.98	0.139	23.10	0.101
REM	26.60	0.176	26.30	0.102
Sopr	23.82	0.241	23.76	0.101
Spff	23.36	0.148	23.45	0.100
Spfg	25.36	0.141	25.26	0.100
Trpt	31.05	0.247	30.74	0.158
Vioo	25.25	0.213	25.30	0.109

Table 5.7: Overall error probabilities for the band-pass filter ($M = 8$ and optimal window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	22.99	0.187	22.88	0.009
Jarre	23.07	0.011	23.11	0.004
REM	26.62	0.020	26.35	0.003
Sopr	23.74	0.226	23.66	0.003
Spff	23.42	0.006	23.48	0.0003
Spfg	25.33	0.028	25.29	0.011
Trpt	31.04	0.324	30.78	0.007
Vioo	25.28	0.148	25.21	0.002

RDM gives $P_e \approx 0.007$. It is worth noting that, since the band-pass filter does not erase the spectral content at any frequency, the minimum achievable error probability is 0.

The overall error probabilities presented in Table 5.8 confirm the better behavior of W-DFT-RDM for the equalizer attack. In fact, for every audio track the BER of W-DFT-RDM is always lower, with an improvement with respect to DFT-RDM that goes from a factor of 1.5 to 7 in terms of error probability, depending on the audio track.

Even though linear filtering does not encompass MPEG Layer-3 (MP3) compression, this can be very roughly seen as a low-pass filtering with cut-off frequency equal to the

Table 5.8: Overall error probabilities for the ten-band equalizer attack ($M = 8$ and optimal window)

Track	DFT-RDM		W-DFT-RDM	
	DWR (dB)	BER	DWR (dB)	BER
Bass	22.96	0.229	22.90	0.0325
Jarre	23.00	0.0380	23.14	0.0180
REM	26.61	0.0688	26.33	0.0130
Sopr	23.79	0.248	23.71	0.0383
Spff	23.34	0.0580	23.44	0.0115
Spfg	25.36	0.0542	25.27	0.0349
Trpt	31.02	0.3514	30.75	0.0512
Vioo	25.33	0.186	25.29	0.0282

Table 5.9: Overall error probabilities for MP3 compression attacks ($M = 8$ and optimal window)

Track	DFT-RDM				W-DFT-RDM			
	DWR (dB)	80 kbps	160 kbps	320 kbps	DWR (dB)	80 kbps	160 kbps	320 kbps
Bass	22.96	0.389	0.346	0.322	22.90	0.339	0.232	0.145
Jarre	23.00	0.409	0.229	0.155	23.14	0.402	0.213	0.140
REM	26.61	0.405	0.258	0.223	26.33	0.360	0.187	0.143
Sopr	23.79	0.399	0.354	0.337	23.71	0.345	0.220	0.146
Spff	23.34	0.292	0.185	0.148	23.44	0.280	0.175	0.138
Spfg	25.36	0.252	0.172	0.144	25.27	0.246	0.167	0.138
Trpt	31.02	0.402	0.371	0.366	30.75	0.389	0.346	0.328
Vioo	25.33	0.389	0.319	0.290	25.29	0.363	0.257	0.167

sampling frequency of the audio track after MP3 compression. Hence, we have carried on several experiments to verify the robustness of DFT-RDM-based techniques to MP3 compression. The real audio tracks, whose sampling frequency is 44.1 kHz, have been marked, compressed using LAME 3.97 [134] to perform MP3 encoding, and, finally, the watermark has been retrieved after decompression.

In Table 5.9 are listed the BERs measured for both the DFT-RDM-based techniques using a spreading factor $M = 8$ and the optimal window. These results have been obtained

for constant bit-rate MP3 encoding of the watermarked audio tracks, but approximately the same error probabilities have been measured for average bit-rate MP3 encoding. It is worth noting that in these experiments the minimum error probability is approximately 0.137, that corresponds to the correct detection of all the watermark samples embedded up to 32 kHz, which is the sampling frequency of the audio tracks compressed by LAME for the considered bit-rates. We want to remark here that the minimum overall error probability $P_e \approx 0.137$ is due to the choice of inserting the hidden information in all the channels up to the sampling frequency of 44.1 kHz, even if it is known that, for audio signals, the high frequencies are not perceptually relevant and consequently they can be likely cut away in the life cycle of the marked content. As a consequence, the data rates listed in Table 5.9 can be quite easily increased designing properly the data hiding system and taking into account perceptual considerations.

From the inspection of the results in Table 5.9, it can be noticed that the error probabilities of W-DFT-RDM are always lower than those of DFT-RDM for the same bit-rate of the MP3 encoded files. Even if the measured error probabilities are considerably dependent on the particular audio track, W-DFT-RDM approaches the minimum error probability for almost all audio tracks and an encoding bit-rate equal to 320 kbps. On the other hand, the BERs measured for DFT-RDM can be far away from the minimum error probability even if the audio tracks are encoded at the maximum allowed bit-rate.

5.5 Concluding remarks

A thorough analysis of the behavior of DFT-RDM for colored Gaussian hosts has been performed. The developed analysis, which is based on a frequency-domain approach, provided an explanation to the performance loss with respect to white Gaussian hosts. This is essentially due to the combination of two facts: 1) on each discrete frequency channel the power of the RDM watermark signal is proportional to the per-channel host signal power, so that different channels are differently protected against additive noise, and 2) the influence of the non-flat psd of the host on the self-noise that in turn is due to a block-DFT operation. Moreover, the per-channel watermark-to-noise ratio has been introduced as a simple measure to evaluate the reliability of each RDM-like channel. The comparison of both the analytically predicted per-channel WNR and the per-channel bit-

error probability with the results of experiments for different attack filters confirmed the validity of the developed generalized analysis.

We have also provided an extension of DFT-RDM for colored hosts without assuming any additional knowledge on the attack filter and without incurring in any penalty in terms of embedding distortion and payload with respect to DFT-RDM. The proposed extension consists in using a fixed whitening filter that captures the average properties of audio signals. The analysis has been validated by experimental results which confirm the performance improvement afforded by the proposed solution. Some interesting results have been also obtained applying W-DFT-RDM to real audio signals. In fact, a BER decrease has been evaluated with respect to basic DFT-RDM and these results encourage to continue on this research line towards the development of data hiding applications based on DFT-RDM.

6

Conclusions

In this thesis, we have focused on the analysis of the robustness requirements for data hiding systems, paying our attention to side-informed methods and desynchronization channels. Since robustness accounts for the capability of the hidden data to survive host signal manipulations, guaranteeing robustness against a wide range of attacks is a fundamental requirement to develop real data hiding applications. Even if a great effort has been spent by the research community in this topic, there are still open issues.

We focused on the desynchronization attacks, that aim at removing the watermark by changing the reference of the watermarked signal, so that the decoder is no more able to retrieve the hidden information. We have shown how the attacks within this class have a different impact on spread spectrum-based systems and on quantization-based systems, in particular the effect of even mild desynchronization attacks on quantization-based decoders can cause a dramatic increase of the error probability

In Chapter 2 a spread spectrum-based watermarking technique robust to geometric distortions for images is presented. The proposed watermarking technique is proven to be robust against rotations and other distortions while preserving the image fidelity, as it is demonstrated by the experimental results. Also, it has been shown that, in order to make the basic correlation decoder able to correctly retrieve the hidden message, the watermark itself has to be nearly invariant against desynchronization attacks. As a consequence, achieving robustness to desynchronization attacks for spread spectrum-based methods is quite simpler than for quantization-based.

Hence, the development of quantization-based algorithms intrinsically invariant to desynchronization attacks, that are modeled by non-additive channels, is the main topic

of this thesis. We departed from the fixed gain attack, that can be considered somewhat solved, and from the rational dither modulation, which is theoretically proven to be invariant to gain scaling and it retains the fundamental advantage of DM, which can be identified in satisfactory trade-off between robustness to the AWGN channel, capacity capability and host signal fidelity. In fact RDM has been usefully exploited as basic tool to account for more complex non-additive channel, taking advantage of the theoretical framework provided in [104] for RDM to identify the achievable performance of the proposed extensions. We focused on the power-law attack, which models nonlinear volumetric distortion such as gamma correction, and on the linear time invariant filtering, which is one of the most common processing for every media content.

The main contributions of this thesis can be thus summarized

- A new class of quantization-based data hiding schemes invariant to power-law attack has been proposed. By hyperbolic angle mapping of the host samples, it is obtained a domain invariant to both gain and exponentiation, where the information embedding can be performed according to any rule within the class of gain invariant QIM-based methods. Therefore any gain invariant QIM-based algorithm can be made invariant to the power-law attack. Then, the performance of these methods have been evaluated and the expected behavior has been verified.
- Within the proposed class of quantization-based methods, Hyperbolic RDM, which essentially consists of performing RDM embedding in the hyperbolic angle domain, has been deeply analyzed. This study has been motivated by the fact that, among the proposed class of power-law attack invariant methods, Hyperbolic RDM has the best data rate against additive noise. The average embedding distortion and the peak embedding distortion have been theoretically derived and the developed analysis has been validated by experiments. Also, exploiting the analysis provided for RDM, it has been provided the analytical expression for decoding error probability to clearly identify the achievable data-rate against additive noise.
- Departing from the analysis in [105] for discrete Fourier transform - rational dither modulation, which was proposed to provide high data rate against LTI filtering, a generalized analysis for non-white hosts has been developed to predict the per-channel decoding error probability. This study was motivated by the loss of performance of DFT-RDM for non-white hosts with respect to white hosts. Hence, using

a frequency domain approach to separately consider each RDM-like channel on each discrete frequency channel, a reasonable explanation of the loss of performance has been identified in the combination of two facts: 1) the power of an RDM watermark signal is proportional to the host signal power, and 2) the influence of the non-flat power spectral density of the host on the self-noise that in turn is due to a block-DFT operation.

- The per-channel watermark-to-noise ratio has been introduced as a simple and intuitive measure to evaluate the reliability of each RDM-like channel, allowing to quickly infer the frequency channel where the hidden information can be retrieved. Moreover the upper bound of the per-channel decoding error probability has been computed as a function of the per channel WNR.
- Whitened DFT-RDM has been proposed to improve the performance of DFT-RDM for non-white Gaussian hosts without assuming any prior knowledge on the attack filter. Performing DFT-RDM on the whitened host signal, the same performance as for DFT-RDM applied to a white host have been evaluated for the same attack filter and the same system parameters. Moreover W-DFT-RDM is proven to not incur in any penalty in terms of embedding distortion and payload with respect to DFT-RDM.
- Whitened DFT-RDM has been tested for real audio signals and satisfactory data rates have been measured against both different attack filters and MP3 compression.

6.1 Future research lines

The work carried out in this thesis leaves a number of open questions which are worth addressing in the future. In the first place, we can enumerate a number of open issues related to Hyperbolic RDM and the whole class of quantization-based methods invariant to power-law attack:

- Hyperbolic RDM has shown a loss of performance against the noise addition with respect to the classical DM-based methods, hence several strategies can be investigated to improve the data rate for Hyperbolic RDM:
 1. Using a detector based on a metric different from the Euclidean distance.

2. The design of a function $h(\cdot) \in \mathcal{H}$ to reject the noise interference at the decoder.
3. Using a dirty-paper trellis code in the embedding function. Dirty-paper trellis code could be adopted to directly embed the data or in conjunction with RDM, as it was proposed in [128].

Also channel coding can be used to improve the BER at price of a payload reduction.

- The experiments carried on Hyperbolic RDM for real images exhibit encouraging results, but a considerable amount of work is still necessary to use this technique in practical applications. Firstly, perceptual considerations has to be incorporated in the method to mask the distortion introduced by the embedding. Moreover, the influence on the performance of varying channel parameters has to be analyzed and practical implementations of Hyperbolic RDM with multimedia signals different from images need yet to be developed.

Also there are some open problems related with Whitened DFT-RDM:

- From the previous results it can be noticed that W-DFT-RDM for audio tracks is not able to fill the performance gap with respect to DFT-RDM for white hosts since a fixed (and non-perfectly matched) average whitening filter is used at both the embedder and the decoder. This is mainly due to the fact that audio signals are nonstationary, hence a further improvement can be the use of a host adaptive, and possibly time-varying, whitening filter at the encoder. Then the host adaptive whitening filter can be transmitted to the decoder or it can reconstructed from the received signal, at least with some approximation.
- Even though encouraging BER results have been obtained for W-DFT-RDM and MP3 compression, an accurate analysis of DFT-RDM-based techniques against compression is needed in order to assess the real bounds.
- To develop a data hiding application based on DFT-RDM for audio signals, the watermark signal should be perceptually masked taking into account the properties of human auditory system.
- The analysis of the DFT-RDM-based algorithms can be extended incorporating an additive noise source in the LTI filtering channel.

6.1 Future research lines

- The DFT-RDM-based techniques can be extended to the case of multi-dimensional signals, evaluating the performance for media content such as images or video, that are nonstationary, non-Gaussian and colored too.

References

- [1] Emil F. Hembrooke, “Identification of sound and like signals,” 1961, United States Patent, 3004104.
- [2] N. Komatsu and H. Tominaga, “Authentication system using concealed image in telematics,” pp. 45–50, 1988.
- [3] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography, 2nd Ed. (The Morgan Kaufmann Series in Multimedia Information and Systems)*, 2nd ed. Morgan Kaufmann, 11 2007.
- [4] A. Bell, “The dynamic digital disk,” *Spectrum, IEEE*, vol. 36, no. 10, pp. 28–35, Oct 1999.
- [5] G. Depovere, T. Kalker, J. Haitsma, M. Maes, L. de Strycker, P. Termont, J. Vandewege, A. Langell, C. Alm, P. Norman, G. O’Reilly, B. Howes, H. Vaanholt, R. Hintzen, P. Donnelly, and A. Hudson, “The VIVA project: digital watermarking for broadcast monitoring,” in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 2, 1999, pp. 202–205 vol.2.
- [6] F. Hartung and M. Kutter, “Multimedia watermarking techniques,” *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1079–1107, Jul 1999.
- [7] “Secure Digital Music Initiative (SDMI),” [http://en.wikipedia.org/wiki/Secure Digital Music Initiative](http://en.wikipedia.org/wiki/Secure_Digital_Music_Initiative).
- [8] “SDMI challenge FAQ,” <http://www.cs.princeton.edu/sip/sdmi/faq.html>.
- [9] C. Herley, “Why watermarking is nonsense,” *Signal Processing Magazine, IEEE*, vol. 19, no. 5, pp. 10–11, Sep 2002.

-
- [10] M. Barni, “What is the future for watermarking? (part I),” *Signal Processing Magazine, IEEE*, vol. 20, no. 5, pp. 55–60, Sep 2003.
- [11] —, “What is the future for watermarking? (part II),” *Signal Processing Magazine, IEEE*, vol. 20, no. 6, pp. 53–59, Nov. 2003.
- [12] P. Moulin, “Comments on ”why watermarking is nonsense”,” *Signal Processing Magazine, IEEE*, vol. 20, no. 6, pp. 57–59, Nov. 2003.
- [13] L. Perez-Freire, P. Comesana, J. R. Troncoso-Pastoriza, and F. Perez-Gonzalez, “Watermarking security: a survey,” in *LNCS Transactions on Data Hiding and Multimedia Security*, 2006.
- [14] I. Cox, M. Miller, and A. McKellips, “Watermarking as communications with side information,” *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul 1999.
- [15] <http://omen.cs.uni-magdeburg.de/ecrypt/>.
- [16] “IEEE signal processing magazine,” Mar 2004.
- [17] <http://www.digimarc.com/>.
- [18] <http://www.teletrax.tv>.
- [19] <http://www.business-sites.philips.com/contentidentification/home>.
- [20] <http://www.thomson.net/globalenglish/solutions/content-tracking-security>.
- [21] <http://www.cinea.com>.
- [22] <http://www.verimatrix.com>.
- [23] <http://www.verance.com>.
- [24] <http://www.geovision.com.tw>.
- [25] <http://www.mediasec.com/>.
- [26] <http://www.tredess.com/en/index.html>.
- [27] <http://www.datamark.com.sg/>.

-
- [28] <http://www.civolution.com/>.
- [29] <http://www.markany.com/>.
- [30] <http://www.widevine.com/>.
- [31] <http://www.msicopycontrol.com/>.
- [32] <http://www.aquamobile.es/>.
- [33] <http://www.digitalwatermarkingalliance.org/>.
- [34] M. Kirstein, “Beyond traditional DRM: Moving to digital watermarking and fingerprinting in media monetization,” January 2006, available at <http://www.multimediantelligence.com>.
- [35] M. Swanson, M. Kobayashi, and A. Tewfik, “Multimedia data-embedding and watermarking technologies,” *Proceedings of the IEEE*, vol. 86, no. 6, pp. 1064–1087, Jun 1998.
- [36] M. Barni and F. Bartolini, *Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications*. Imprint unknown, 1 2004.
- [37] S. Voloshynovskiy, S. Pereira, T. Pun, J. Eggers, and J. Su, “Attacks on digital watermarks: classification, estimation based attacks, and benchmarks,” *Communications Magazine, IEEE*, vol. 39, no. 8, pp. 118–126, Aug 2001.
- [38] J. Bloom, I. Cox, T. Kalker, J.-P. Linnartz, M. Miller, and C. Traw, “Copy protection for DVD video,” *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1267–1276, Jul 1999.
- [39] P. Moulin and R. Koetter, “Data-hiding codes,” *Proceedings of the IEEE*, vol. 93, no. 12, pp. 2083–2126, Dec. 2005.
- [40] P. Campisi, D. Kundur, D. Hatzinakos, and A. Neri, “Compressive data hiding: An unconventional approach for improved color image coding,” in *EURASIP Journal on Applied Signal Processing Special Issue on Emerging Applications of Data Hiding*, vol. 2002, no. 2, 2002, pp. 152–163.

-
- [41] J. Fridrich and M. Goljan, “Images with self-correcting capabilities,” in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 3, 1999, pp. 792–796 vol.3.
- [42] F. Bartolini, A. Manetti, A. Piva, and M. Barni, “A data hiding approach for correcting errors in H.263 video transmitted over a noisy channel,” in *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on*, 2001, pp. 65–70.
- [43] S. Decker, “Engineering considerations in commercial watermarking,” *Communications Magazine, IEEE*, vol. 39, no. 8, pp. 128–133, Aug 2001.
- [44] J. Dittmann, M. Steinebach, P. Wohlmacher, and R. Ackermann, “Digital watermarks enabling e-commerce strategies: conditional and user specific access to services and resources,” *EURASIP J. Appl. Signal Process.*, vol. 2002, no. 1, pp. 174–184, 2002.
- [45] T. Kalker, G. Depovere, J. Haitzma, and M. J. Maes, “Video watermarking system for broadcast monitoring,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, P. W. Wong and E. J. Delp, Eds., vol. 3657, Apr. 1999, pp. 103–112.
- [46] P. Campisi, M. Carli, G. Giunta, and A. Neri, “Blind quality assessment system for multimedia communications using tracing watermarking,” *Signal Processing, IEEE Transactions on*, vol. 51, no. 4, pp. 996–1002, Apr 2003.
- [47] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, “Fundamentals of data hiding security and their application to spread-spectrum analysis,” in *Information Hiding*, ser. Lecture Notes in Computer Science, M. Barni, J. Herrera-Joancomartí, S. Katzenbeisser, and F. Pérez-González, Eds., vol. 3727. Springer, 2005, pp. 146–160.
- [48] F. Cayre, C. Fontaine, and T. Furon, “Watermarking security: theory and practice,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 10, pp. 3976–3987, Oct. 2005.
- [49] T. Kalker, “Considerations on watermarking security,” in *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on*, 2001, pp. 201–206.

-
- [50] M. Kutter, S. Voloshynovskiy, and A. Herrigel, “The watermark copy attack,” 2000, pp. 371–380.
- [51] G. Doerr and J.-L. Dugelay, “A guide tour of video watermarking,” *Signal Processing: Image Communication*, vol. 18, pp. 263–282(20), April 2003.
- [52] M. Steinebach, F. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, S. Seibel, N. Fates, and L. Ferri, “StirMark benchmark: audio watermarking attacks,” in *Information Technology: Coding and Computing, 2001. Proceedings. International Conference on*, Apr 2001, pp. 49–54.
- [53] M. Scagliola and P. Guccione, “Rotation invariant feature extraction for watermarking,” in *SIGMAP*, P. A. A. Assunção and S. M. M. de Faria, Eds. INSTICC Press, 2008, pp. 229–235.
- [54] —, “Geometric distortion resilient watermarking based on a single robust feature for still images,” in *Communications in Computer and Information Science*, vol. 48, November 2009, pp. 345–357.
- [55] —, “Providing invariance to nonlinear volumetric scaling for quantization based watermarking,” in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 1489–1492.
- [56] P. Guccione and M. Scagliola, “Hyperbolic RDM for nonlinear volumetric distortions,” *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 1, pp. 25–35, March 2009.
- [57] M. Scagliola, F. Perez-Gonzalez, and P. Guccione, “An extended analysis of discrete fourier transform - rational dither modulation for non-white hosts,” *to be presented at Information Forensics and Security. WIFS 2009. IEEE Workshop on*.
- [58] —, “High-rate data-hiding robust to linear filtering for colored hosts,” *to appear on EURASIP Journal on Information Security*.
- [59] V. S. Tirkel, R. G. V. Schyndel, A. Z. Tirkel, and C. F. Osborne, “Towards a robust digital watermark,” in *Nanyang Technological University Singapore*, 1995, pp. 504–508.

-
- [60] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” *IBM Syst. J.*, vol. 35, no. 3-4, pp. 313–336, 1996.
- [61] I. Cox, J. Kilian, F. Leighton, and T. Shamoan, “Secure spread spectrum watermarking for multimedia,” *Image Processing, IEEE Transactions on*, vol. 6, no. 12, pp. 1673–1687, Dec 1997.
- [62] J. R. Barry, D. G. Messerschmitt, and E. A. Lee, *Digital Communication: Third Edition*, 3rd ed. Springer, 9 2003.
- [63] J. Hernandez, M. Amado, and F. Pérez-González, “DCT-domain watermarking techniques for still images: detector performance analysis and a new structure,” *Image Processing, IEEE Transactions on*, vol. 9, no. 1, pp. 55–68, Jan 2000.
- [64] C. E. Shannon, “A mathematical theory of communication,” *Bell system technical journal*, vol. 27, 1948.
- [65] C. Shannon, “Communication in the presence of noise,” *Proceedings of the IEEE*, vol. 72, no. 9, pp. 1192–1201, Sept. 1984.
- [66] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*, 2nd ed. Wiley-Interscience, 7 2006.
- [67] B. Chen and G. Wornell, “Quantization index modulation: a class of provably good methods for digital watermarking and information embedding,” *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1423–1443, May 2001.
- [68] P. Moulin and J. O’Sullivan, “Information-theoretic analysis of information hiding,” in *Information Theory, 2000. Proceedings. IEEE International Symposium on*, 2000, pp. 19–.
- [69] J. Eggers, R. Bauml, and B. Girod, “Digital watermarking facing attacks by amplitude scaling and additive white noise,” in *in 4th Int. ITG Conf. on Source and Channel Coding*, 2002, pp. 28–30.
- [70] P. Moulin, “The role of information theory in watermarking and its application to image watermarking,” *Signal Process.*, vol. 81, no. 6, pp. 1121–1139, 2001.

-
- [71] A. Cohen and A. Lapidoth, “The gaussian watermarking game,” *Information Theory, IEEE Transactions on*, vol. 48, no. 6, pp. 1639–1667, Jun 2002.
- [72] M. Costa, “Writing on dirty paper (corresp.),” *Information Theory, IEEE Transactions on*, vol. 29, no. 3, pp. 439–441, May 1983.
- [73] J. Eggers, R. Bauml, R. Tzschoppe, and B. Girod, “Scalar costas scheme for information embedding,” *Signal Processing, IEEE Transactions on*, vol. 51, no. 4, pp. 1003–1019, Apr 2003.
- [74] M. Miller, G. Doerr, and I. Cox, “Applying informed coding and embedding to design a robust high-capacity watermark,” *Image Processing, IEEE Transactions on*, vol. 13, no. 6, pp. 792–807, June 2004.
- [75] A. Abrardo and M. Barni, “Orthogonal dirty paper coding for informed data hiding,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, E. J. Delp III & P. W. Wong, Ed., vol. 5306, Jun. 2004, pp. 274–285.
- [76] J. K. Su, B. Girod, and J. J. Eggers, “Analysis of digital watermarks subjected to optimum linear filtering and additive noise,” in *Signal Processing, Special Issue on Information-Theoretic Issues in Digital Watermarking*, 2000, pp. 1141–1175.
- [77] R. Bäuml, Joachim, and J. Huber, “A channel model for watermarks subject to desynchronization attacks,” vol. 4675, 2002, pp. 281–292.
- [78] F. Bartolini, M. Barni, and A. Piva, “Performance analysis of spread transform dither modulation (st-dm) watermarking in presence of nonadditive attacks,” *IEEE Trans. Signal Proc.*, vol. 52, no. 10, pp. 2965–2974, Oct. 2004.
- [79] F. Petitcolas and R. Anderson, “Evaluation of copyright marking systems,” in *Multimedia Computing and Systems, 1999. IEEE International Conference on*, vol. 1, Jul 1999, pp. 574–579 vol.1.
- [80] M. Barni, A. D’Angelo, and N. Merhav, “Expanding the class of watermark desynchronization attacks,” in *MM&Sec ’07: Proceedings of the 9th workshop on Multimedia & security*. New York, NY, USA: ACM, 2007, pp. 195–204.

-
- [81] V. Licks and R. Jordan, "Geometric attacks on image watermarking systems," *Multimedia, IEEE*, vol. 12, no. 3, pp. 68–78, July-Sept. 2005.
- [82] C.-Y. Lin, M. Wu, J. Bloom, I. Cox, M. Miller, and Y. Lui, "Rotation, scale, and translation resilient watermarking for images," *Image Processing, IEEE Transactions on*, vol. 10, no. 5, pp. 767–782, May 2001.
- [83] M. Alghoniemy and A. Tewfik, "Geometric invariance in image watermarking," *Image Processing, IEEE Transactions on*, vol. 13, no. 2, pp. 145–153, Feb. 2004.
- [84] M. Barni, "Effectiveness of exhaustive search and template matching against watermark desynchronization," *Signal Processing Letters, IEEE*, vol. 12, no. 2, pp. 158–161, Feb. 2005.
- [85] D. Zheng, Y. Liu, and J. Zhao, "A survey of RST invariant image watermarking algorithms," *Electrical and Computer Engineering, 2006. CCECE '06. Canadian Conference on*, pp. 2086–2089, May 2006.
- [86] J.-L. Dugelay, S. Roche, C. Rey, and G. Doerr, "Still-image watermarking robust to local geometric distortions," *Image Processing, IEEE Transactions on*, vol. 15, no. 9, pp. 2831–2842, Sept. 2006.
- [87] P. Agarwal and B. Prabhakaran, "Robust blind watermarking of point-sampled geometry," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 1, pp. 36–48, March 2009.
- [88] P. Moulin, "Universal decoding of watermarks under geometric attacks," in *Information Theory, 2006 IEEE International Symposium on*, July 2006, pp. 2353–2357.
- [89] R. Lagendijk and I. Shterev, "Estimation of attacker's scale and noise variance for QIM-DC watermark embedding," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 1, Oct. 2004, pp. 55–58 Vol. 1.
- [90] I. Shterev and R. Lagendijk, "Amplitude scale estimation for quantization-based watermarking," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4146–4155, Nov. 2006.

-
- [91] P. Moulin, A. Briassouli, and H. Malvar, “Detection-theoretic analysis of desynchronization attacks in watermarking,” in *Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on*, vol. 1, 2002, pp. 77–84 vol.1.
- [92] F. Balado, K. Whelan, G. Silvestre, and N. Hurley, “Joint iterative decoding and estimation for side-informed data hiding,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 10, pp. 4006–4019, Oct. 2005.
- [93] K. Whelan, F. Balado, G. Silvestre, and N. Hurley, “PLL-based synchronization of dither-modulation data hiding,” in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 2, May 2006, pp. II–II.
- [94] J. Wang, I. D. Shterev, and R. L. Lagendijk, “Scale estimation in two-band filter attacks on QIM watermarks,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, E. J. Delp, III and P. W. Wong, Eds., vol. 6072, Feb. 2006, pp. 118–127.
- [95] S. Sadasivam and P. Moulin, “On estimation accuracy of desynchronization attack channel parameters,” *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 3, pp. 284–292, Sept. 2009.
- [96] S. Pereira and T. Pun, “Fast robust template matching for affine resistant image watermarks,” *IEEE Trans. on Image Processing*, vol. 9, pp. 1123–1129, 1999.
- [97] M. Álvarez Rodríguez and F. Pérez-González, “Analysis of pilot-based synchronization algorithms for watermarking of still images,” *Signal Processing: Image Communication*, vol. 17, no. 8, pp. 611 – 633, 2002.
- [98] P. Moulin, “Embedded-signal design for channel parameter estimation part I: linear embedding,” in *Statistical Signal Processing, 2003 IEEE Workshop on*, Sept.-1 Oct. 2003, pp. 38–41.
- [99] —, “Embedded-signal design for channel parameter estimation part II: quantization embedding,” in *Statistical Signal Processing, 2003 IEEE Workshop on*, Sept.-1 Oct. 2003, pp. 42–45.

-
- [100] P. Moulin and A. Ivanovic, "The Fisher information game for optimal design of synchronization patterns in blind watermarking," in *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 2, Oct 2001, pp. 550–553 vol.2.
- [101] M. Feder and A. Lapidoth, "Universal decoding for channels with memory," *Information Theory, IEEE Transactions on*, vol. 44, no. 5, pp. 1726–1745, Sep 1998.
- [102] M. Kutter, "Watermarking resisting to translation, rotation, and scaling," 1998, pp. 423–431.
- [103] J. J. ÓRuanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Process.*, vol. 66, no. 3, pp. 303–317, 1998.
- [104] F. Pérez-González, C. Mosquera, M. Barni, and A. Abrardo, "Rational dither modulation: a high-rate data-hiding method invariant to gain attacks," *IEEE Trans. Signal Proc.*, vol. 53, no. 10, pp. 3960–3975, Oct. 2005.
- [105] F. Pérez-González and C. Mosquera, "Quantization-based data hiding robust to linear-time-invariant filtering," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 2, pp. 137–152, June 2008.
- [106] F. Ourique, V. Licks, R. Jordan, and F. Pérez-González, "Angle QIM: a novel watermark embedding scheme robust against amplitude scaling distortions," *Proc. ICASSP*, vol. 2, pp. 797–800, Mar. 2005.
- [107] P. Comesaña and F. Pérez-González, "On a watermarking scheme in the logarithmic domain and its perceptual advantages," *Proc. of ICIP*, vol. 2, pp. 145–148, Sept. 2007.
- [108] G. Beylkin, "Discrete radon transform," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, no. 2, pp. 162–172, Feb 1987.
- [109] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Rotation-invariant multiresolution texture analysis using Radon and wavelet transforms," *Image Processing, IEEE Transactions on*, vol. 14, no. 6, pp. 783–795, June 2005.

REFERENCES

- [110] J. Lee, R. Haralick, and S. L.G., “Morphological edge detector,” *Robotics and Automation, IEEE Transactions on*, vol. 3, no. 2, pp. 142–153, Apr. 1987.
- [111] R. C. Gonzalez and R. E. Woods, *Digital image processing*. Prentice Hall, 2002.
- [112] Stone, H.S. and Tao, B. and McGuire, M., “Analysis of image registration noise due to rotationally dependent aliasing,” Nec Research Institute Tech. Rep., Tech. Rep. TR 98-018, 1998.
- [113] M. Barni, F. Bartolini, V. Cappellini, and A. Piva, “A DCT-domain system for robust image watermarking,” *Signal Process.*, vol. 66, no. 3, pp. 357–372, 1998.
- [114] Q. Li and I. J. Cox, “Using perceptual models to improve fidelity and provide resistance to volumetric scaling for quantization index modulation watermarking,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 2, pp. 127–139, June 2007. [Online]. Available: <http://dx.doi.org/10.1109/TIFS.2007.897266>
- [115] A. B. Watson, “DCT quantization matrices visually optimized for individual images,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, J. P. Allebach and B. E. Rogowitz, Eds., vol. 1913, Sep. 1993, pp. 202–216.
- [116] B. Chen and G. W. Wornell, “Dither modulation: a new approach to digital watermarking and information embedding,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, P. W. Wong & E. J. Delp, Ed., vol. 3657, Apr. 1999, pp. 342–353.
- [117] L. Schuchman, “Dither signals and their effect on quantization noise,” *Communication Technology, IEEE Transactions on*, vol. 12, no. 4, pp. 162–165, December 1964.
- [118] L. Perez-Freire and F. Perez-Gonzalez, “Security of lattice-based data hiding against the watermarked-only attack,” *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 4, pp. 593–610, Dec. 2008.

-
- [119] F. Perez-Gonzalez, F. Balado, and J. Martin, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *Signal Processing, IEEE Transactions on*, vol. 51, no. 4, pp. 960–980, Apr 2003.
- [120] P. Bas, "A quantization watermarking technique robust to linear and non-linear volumetric distortions using a fractal set of floating quantizers," in *Information Hiding*, 2005, pp. 106–117.
- [121] J. J. Eggers, R. Baeuml, and B. Girod, "Estimation of amplitude modifications before SCS watermark detection," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, E. J. Delp & P. W. Wong, Ed., vol. 4675, Apr. 2002, pp. 387–398.
- [122] K. Lee, D. Kim, T. Kim, and K. Moon, "EM estimation of scale factor for quantization-based audio watermarking," in *Proc. Second Int. Workshop Digital Watermarking*, I. Cox, T. Kalker, and Y. M. Ro, Eds., Seoul, Korea, Oct. 2003, pp. 316–327.
- [123] A. Abrardo and M. Barni, "Informed watermarking by means of orthogonal and quasi-orthogonal dirty paper coding," *Signal Processing, IEEE Transactions on*, vol. 53, no. 2, pp. 824–833, Feb. 2005.
- [124] J. C. Oostveen, T. Kalker, and M. Staring, "Adaptive quantization watermarking," in *Proc. SPIE Security Watermarking Multimedia Contents*, E. J. Delp and P. W. Wong, Eds., vol. 5306, San Jose, CA, Jun. 2004, pp. 296–303.
- [125] V. Licks, F. Ourique, R. Jordan, and F. Perez-Gonzalez, "An exact expression for the bit error probability in angle QIM watermarking under simultaneous amplitude scaling and awgn attacks," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, 18-23, 2005, pp. 801–804.
- [126] F. Ourique, V. Licks, and R. Jordan, "Angle QIM: on document to watermark ratio analysis," in *Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on*, vol. 1, 28-31, 2005, pp. 111–114.

REFERENCES

- [127] K. Whelan, G. Silvestre, and N. Hurley, “Iterative decoding of scale invariant image data-hiding,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 1, Sept. 2005, pp. I-989–92.
- [128] A. Abrardo, M. Barni, F. Perez-Gonzalez, and C. Mosquera, “Improving the performance of RDM watermarking by means of trellis coded quantisation,” *Information Security, IEE Proceedings*, vol. 153, no. 3, pp. 107–114, Sept. 2006.
- [129] J. J. Garcia-Hernandez, M. Nakano-Miyatake, and H. Perez-Meana, “Data hiding in audio signal using rational dither modulation,” *IEICE Electronics Express*, vol. 5, no. 7, pp. 217–222, 2008.
- [130] P. Campisi and A. Neri, “Video watermarking in the 3D-DWT domain using perceptual masking,” in *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, vol. 1, Sept. 2005, pp. I-997–1000.
- [131] M. Deng, T. Bianchi, A. Piva, and B. Preneel, “An efficient buyer-seller watermarking protocol based on composite signal representation,” in *MM&Sec ’09: Proceedings of the 11th ACM workshop on Multimedia and security*. New York, NY, USA: ACM, 2009, pp. 9–18.
- [132] F. Guerrini, R. Leonardi, and M. Barni, “Image watermarking robust against non-linear value-metric scaling based on higher order statistics,” *Proc. ICASSP*, vol. 5, pp. 397–400, May 2006.
- [133] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing (4th Edition)*, 4th ed. Prentice Hall, 4 2006.
- [134] The Lame Project, <http://lame.sourceforge.net/>.