On Distortion-Compensated Dither Modulation Data-Hiding with Repetition Coding

Pedro Comesaña[†], Fernando Pérez-González[†]* and Félix Balado[‡]

Abstract

An exhaustive analysis of the distortion-compensated dither modulation (DC-DM) data-hiding method with repetition coding is presented. Two decoding strategies, ML lattice decoding and Euclidean distance decoding, are discussed and some simplifications presented. An exact performance analysis in terms of the bit error rate (BER) is given; such an exact analysis is currently not available in the literature. Two methods for computing the exact BER and several approximations and bounds, most of them in closed form, are provided. These approximations are employed to propose two novel improvements on the standard DC-DM method with repetition: the use of a weighted Euclidean distance, with optimizable weights, and a vector form of the distortion compensation parameter. Both account for significant performance improvements. DC-DM is compared with quantization methods in the projected domain, showing worse performance against additive noise attacks, but higher robustness to cropping attacks. A performance analysis of DC-DM under coarse quantization, that can be specialized to JPEG compression is also supplied. All our results are validated with numerical simulations with both synthetic data and real images.

Index Terms

Watermarking, quantization-based data hiding, side information, lattice decoding.

† Dept. Teoría de la Señal y Comunicaciones, ETSI Telecom., Universidad de Vigo, 36200 Vigo, Spain. Phone: +34 986 812124. Fax: +34 986 812116. E-mails: pcomesan@gts.tsc.uvigo.es and fperez@gts.tsc.uvigo.es

‡ Department of Computer Science, University College Dublin (National University of Ireland), Belfield Campus, Dublin 4, Ireland. Phone: +353 1 716 2454. Fax: +353 1 269 7262. E-mail: fiz@ihl.ucd.ie

* Corresponding author. E-mail: fperez@gts.tsc.uvigo.es

This work was partially funded by *Xunta de Galicia* under projects PGIDT04 TIC322013PR and PGIDT04 PXIC32202PM; MEC project DIPSTICK, reference TEC2004-02551/TCM; FIS project IM3, reference G03/185, Enterprise Ireland (Research Grant 2002/230) and European Comission through the IST Programme under Contract IST-2002-507932 ECRYPT.

On Distortion-Compensated Dither Modulation Data-Hiding with Repetition Coding

I. Introduction

Research in data hiding has redoubled efforts since the turning point entailed by the embracement of the side-informed approach. The first rigorous appearance of side information in watermarking only took place when Chen and Wornell [1] demonstrated that their Distortion-Compensated Quantization Index Modulation (DC-QIM) method embodied the same desirable features as a scheme for canceling known interference introduced more than a decade before by Costa [2]. Thus, it was first shown that data hiding with side information *only* at the encoder —i.e., blind data hiding— was possible with the same performance attainable when that side information was also available at the decoder.

The basic procedure of DC-QIM involves the quantization of a given host signal using a multidimensional quantizer selected from a finite set by the message to be embedded. A fundamental feature is that the watermarked signal is obtained by adding back to the quantized host signal the quantization error scaled depending on an optimizable parameter. This *distortion compensation* is what makes DC-QIM equivalent to Costa's scheme, as a proper choice of the parameter is known to yield the non-blind achievable rate under additive white Gaussian distortion independent of the host [3], [2]. Chen and Wornell also gave the first proposal to put DC-QIM in practice with Distortion-Compensated Dither Modulation (DC-DM), a particular case in which the set of quantizers are dithered (shifted) versions of a basic one. Due to the implementation and design issues associated to multidimensional quantizers, this basic quantizer usually relies on the Cartesian product of scalar lattice quantizers. DC-DM based on uniform scalar quantization is straightforward to implement and more easily amenable to analysis than other more complex settings. A number of additional works have also aimed at building practical methods based on Costa's result. Among them we have the Scalar Costa Scheme (SCS) [4] and the Scaled Bin Encoding (SBE) [5] — which are completely equivalent to DC-DM with uniform scalar quantizers—, the continuous periodic functions for self-noise suppression (CP-SNS) [6], and others.

Although DC-DM with uniform scalar quantizers is a suboptimal side-informed scheme, it is well known that has an achievable rate often acceptably close to the ideal limit [3], [4]. Nevertheless, performance analyses for the probability of decoding error of DC-DM are scarce, and usually either incomplete or inexact. Among previous attempts, we may cite first those ones devoted to determine the

decoding performance of DM, i.e., without distortion compensation [7], [8], [9]. Also, upperbounding strategies to DC-DM with repetition coding were studied in [10], whereas an approximation to the bit error rate of generic DC-QIM methods is also given in [3]. In order to contribute to this research area, the main objective of this paper is to provide a thorough analysis of DC-DM with uniform scalar quantizers and repetition coding, presenting accurate theoretical approximations and bounds to the bit error rate at the decoder. Building on our analysis, we also propose enhancements on this standard scheme, both by means of optimizable weights on the standard Euclidean-distance lattice decoder, and by introducing a novel vectorial structure for the distortion-compensation parameter. Finally, we analyze the behavior of the method under coarse quantization.

a) Notation and Framework: We will denote scalar random variables with capital letters (e.g., X), and their outcomes with lowercase letters (e.g., x). The same notation criterion applies to random vectors and their outcomes, denoted in this case by bold letters (e.g., $\mathbf{X}$, $\mathbf{x}$). We assume without loss of generality that the host signal is represented by a zero-mean random vector $\mathbf{X}^o = (X_1^o, \cdots, X_N^o)^T$. If necessary, these particulars can always be achieved by subtracting any non-zero mean from the host, and by using an arbitrary bijective transformation from the original arrangement of the host signal samples to a unidimensional one. Before embedding we apply a key-dependent pseudorandom permutation $\Pi(\cdot)$ to $\mathbf{X}^o$. The permuted host $\mathbf{X} \triangleq \Pi(\mathbf{X}^o)$ is partitioned into M subvectors $\mathbf{X}_j \triangleq (X_{L\cdot(j-1)+1}, \cdots, X_{L\cdot(j-1)+L})^T$, for $j = 1, \cdots, M$, and assuming for notational simplicity that $L \triangleq N/M$ is integer. Apart from the security increase due to the uncertainty that this permutation procedure causes to an attacker unaware of the key, an important side advantage is that of facilitating the analysis. This is due to the fact that the pseudorandom selection of the elements in each subvector $\mathbf{X}_j$ approximately grants their statistical independence. This hypothesis of approximate independence usually holds true for natural signals, as long as L is not of the same order as N. Moreover, we will show in Section IV-D that the pseudorandom partitions above are also advantageous from a performance point of view.

The watermarked signal $\mathbf{Y}$ will be obtained from both the host signal $\mathbf{X}$ and the information message $\mathbf{b}$ to be conveyed. We will assume, once again without loss of generality, that $\mathbf{b} = (b_1, \cdots, b_M)^T$ is a P-ary vector, with b_j taking values uniformly in $\{0, \cdots, P-1\}$ for $j = 1, \cdots, M$. A particular symbol b_j will be embedded using the subvector $\mathbf{X}_j$ to get $\mathbf{Y}_j$. As all the subvectors are obtained the same way, notice that we will only need to focus our attention on one arbitrary subvector for analytical purposes. In particular, note that the average host signal power in each partition will tend to be approximately the same as L increases. Denoting this value as D_h, and using the intra-partition independence assumption,

we can write $D_h \approx D_h^{(j)} = \frac{1}{L}\sum_{i=(j-1)\cdot L+1}^{j\cdot L}\sigma_{X_i}^2, \quad j = 1, \cdots, M$, where $\sigma_{X_i}^2 \triangleq \text{Var}\{X_i\}$ and $D_h^{(j)}$ denotes the average host signal power in the j-th partition.

The imperceptibility of the differences between $\mathbf{X}$ and $\mathbf{Y}$ has to be guaranteed by means of a perceptual analysis of the host signal previous to the embedding operation. This procedure is intrinsically dependent on the type of host signal in question. Due to this fact, we will consider henceforth that the host is a multimedia signal given in a certain domain of interest. The only requirement is that the domain chosen is suited to compute a *perceptual mask* $\boldsymbol{\alpha}$, taking into account human perceptual features. We assume in the following that the maximum energy for an unnoticeable modification of the corresponding host signal sample X_i is proportional to α_i^2.

Before closing this section, we need some basic concepts about lattices [11]. Let $\|\cdot\|$ denote the Euclidean norm. Given a (possibly translated) lattice Λ in an L-dimensional Euclidean space, we associate to it its nearest-neighbor quantizer $Q_\Lambda(\cdot)$ which is defined in such a way that, for an arbitrary vector $\mathbf{x}$, it yields $Q_\Lambda(\mathbf{x}) \in \Lambda$, such that $\|\mathbf{x} - Q_\Lambda(\mathbf{x})\|$ is minimum. Given a lattice Λ, let $\mathcal{V}(\Lambda)$ denote the quantization region associated with that centroid of Λ located at the origin. Then, we will write $\mathbf{x} \mod \Lambda$ to denote the vector $(\mathbf{x} - Q_\Lambda(\mathbf{x})) \in \mathcal{V}(\Lambda)$.

The remainder of this paper is organized as follows: Section II presents the standard DC-DM method with uniform scalar quantizers and repetition coding, and discusses two main decoding strategies: maximum likelihood (ML) lattice decoding and Euclidean distance decoding, with some useful approximations. Section III is devoted to providing a complete analysis of the performance of the scheme in terms of its bit error rate (BER), with several approximations and bounds. Standard DC-DM is improved in Section IV with the proposal of weighted Euclidean distance decoding and a vectorial distortion compensation parameter. We show that the weighting allows for near-ML decoding, and we give a geometrical interpretation of this improvement. In addition we show that the vectorial compensation parameter is profitable in the realistic case of varying watermark-to-noise power ratio at each host signal sample. Section V focuses on the adaptation of our theoretical analysis to coarse-quantization attacks, mainly JPEG compression. Empirical results validating our theoretical derivations and a comparison with trellis-based embedding are presented in Section VI, and our main conclusions summarized in Section VII.

II. DC-DM WITH UNIFORM SCALAR QUANTIZERS

We describe next the implementation of DC-DM, generalizing Chen and Wornell's proposal [3] to account for perceptual constraints as done in [9]. We restrict our presentation to any of the L-dimensional subvectors inside which the host signal samples can be assumed independent, dropping the subindex j in

the sequel for notational simplicity. Let us assume that the information symbol b is hidden using DC-DM inside the host $\mathbf{X}$. Then, we denote by

$$\mathbf{E} \triangleq Q_b(\mathbf{X}) - \mathbf{X}, \tag{1}$$

the quantization error resulting from quantizing $\mathbf{X}$ with the quantizer $Q_b(\cdot)$ corresponding to the b-th symbol, which is based on a minimum Euclidean distance criterion. The watermarked signal $\mathbf{Y}$ is then obtained as

$$\mathbf{Y} \;\; = \;\; \mathbf{X} + \nu\mathbf{E} = Q_b(\mathbf{X}) - (1-\nu)\mathbf{E}, \tag{2}$$

The distortion-compensation parameter ν, $0 < \nu \leq 1$, is an optimizable variable akin to the one in Costa's paper. The component $(1-\nu)\mathbf{E}$ may be termed as self-noise, since it is caused by the watermarking process itself due to the distortion compensation. As we will see in Section IV-A, performance improvements are obtained by using $\nu < 1$, i.e., allowing a certain degree of self-noise.

Dither modulation means that all the quantizers $Q_b(\cdot)$ are just shifted versions of a basic quantizer $Q_\Lambda(\cdot)$. The offset for obtaining each one of these quantizers is a dither vector $\mathbf{v}(b)$ that depends on both a secret key and the message to be sent b. Then, the quantizer $Q_b(\cdot)$ can be put as

$$Q_b(\mathbf{X}) = Q_\Lambda\left(\mathbf{X} - \mathbf{v}(b)\right) + \mathbf{v}(b). \tag{3}$$

As aforementioned, the simplest and more widespread implementation of DC-DM is the one by means of uniform scalar quantizers [3], [7], [12], [4], [8]. In this case $Q_\Lambda(\cdot)$ may be defined as the quantizer whose quantization centroids are given by the points in the lattice $\Lambda = P\Lambda'$, with $\Lambda' \triangleq (\Delta_1\mathbb{Z}, \Delta_2\mathbb{Z}, \cdots, \Delta_L\mathbb{Z})^T$. We will impose the criterion that the dither vectors $\mathbf{v}(b)$ are such that the distance between the closest centroids of the quantizers corresponding to any two different symbols is maximized. This just means that, for instance, $\mathbf{v}(b) = b \cdot (\Delta_1, \cdots, \Delta_L)^T + \mathbf{d}$, where $\mathbf{d}$ is a key-dependent vector deterministically known to both encoder and decoder. This strategy increases the robustness of the embedding by placing the centroids as far away as possible. Also, the resultant symmetry allows to assume an arbitrary embedded symbol b for the analysis, as we will see later.

Notice that, for $L > 1$, this particular choice of the dither vectors amounts to using a repetition code. It is well known that, even though it is useful in many practical situations (e.g., see [7], [12], [8]), this channel coding strategy is not the optimal one. It is pertinent to note that an empirical study on the concatenation of repetition coding for SCS (DC-DM) with near-optimal turbo codes was given in [4]. From the results in that work, it is possible to conclude that the concatenation of turbo codes and repetition is quite close to the capacity limit for Gaussian channels at low embedding rates. Then, the

appeal of this scheme lies in the fact that it presents evident advantages from the complexity point of view yet keeping quite a good performance. This result adds an interesting practical perspective to the analysis of DC-DM with repetition coding. In Section VI we will provide additional empirical evidence on the usefulness of the concatenation of this scheme with an outer turbo code, using a channel model resulting from our analysis.

In order to keep the exposition simple we will only study the case $P = 2$ (i.e., binary), but the approach we will follow can be extended for arbitrary alphabet sizes. We remark that, to the best of our knowledge, a rigorous performance analysis in terms of probability of error is not available even for this relatively simple case. For the binary case, the quantization centroids for $Q_b(\cdot)$ will be given by the shifted lattice $\Lambda_b = 2\Lambda' + b \cdot (\Delta_1, \cdots, \Delta_L)^T + \mathbf{d}$, for $b \in \{0, 1\}$.

The use of scalar lattices inherently introduces an amplitude-limited embedding distortion. Since we can write (1) as $\mathbf{E} = (\mathbf{v}(b) - \mathbf{X}) \mod \Lambda$, it follows that $\mathbf{E}$ will be uniformly distributed over $\mathcal{V}(\Lambda)$ when $\mathbf{v}(b)$ is a deterministic vector if and only if $\mathbf{X} \mod \Lambda$ satisfies the same condition (uniform condition). For typical continuous distributions this will be the case if $\sigma_{X_i} \gg \Delta_i$ for all i. Due to perceptual constraints, for most watermarking scenarios the uniform condition will approximately hold, and hence we will assume hereafter that $\sigma_{X_i} \gg \Delta_i$ for all i, and so that $E_i \sim U(-\Delta_i, \Delta_i]$.

Noticing that the watermark signal is given by $\mathbf{W} = \mathbf{Y} - \mathbf{X} = \nu\mathbf{E}$, it is clear that its energy per dimension will be $\mathrm{E}\{W_i^2\} = \nu^2\Delta_i^2/3$. According to the perceptual mask assumed in the previous section, we can achieve the maximum unnoticeable embedding distortion by choosing Δ_i to be proportional to α_i. Last, the embedding distortion in the subvector under analysis is defined as $D_w \triangleq \frac{1}{L}\sum_{i=1}^{L} \mathrm{E}\{W_i^2\}$.

A. Attack Channel

Decoding is accomplished by the receiver after the watermarked signal $\mathbf{y}$ has undergone an attack channel. Throughout most of the paper —with the remarkable exception of Section V— we will assume that this channel is a zero-mean additive probabilistic channel independent of $\mathbf{X}$ and $\mathbf{b}$, yielding a received signal $\mathbf{Z} = \mathbf{Y} + \mathbf{N}$. This type of channel model has been consistently used for benchmarking purposes in most relevant data hiding research. Recalling that the elements of the L-length subvectors are pseudorandomly chosen through the permutation $\Pi(\cdot)$, we may also assume that the samples in $\mathbf{N}$ are mutually independent, with diagonal covariance matrix $\mathbf{\Gamma} = \mathrm{diag}(\sigma_{N_1}^2, \cdots, \sigma_{N_L}^2)$. The *channel distortion* D_c can be then defined in a similar fashion as the embedding distortion, i.e., $D_c \triangleq \frac{1}{L}\sum_{i=1}^{L} \sigma_{N_i}^2$. We would like to remark that this kind of measurement would in principle allow to concentrate all the attacking distortion on a single sample of $\mathbf{Y}$ or spread it over all the vector. This freedom to distribute

distortion hints at the poor connection existing between perceptual issues and this kind of mean square error (MSE) distortion measurements.

Nevertheless, we will undertake all subsequent analyses using MSE, as this criterion has been the most employed in the literature so far for the sake of tractability. Notice, for instance, that the hypotheses of Costa's result are stated for this type of restriction. In any case, an attacker may try to partially relieve the intrinsic inconveniences of MSE in order to comply with the usual requirement of minimal perceptual impact of the attack. Assuming the adequacy of the perceptual mask, it is clear that one way to meet this condition is to *perceptually shape* the added noise, such that its variance at each dimension is proportional to the corresponding allowable perceptual energy. Last, we will find it useful to introduce the *watermark-to-noise ratio* as WNR $\triangleq \frac{D_w}{D_c}$, that relates the power of the embedding and channel distortion, establishing a working point similar to the signal-to-noise ratio (SNR) in communications.

B. Modulo-Lattice Equivalent Noise

Without loss of generality, and assuming hereafter that all embedded symbols are equally likely, we will focus our analysis on any given symbol b. The optimal decoding criterion that minimizes the bit error rate is the ML decision given by

$$\hat{b} = \arg\max_{b\in\{0,1\}} f_{\mathbf{Z}|B}(\mathbf{z}|b). \tag{4}$$

According to the preceding exposition, the samples in $\mathbf{Z}$ can be assumed to be mutually independent, so we can expand (4) as

$$\begin{aligned}\hat{b} &= \arg\max_{b\in\{0,1\}} \prod_{i=1}^{L} f_{Z_i|B}(z_i|b) \\ &= \arg\max_{b\in\{0,1\}} \prod_{i=1}^{L} \int_{-\infty}^{\infty} f_{Y_i|B}(z_i - r_i|b) f_{N_i}(r_i) dr_i,\end{aligned}$$

where $f_{Y_i}(\cdot)$ and $f_{N_i}(\cdot)$ are the probability density functions (pdf) of the independent random variables Y_i and N_i, respectively. The main drawback of this approach is that it requires prior knowledge about the host signal pdf. Also, the ML approach to DC-DM decoding can be too costly since we have to take into account f_{X_i}, the sent bit and the dither in order to compute f_{Y_i}. Therefore, simplifications to it are desirable. One such simplification, which we will assume throughout the remainder of the paper, is lattice decoding. Lattice decoding rules can be seen as operating over variables that are reduced modulo-Λ. In our case, the decision will be based on the statistics $\tilde{z}_i \triangleq \frac{z_i - Q_0(z_i)}{\Delta_i}$, $i = 1, \cdots, L$, where $Q_0(z_i)$ is the i-th component of $Q_0(\mathbf{z})$. From the way it is constructed, it is clear that $\tilde{z}_i \in (-1, 1]$; this leads to considering

modulo-$2\mathbb{Z}^L$ vector reductions, for which the result belongs to $(-1,1]^L$. Also, the normalization in the definition of $\tilde{z}_i$ is reasonable if we assume that the channel noise is perceptually shaped, as in that case its variance will be roughly proportional to Δ_i^2.

Let $f_{\tilde{\mathbf{Z}}}(\tilde{\mathbf{z}})$ denote the pdf of $\tilde{\mathbf{Z}}$. Then, the ML lattice decoder will choose $\hat{b}$ according to the rule

$$\hat{b} = \underset{b\in\{0,1\}}{\arg\max}\, f_{\tilde{\mathbf{Z}}|B}(\tilde{\mathbf{z}}|b). \tag{5}$$

Erez and Zamir showed in [13] that lattice decoding (as a less computationally demanding alternative to maximum likelihood decoding) can achieve capacity under certain conditions. In fact they proved this result for both the ML lattice decoder (also termed by the authors *noise-matched lattice decoder*) and the Euclidean lattice decoder.

As we shall see in Section III, performance analysis requires to determine the distribution of the noise in the decision statistics, which we tackle next. Let us define the *total noise* random variable as

$$T_i \triangleq \frac{-(1-\nu)E_i + N_i}{\Delta_i}. \tag{6}$$

Recalling that if X, Y are two random variables related by $Y = aX$, their pdfs satisfy $f_Y(y) = |a|^{-1} f_X(y/a)$, and that the pdf of the sum of two independent random variables is the convolution of the respective pdfs, we can write

$$f_{T_i}(t_i) = \frac{\Delta_i^2}{(1-\nu)} \left[f_{N_i}(t_i\Delta_i) * f_{E_i}(t_i\Delta_i/(1-\nu)) \right], \tag{7}$$

where $*$ denotes convolution, and $E_i \sim U(-\Delta_i, \Delta_i]$, for all $i = 1, \cdots, L$. Now, the *modular total noise* random variable $\mathbf{U}$ is simply defined as $\mathbf{U} \triangleq \mathbf{T} \mod 2\mathbb{Z}^L$.[1] Consequently, the support of U_i will be the interval $(-1, 1]$, for all $i = 1, \cdots, L$.

Considering (6), the pdf of U_i can be written as

$$f_{U_i}(u_i) \;=\; \begin{cases} \sum_{l=-\infty}^{\infty} f_{T_i}(u_i - 2l), & \text{if } u_i \in (-1, +1] \\ 0, & \text{otherwise} \end{cases}. \tag{8}$$

Alternatively, $f_{U_i}(u_i)$ can be written as $f_{U_i}(u_i) = \frac{\Delta_i^2}{(1-\nu)} \left[f_{N_i}(u_i\Delta_i) \circledast_2 f_{E_i}(u_i\Delta_i/(1-\nu)) \right]$, with $\circledast_2$ the circular convolution over $(-1, 1]$, which includes the aliasing effect evident in (8). For any two arbitrary

[1]Note that $\mathbf{U}$ is not the same as $\tilde{\mathbf{Z}}$, since the latter will depend on the bit sent.

pdfs $f_B(x)$ and $f_C(x)$ this operation is defined as

$$f_B(x) \circledast_2 f_C(x) \triangleq \begin{cases} \sum_{l=-\infty}^{\infty} \int_{-\infty}^{\infty} f_B(y-2l) f_C(x-y) dy, \\ \qquad -1 < x \leq +1, \\ 0, \\ \qquad \text{otherwise} \end{cases},$$

A similar technique has been used in [4] to show the independence of the quantization error and the host signal when a uniform dither is used. In [4] the role of the circular convolution is played by the sampling of the characteristic function with period π. This sampling has an aliasing effect, since it is equivalent to the convolution in the time domain with an impulse train with period 2.

When the symbol b is embedded, it is clear from (2) that the decision statistics will take the form $\tilde{z}_i = \frac{Q_b(x_i)-(1-\nu)e_i+n_i-Q_0(z_i)}{\Delta_i} \mod 2\mathbb{Z} = (u_i + b) \mod 2\mathbb{Z}$, for all $i = 1, \cdots, L$, or, in short, $\tilde{\mathbf{z}} = (\mathbf{u} + b\mathbf{1}) \mod 2\mathbb{Z}^L$, where $\mathbf{1}$ is a vector of L ones. Equivalently, $\mathbf{u} = (\tilde{\mathbf{z}} - b\mathbf{1}) \mod 2\mathbb{Z}^L$. Then, the decision rule in (5) is equivalent to deciding $\hat{b} = 0$ whenever

$$f_{\mathbf{U}}(\tilde{\mathbf{z}}) > f_{\mathbf{U}}\left((\tilde{\mathbf{z}} - \mathbf{1}) \mod 2\mathbb{Z}^L\right), \tag{9}$$

and $\hat{b} = 1$ otherwise.

1) An Approximation to the ML Lattice Decoder: In [14] Forney et al. provided useful approximations to the pdf of a modulo-reduced —or aliased— Gaussian pdf. The same approach can be followed to write an approximation of the pdf of the *modular total noise* random variable U_i, defined in (8) and needed in (9) for ML lattice decoding.

Recall from (7) that, for $N_i \sim \mathcal{N}(0, \sigma_{N_i}^2)$, the pdf of T_i is just the convolution of a Gaussian with zero-mean and variance $\sigma_{N_i}^2/\Delta_i^2$, and a uniform pdf in $(-(1-\nu), +(1-\nu)]$. Pursuing the sort of approximations proposed in [14], it is possible to conclude that:

- For $\sigma_{T_i} \ll 1$, the contributions in the summation in (8) for $l \neq 0$ are negligible. The most significant part of the pdf of T_i is concentrated in the interval $(-1, +1]$, so the aliasing effect can be neglected. Therefore $f_{U_i}(u_i)$ can be well approximated by $f_{T_i}(t_i)$.
- For $\sigma_{T_i} \gg 1$, it is possible to consider that T_i follows a Gaussian distribution with $\sigma_{T_i}^2 = \sigma_{N_i}^2/\Delta_i^2 + (1-\nu)^2/3$. Now the pdf $f_{U_i}(u_i)$ becomes nearly constant due to the strong aliasing. Observe that since $\sum_l f_{T_i}(u_i - 2l)$ in (8) is periodic if we do not restrict u_i to lie on $(-1, +1]$, it makes sense to expand it in terms of its Fourier series and then truncate it to this interval. Forney et al. suggested approximating this function by keeping the low-frequency terms of this expansion.

The computation of the Fourier series expansion of a periodic function on a lattice can be performed by using the dual of that lattice [11]. As in our case U_i is obtained by means of the lattice $2\mathbb{Z}$, the corresponding dual lattice is simply given by $\mathbb{Z}/2$, and so the desired pdf can be expanded as [14]

$$\begin{aligned} f_{U_i}(u_i) &= \frac{1}{2}\sum_{k\in\mathbb{Z}} \exp\left(-\pi^2\sigma_{T_i}^2 k^2/2\right) e^{j2\pi k u_i/2}, \\ & \quad -1 < u_i \leq 1. \end{aligned} \tag{10}$$

The DC and fundamental frequency terms in this expansion correspond to $k = 0$ and $k = \pm 1$, respectively. Keeping just these two terms in (10), we can write $f_{U_i}(u_i) \approx \frac{1}{2}\left(1 + 2\mathrm{e}^{(-\pi^2\sigma_{T_i}^2)/2}\cos(\pi u_i)\right)$, for $-1 < u_i \leq 1$. The usefulness of this approximation is illustrated in Section IV-C, where a geometrical interpretation of lattice ML decision regions is provided.

2) Euclidean and Weighted Euclidean Distance-based Lattice Decoder: Despite the complexity reduction from ML decoding to ML lattice decoding and the further diminution brought about by Forney's approximation, it is desirable to seek even simpler decoding strategies. In this section we discuss lattice decoding based on the Euclidean distance.

When each dimension is normalized by its quantization step, Euclidean lattice decoding can be written as

$$\hat{b} = \underset{b\in\{0,1\}}{\arg\min} \|\mathbf{\Delta}^{-1}(\mathbf{z} - Q_b(\mathbf{z}))\|^2, \tag{11}$$

where $\mathbf{\Delta} \triangleq \mathrm{diag}(\Delta_1, \cdots, \Delta_L)$. This approximation is tantamount to choosing the b whose associated shifted lattice Λ_b yields the minimum normalized quantization error. Minimum distance decoding of DC-DM was in fact part of the original proposal of DC-DM in [3]. It is also used in [7], [8], and in [4] for the equivalent Scalar Costa Scheme (SCS).

To see the relationship between this decoding strategy and ML lattice decoding, let $\mathcal{S} \triangleq \{\pm 1\}^L$. Recalling the definition of $\tilde{\mathbf{z}}$, it is clear that for $b = 0$, $\|\mathbf{\Delta}^{-1}(\mathbf{z} - Q_b(\mathbf{z}))\| = \|\tilde{\mathbf{z}}\|$, while for $b = 1$, $\|\mathbf{\Delta}^{-1}(\mathbf{z} - Q_b(\mathbf{z}))\|$ becomes $\|\tilde{\mathbf{z}} - \mathbf{s}\|$, where $\mathbf{s} \in \mathcal{S}$ is such that $\|\tilde{\mathbf{z}} - \mathbf{s}\|$ is minimum. Putting this together, we can rephrase the decoding rule in (11) as deciding $\hat{b} = 0$ if

$$\|\tilde{\mathbf{z}}\|^2 < \min_{\mathbf{s}\in\mathcal{S}} \|\tilde{\mathbf{z}} - \mathbf{s}\|^2, \tag{12}$$

and $\hat{b} = 1$ otherwise.

But now, the vector $\mathbf{s}$ minimizing the norm in the right hand side of (12) is such that it also satisfies $(\tilde{\mathbf{z}} - \mathbf{s}) = (\tilde{\mathbf{z}} - \mathbf{s}) \mod 2\mathbb{Z}^L$. Thus, the parallelism between (12) and (9) is clear if one considers that the modulo-$2\mathbb{Z}^L$ operation maps the set $\mathcal{S}$ onto vector $\mathbf{1}$, i.e., for all $\mathbf{s} \in \mathcal{S}$, $(\tilde{\mathbf{z}} - \mathbf{s}) \mod 2\mathbb{Z}^L = (\tilde{\mathbf{z}} - \mathbf{1}) \mod 2\mathbb{Z}^L$.

In fact, the two decoding rules would be equivalent if the modular total noise $\mathbf{U}$ had Gaussian independent and identically distributed (i.i.d.) components. It is convenient to examine under which conditions this latter property would hold. First, in order to neglect overlaps of the shifted versions of f_{T_i} in the construction of f_{U_i} in (8), it is required that $\sigma_{T_i} \ll 1$ for all i. Second, for the T_i to be i.i.d. Gaussian, a necessary and sufficient condition would be that $\nu = 1$ (i.e., there is no self-noise) and that the noise components N_i are independent Gaussian, with variances proportional to Δ_i^2 for all i. Notice that in general these conditions will not be satisfied, so Euclidean distance decoding will be suboptimal.

When those conditions are not met, it is useful to modify minimum distance decoding while still retaining a relatively simple decoding approach by comparison with ML lattice decoding. To this end, we introduce a weighted Euclidean distance, for which the decoding rule becomes

$$\hat{b} \;=\; \underset{b\in\{0,1\}}{\arg\min} \left\{ (\mathbf{z} - Q_b(\mathbf{z}))^T \, \boldsymbol{\Delta}^{-1}\mathbf{B}\boldsymbol{\Delta}^{-1} \, (\mathbf{z} - Q_b(\mathbf{z})) \right\}, \tag{13}$$

where the weighting matrix $\mathbf{B}$ is defined as

$$\mathbf{B} \triangleq \; \text{diag}(\beta_1, \cdots, \beta_L).$$

The purpose of these weights is to introduce additional degrees of freedom to improve decoding in practice when minimum distance decoding is just too far away from optimality. We will show in Section IV-B how a proper design of the parameter vector $\boldsymbol{\beta}$ allows to improve decoding when additional information about the channel noise is available. Also, it should be taken into account that the normalization by Δ_i in (11) does not entail any loss of generalization or loss in performance, since its effect could be canceled by β_i in any case. Whenever no optimization is attempted, $\beta_i = 1$ will be set for all i. In this case (13) becomes equivalent to (11).

III. Performance Analysis

Next, we will analyze the performance of binary repetition DC-DM in terms of the bit error rate (BER) at the decoder output. In this section we will consider only the unweighted minimum Euclidean distance approach to DC-DM decoding, i.e., $\mathbf{B} = \mathbf{I}_{L\times L}$, the identity matrix of size L, leaving to Section IV-B the study of the effect of the weights $\boldsymbol{\beta}$.

Costa's framework considers i.i.d. channel noise and signal (in our case watermark), which naturally induces the same distortion compensation parameter ν for all dimensions. On the other hand, perceptual considerations motivate that in our scheme the variances of both the channel noise and the watermark

be in general different for each of the dimensions. Although this setting suggests a vectorial distortion compensation parameter $\boldsymbol{\nu}$ —i.e., dimension-dependent—, for the sake of simplicity we will only deal with a scalar ν in this section, and explore the vectorial possibility in Section IV-A.

From (13) we can see that decoding is equivalent to quantizing $\mathbf{z}$ with both the shifted lattice Λ_0 and Λ_1 and then assigning the value of the bit that yields the smallest (in an Euclidean distance sense) normalized quantization error. Obviously, this is completely equivalent to quantizing $(\boldsymbol{\Delta}^{-1}\mathbf{z})$ with $(\boldsymbol{\Delta}^{-1}\Lambda_0) \cup (\boldsymbol{\Delta}^{-1}\Lambda_1)$ following also a minimum Euclidean distance criterion.

It can be readily seen that the probability of decoding error does not depend on the actual embedded bit. Let us assume then that $b = 0$ is sent, so $\tilde{\mathbf{Z}} = \mathbf{U}$. Hence, taking (12) into account, an error happens whenever

$$||\mathbf{u}||^2 > \min_{\mathbf{s}\in\mathcal{S}} ||\mathbf{u} - \mathbf{s}||^2. \tag{14}$$

The minimization in (14) is equivalent to seeking the closest centroid to $\mathbf{u}$ among the shifted lattice corresponding to $b = 1$. The decoding region given by (14) is a generalized octahedron [11] whose vertices are those vectors having only one non-zero component with value $\pm L/2$.

Therefore, a decoding error will happen if and only if $\mathbf{u}$ lies out of this generalized octahedron. Due to the symmetry of the octahedron in all the orthants with respect to the origin, it is reasonable to project the random variable $\mathbf{U}$ onto the positive one, to construct $U_i^+ \triangleq |U_i|, 1 \leq i \leq L$, and then proceed to determine the probability of being closer to the vertex $\mathbf{s}_1 \triangleq \mathbf{1} \in \mathcal{S}$, than to the origin. This probability is thus the probability of bit error, which can be written as

$$P_e \;=\; \Pr\{\|\mathbf{U}^+\|^2 > \|\mathbf{U}^+ - \mathbf{1}\|^2\} = \Pr\left\{\sum_{i=1}^{L} U_i^+ > L/2\right\}. \tag{15}$$

The evaluation of this expression requires the pdf of U_i^+, $1 \leq i \leq L$, which is just

$$f_{U_i^+}(u_i^+) \;\triangleq\; \begin{cases} \left[f_{U_i}(u_i^+) + f_{U_i}(-u_i^+)\right], & \text{if } 0 \leq u_i^+ \leq 1 \\ 0, & \text{otherwise} \end{cases}, \quad 1 \leq i \leq L. \tag{16}$$

Therefore, if we define the variable

$$R \triangleq \sum_{i=1}^{L} U_i^+, \tag{17}$$

then from (15), the computation of P_e is equivalent to integrating the tail of the pdf of R from $L/2$ to L.

Even though formula (15) allows us to determine the *exact* probability of bit error, its computation is very expensive for large L. This motivates the proposal of two numerical approaches for its calculation, which are discussed in Sections III-A and III-B. On the other hand, neither formula (15) nor these practical methods provide closed-form expressions, making it difficult to extract conclusions of theoretical value. For this reason, Sections III-C and III-D are devoted to discussing analytical approximations and bounds respectively.

A. Beaulieu's Approach

In this section, we adapt a technique proposed by Beaulieu [15] for computing the tail probability of the summation of L i.i.d. random variables, as it occurs in (15). This technique was already used in [9] to upperbound the bit error probability of DM. Let $\omega_l \triangleq \frac{2\pi l}{T}$ for any positive integer l, with T a large enough real number, and let $F_{U_i^+}(\omega)$ be the characteristic function of U_i^+, given by $F_{U_i^+}(\omega) = \int_0^1 \mathrm{e}^{j\omega u} f_{U_i^+}(u)du$. Then, the computation of P_e is made, following [15], as

$$P_e \approx \frac{1}{2} + \frac{2}{\pi} \sum_{\substack{l=1 \\ l \text{ odd}}}^{\infty} \frac{\prod_{i=1}^{L} |F_{U_i^+}(\omega_l)| \sin(\sum_{i=1}^{L} \phi_i(\omega_l))}{l}, \tag{18}$$

where $\phi_i(\omega)$ is defined as $\phi_i(\omega) \triangleq \arg\{F_{U_i^+}(\omega)\} - \omega/2$, with $\arg(\cdot)$ denoting the four-quadrant phase. The main drawback of this method is that it is rather computationally demanding, apart from the fact that it may present numerical problems due to the large values that could be involved in the summation of a truncated version of the series in (18). In the appendix the expressions of the functions required for computing (18) for a Gaussian channel noise are derived.

B. DFT Method

Since the U_i^+ in (17) are independent random variables, the pdf of R is just the convolution of the pdfs of U_i^+, $1 \leq i \leq L$. This computation can be efficiently done in the Discrete Fourier Transform (DFT) domain. To that end, let $\Phi_{U_i^+} \triangleq \mathrm{DFT}_{LK}\left(K \cdot f_{U_i^+}(k/K)\right)$ be the $L \times K$-point DFT of the sequence obtained by sampling $f_{U_i^+}(\tau)$ at $\tau = \frac{k}{K}$, with $k = 0, \cdots, K-1$. Using this definition it is straightforward to write $\Phi_R[m] = \prod_{i=1}^{L} \Phi_{U_i^+}[m], \quad m = 0, \cdots, LK-1$. Finally, the discretized pdf of R is obtained using the Inverse Discrete Fourier Transform (IDFT) as $f_R = \mathrm{IDFT}_{LK}(\Phi_R)$, and (15) is computed as

$$P_e \approx \sum_{k=\left\lceil \frac{L(K-1)+1}{2} \right\rceil}^{LK-1} f_R[k],$$

where the limits of this summation stand for $R = L/2$ and $R = L$ in the corresponding integral. The accuracy of the computation can be increased by using a larger value of K, i.e., by sampling more finely the pdfs involved in the calculation.

This technique resembles Beaulieu's approach in that both of them work in a transform domain. Nevertheless, the DFT method presents a much lower computational cost, without any of the numerical problems shown by Beaulieu's approach. This fact makes the DFT method an enticing approach to assess the performance to any degree of accuracy required.

C. Central Limit Theorem-based Approximation

A third option consists in taking advantage of the independence of the random variables U_i^+ in the summation (17) to invoke the Central Limit Theorem (CLT). This result states that the distribution of R will tend to a Gaussian as $L \to \infty$, in which case we may approximate the probability of error as

$$P_e \approx \mathcal{Q}\left(\frac{\frac{L}{2} - \sum_{i=1}^{L} \mathrm{E}\{U_i^+\}}{\sqrt{\sum_{i=1}^{L} \mathrm{Var}\{U_i^+\}}}\right), \tag{19}$$

where $\mathcal{Q}(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{\tau^2}{2}} d\tau$.

The main advantage of CLT-based approximation is that it gives a closed expression for P_e, which can be exploited for analytical purposes (see Section IV-B). Although this method to compute P_e is much simpler than the previous ones, some remarks are due. First, for small values of L it could entail problems in the convergence of R to a Gaussian. One factor that speeds up convergence is the similarity between the distributions of the summands. Also, and as discussed in [10], note that the process of building the one-sided distributions $f_{U_i^+}(u_i^+)$ may produce highly skewed pdfs whose convolution converges very slowly to a Gaussian distribution. If this is the case, the Gaussian approximation to P_e may underestimate the importance of the tails of $f_R(r)$.

Last, although analytical expressions of $\mathrm{E}\{U_i^+\}$ and $\mathrm{Var}\{U_i^+\}$ are available in closed form when N_i follows a uniform distribution, in general the explicit computation of these statistics may require numerical integration. Therefore, it is recommended to use the DFT method for obtaining numerical results, as it gives a higher degree of accuracy.

D. Bounds on P_e

In this section we discuss several other known bounds on the bit error probability.

1) Erez and Zamir's Bound: Erez and Zamir have recently proposed a method that can be accommodated to upperbound the probability of error of binary DC-DM when performing Euclidean lattice decoding —i.e., minimum distance decoding— under Additive White Gaussian Noise (AWGN) channel distortion [13]. Let $\mathcal{W} = \mathcal{V}\left(\{2\mathbb{Z}^L\} \cup \{2\mathbb{Z}^L + \mathbf{1}\}\right)$ denote the region associated to a right decision, i.e., $P_e = \Pr\{\mathbf{U} \notin \mathcal{W}\}$. Then, it is possible to write $P_e \leq \Pr\{\mathbf{T} \notin \mathcal{W}\}$. Erez and Zamir's procedure may be used to construct an upper bound on the latter probability that depends on D_w, D_c, Λ and L. In turn, this obviously upperbounds P_e. Unfortunately, the bound turns out to be rather loose for our particular problem (see Section VI); for this reason, we will omit the details of its implementation.

2) Union Bound and Nearest Neighbor Approximation: The classical *union bound* (UB) is based on adding the pairwise probability of mistaking the transmitted centroid with each of its nearest neighbors corresponding to a wrong decision. The possible overlaps of the error regions associated to each of these error events are disregarded in this computation, and this is the reason why it produces an upper bound. When the WNR is increased these overlaps diminish, and so the bound gets closer to the true value. As in our implementation of DC-DM we are using uniform scalar quantizers, there are 2^L nearest error neighbors. Thus, assuming that the pdf of the channel distortion is symmetric, the union bound may be computed as

$$\begin{aligned} P_e &\leq P_{\text{union}} = 2^L \cdot \Pr\{\|\mathbf{U}\|^2 > \|\mathbf{U} - \mathbf{1}\|^2\} \\ &= 2^L \cdot \Pr\left\{\sum_{i=1}^{L} U_i > L/2\right\}, \end{aligned}$$

where the last probability can be obtained by means of any of the methods in Sections III-A–III-C, similarly to what is done with (15). Alternatively, for L large enough, we can compute an approximation applying the Central Limit Theorem. To this end, we just need to compute the variance of the zero-mean random variable whose pdf is the circular convolution of the channel noise and the self-noise. Note that due to the approximation implicit in the CLT, we can no longer ensure that the result is a bound, but an approximation to the bound, which will be asymptotically good as $L \to \infty$. This approximation is given by

$$P_e \approx 2^L \cdot \mathcal{Q}\left(\frac{L}{2\sqrt{\sum_{i=1}^{L} \text{Var}\{U_i\}}}\right). \tag{20}$$

In contrast to Section III-C, if N_i is symmetric about the origin the involved pdfs (i.e., those of $f_{U_i}(u_i)$) are also symmetric, so their convolution will converge more quickly to a Gaussian distribution.

Following the previous guidelines for the union bound we may also approximate the bit error probability using the nearest neighbor distance sketched in [3]. The estimate therein assumes Quantization Index

Modulation without distortion compensation and additive white Gaussian noise. This result may be improved by replacing the real Gaussian pdf with a Gaussian with variance the sum of those corresponding to the channel noise and the self-noise, what yields $P_e \sim \mathcal{Q}\left(\frac{L}{2\sqrt{\sum_{i=1}^{L} \mathrm{Var}\{U_i\}}}\right)$.

Following the discussion in [16] on the validity of the CLT, it is necessary to check against empirical results all the CLT-based approximations and bounds that we have given in Sections III-C and III-D.2. This task is undertaken in Section VI-A.

IV. Improvements on Standard DC-DM

In this section we introduce some improvements in the performance of the DC-DM scheme studied so far. Specifically, we will deal with the distortion compensation parameter as well as with the decoding weights.

A. Study of the Distortion Compensation Parameter

The distortion compensation parameter ν, may be used in two equivalent ways. Namely, it may reduce the embedding power by a factor ν^2 for a fixed lattice, or, alternatively, it may afford an expansion of the lattice by a factor $\frac{1}{\nu}$ when the power of the watermark is kept constant. Interestingly, it can be shown that both lead to the same bit error probability for a given WNR when the power spectral density of the noise sequence is fixed, save for a multiplicative constant. Therefore, although throughout this paper we are using a fixed lattice, we should be aware that, when the stated conditions are met, this is equivalent to the expansion of that lattice for a fixed D_w.

The determination of the distortion compensation parameter may be tackled under a number of different optimization criteria. Obviously, these criteria will in general lead to different values of ν. Probably the simplest, but also one of the most used, is the minimum mean square error (MMSE) criterion (see [17]). This criterion was for instance used in [13]. We may also think of optimizing this parameter depending on the bit error rate. The problem in this case is the lack of closed-form expressions that would allow to face the optimization problem in an analytical way. Following MMSE, the initial intention would be to minimize $\sum_{i=1}^{L} \sigma_{U_i}^2$; however, due to the aliasing effect, this becomes an unsurmountable problem. Considering that for large WNRs and large values of ν the modulo operation can be neglected, it is reasonable to address instead the minimization of

$$\begin{aligned} \varphi(\nu) &\triangleq \sum_{i=1}^{L} \sigma_{T_i}^2 = \sum_{i=1}^{L} \left\{ \frac{\sigma_{N_i}^2}{\Delta_i^2} + \frac{(1-\nu)^2}{3} \right\} \\ &= \sum_{i=1}^{L} \left\{ \frac{\nu^2 \xi_i}{3} + \frac{(1-\nu)^2}{3} \right\}, \end{aligned} \tag{21}$$

for a fixed $\xi_i \triangleq \sigma^2_{N_i}/\mathrm{E}\{W_i^2\}$, $i = 1, \cdots, N$. Note that ξ_i can be regarded to as a *noise to watermark ratio* for the i-th dimension. Function $\varphi(\nu)$ above can be easily seen to be minimized at

$$\nu^* = \frac{1}{1 + \frac{1}{L}\sum_{i=1}^{L} \xi_i}.$$

Alternatively, one may also consider using a different value of ν for each dimension. This yields a vector of distortion compensation parameters $\boldsymbol{\nu} \triangleq (\nu_1, \cdots, \nu_L)$, so (21) takes now the shape

$$\varphi(\boldsymbol{\nu}) \quad = \quad \sum_{i=1}^{L} \left\{ \frac{\nu_i^2 \xi_i}{3} + \frac{(1-\nu_i)^2}{3} \right\}, \tag{22}$$

where, as above, the noise to watermark ratio in the i-th coefficient, ξ_i, is kept fixed. The vector of distortion compensation parameters that minimizes (22) is given now by

$$\nu_i^* \triangleq \frac{1}{1+\xi_i},$$

for all $i = 1, \cdots, L$. Clearly, $\varphi(\boldsymbol{\nu}^*) \leq \varphi(\nu^*)$, since the first minimization is a particular case of the second constrained to a vector with equal components.

It is possible to regard the distortion compensation effect of the vector case as a Wiener filtering with matrix $\mathbf{A}^* \triangleq \mathrm{diag}(\boldsymbol{\nu}^*)$. This is so because all the self-noise elements corresponding to the components of $\boldsymbol{\nu}^*$ are mutually independent, what implies a diagonal filter. In fact, similar solutions have been proposed by Yu et al. in [18] from an information-theoretic point of view.

Finally, we would like to make some remarks. The performance improvement achieved by replacing ν with $\boldsymbol{\nu}$ is compatible with the gain due to using the decoding weights in (13). Whereas $\boldsymbol{\nu}$ modifies the pdfs independently at each dimension, we will see in the next section that $\boldsymbol{\beta}$ modifies the weighting of the dimensions when they are considered together. This fact will be duly shown in the next Section VI-B.

B. Derivation of the Improved Decoding Weights

We turn next our attention to the problem of optimizing the weights introduced in (13). Recall that the objective of this approach is to improve the performance of the minimum distance decoder using additional knowledge about the channel distortion eventually available at the decoder.

Adapting the method followed in Section III to the decoder in (13), it turns out that now P_e can be written as

$$P_e = \Pr\left\{ \sum_{i=1}^{L} \beta_i U_i^+ > \frac{1}{2} \sum_{i=1}^{L} \beta_i \right\},$$

which obviously reduces to (15) for $\boldsymbol{\beta} = \mathbf{1}$. Taking into account that any analytical optimization of the weights requires the availability of a closed-form approximation to P_e, we will discuss here the

minimization of (19) and (20) when weights are introduced. Starting with the CLT-based approximation, which we will see that it is very accurate for low values of WNR in Section VI, and under the same assumptions as in Section III-C, it is possible to write

$$P_e \approx P_{s_1} = \mathcal{Q}\left(\frac{\frac{1}{2}\sum_{i=1}^{L}\beta_i - \sum_{i=1}^{L}\beta_i \mathrm{E}\{U_i^+\}}{\sqrt{\sum_{i=1}^{L}\beta_i^2 \mathrm{Var}\{U_i^+\}}}\right). \tag{23}$$

Recalling that the $\mathcal{Q}(\cdot)$ function is monotonically decreasing, it follows that P_{s_1} is minimized when its argument is maximized. Then, the improved decoding weights can be found by differentiating the argument of $\mathcal{Q}(\cdot)$ in (23) with respect to $\beta_i, 1 \leq i \leq L$. Then, the decoding weights minimizing P_{s_1} are

$$\beta_i^* = K \cdot \frac{\left(\frac{1}{2} - \mathrm{E}\{U_i^+\}\right)}{\mathrm{Var}\{U_i^+\}},\ 1 \leq i \leq L, \tag{24}$$

where K is an irrelevant positive real constant, since the weights vector can be scaled without any impact on performance. Also, it is very interesting to note that some of the β_i^* may be negative. This will happen when $\mathrm{E}\{U_i^+\} > 1/2$, which may occur for large distortions. The effect of a negative weight can be interpreted as a swapping of the centroids assigned to each symbol.

As it can be inferred from (24), in order to compute the improved decoding weights, knowledge of $\mathrm{E}\{U_i^+\}$ and $\mathrm{Var}\{U_i^+\}$ is required. Note that due to the aliasing and truncation effects that show up in the construction of $\mathbf{U}^+$, this information is not directly derivable from the first and second order moments of the total noise random variable.

a) High WNR: As we will see in Section VI (Figure 6), the CLT-based approximation moves away from the empirical results as the WNR increases. In this case we can consider to use the union bound (20) to compute the improved decoding weights, since it is a better approximation to the P_e in the present scenario. Accordingly, the function that we have to minimize now is

$$P_e \approx P_{s_2} = 2^L \cdot \mathcal{Q}\left(\frac{\sum_{i=1}^{L}\beta_i}{2\sqrt{\sum_{i=1}^{L}\beta_i^2 \mathrm{Var}\{U_i\}}}\right), \tag{25}$$

which can be shown to be equivalent to the minimization of $\sum_{i=1}^{L}\beta_i^2 \mathrm{Var}\{U_i\}$ constrained to $\sum_{i=1}^{L}\beta_i = G$, for some arbitrary G. Applying Lagrange multipliers we may write the optimization functional as $\varphi(\boldsymbol{\beta}) = \sum_{i=1}^{L}\beta_i^2 \mathrm{Var}\{U_i\} - \lambda\left(\sum_{i=1}^{L}\beta_i - G\right)$. Differentiating it with respect to β_i and equating to zero it is straightforward to see that the minimum of (25) is obtained for

$$\beta_i^{**} = K\frac{1}{\mathrm{Var}\{U_i\}},$$

for $1 \leq i \leq L$ and any positive constant K. Interestingly, it is possible to show analytically that for large WNRs $\boldsymbol{\beta}^*$ will be nearly proportional to $\boldsymbol{\beta}^{**}$, which justifies the use of $\boldsymbol{\beta}^*$ also for large WNRs in spite of the looser approximation employed for its computation.

Notice that, after the optimal weights for the CLT-based approximations have been obtained, it is possible to resort to a more accurate computation of P_e (such as the Beaulieu's method or the DFT approach) by slightly modifying it to take the weights into account. The improvements afforded by $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^{**}$ will be empirically shown in Section VI-B.

C. A Geometric Interpretation of the Decoding Strategies

Here we provide a geometric interpretation of the various decoding strategies we have discussed, which will help to understand the role of the decoding weights and the goodness of Forney's approximation. For pictorial reasons, the case $L = 2$ is considered here. First of all, we derive the ML decision boundary based on Forney's approach when $\sigma^2_{T_i}$ is large. Noticing that from (9) the true ML lattice decoding boundary is the locus of the points $(u_1, u_2)^T$ for which $f_U(u_1, u_2) = f_U((u_1 - 1) \mod 2\mathbb{Z}, (u_2 - 1) \mod 2\mathbb{Z})$, and making use of the approximation in Section II-B.1, we can conclude that in the positive quadrant this boundary is approximately given by

$$\begin{aligned}\phi = \Big\{ & (u_1, u_2)^T \in [0,1] \times [0,1] : \\ & \Big(1 + 2\mathrm{e}^{-\pi^2 \sigma^2_{T_1}/2} \cos(\pi u_1)\Big) \cdot \Big(1 + 2\mathrm{e}^{-\pi^2 \sigma^2_{T_2}/2} \cos(\pi u_2)\Big) \\ = & \Big(1 - 2\mathrm{e}^{-\pi^2 \sigma^2_{T_1}/2} \cos(\pi u_1)\Big) \cdot \\ & \Big(1 - 2\mathrm{e}^{-\pi^2 \sigma^2_{T_2}/2} \cos(\pi u_2)\Big) \Big\}, \end{aligned} \tag{26}$$

with straightforward extensions to all other quadrants.

Figure 1 shows for the positive quadrant the true ML lattice decoder decision region for $\hat{b} = 0$ (shaded area) and the approximate decision boundary given by (26). The parameters of this plot are: $\sigma_{N_1}/\Delta_1 = 0.4113$, $\sigma_{N_2}/\Delta_2 = 0.2530$ and $\nu = 0.5$, so $\sigma_{T_1} = 0.5025$ and $\sigma_{T_2} = 0.3838$. As it can be perceived, Forney's approximation gives a very good estimate of the real boundary. Figure 1 also plots the decision boundaries that result using (13) with $\boldsymbol{\beta} = \mathbf{1}$, and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, which with the above parameters becomes $\beta^*_1 = 1.5936$ and $\beta^*_2 = 3.9005$. Observe how the use of β^*_i leads to a linear approximation of the true lattice ML decision boundary. Note however, that the ultimate purpose of the weights $\boldsymbol{\beta}^*$ is not to yield the best linear approximation of this boundary but to minimize an approximation of the bit error probability.

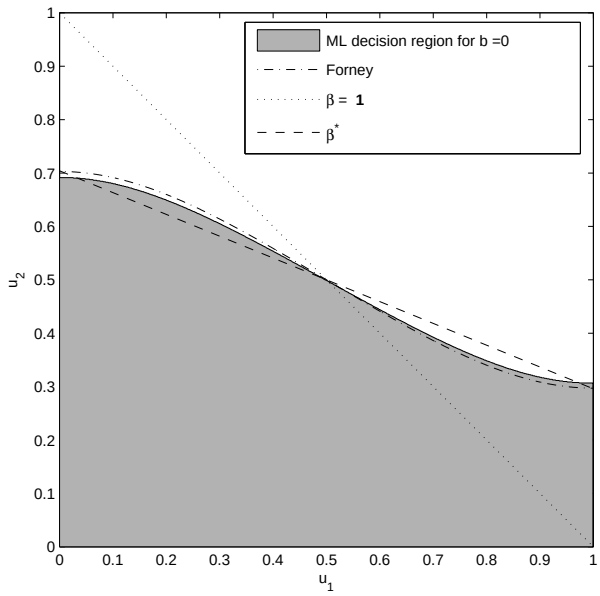


Fig. 1. Comparison of the decision regions for DC-DM ($L = 2$) obtained using Forney's approximation, the ML lattice decoder, and the Euclidean distance decoder with $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta} = \mathbf{1}$.

D. Discussion about the Pseudorandom Choice of the Partitions

Throughout this paper we have been assuming that the samples comprising the j-th host subvector $\mathbf{X}_j$ were pseudorandomly chosen. Starting from our CLT-based approximations, and using the law of large numbers, it is possible to theoretically justify the use of such pseudorandom assignment. Due to the lack of space, here we will only provide an empirical justification.

With this aim, we will consider the particular case of applying DC-DM watermarking to an image on the mid-frequencies of its 8×8-block DCT, the transform used in the JPEG standard. Moreover, we will let the channel noise variance be proportional to the squared JPEG quantization step (quality factor QF = 80) in each dimension, being this noise uniform. This quality factor is a scalar ranging from 0 (poor quality) to 100 (high quality) used by some implementations of the JPEG compression algorithm to indicate the quantization table. We have chosen this attack because it is assumed to have a perceptually-based power distribution (as JPEG quantization steps stem from perceptual considerations), although it does not follow the same power allocation as the watermark. We will consider two cases for defining the subvectors $\mathbf{X}_j$: global pseudorandom partitions (i.e., all available coefficients in the same pool), and frequency-dependent pseudorandom partitions (i.e., each pool consists of those coefficients with the same frequency indices that come from different blocks).

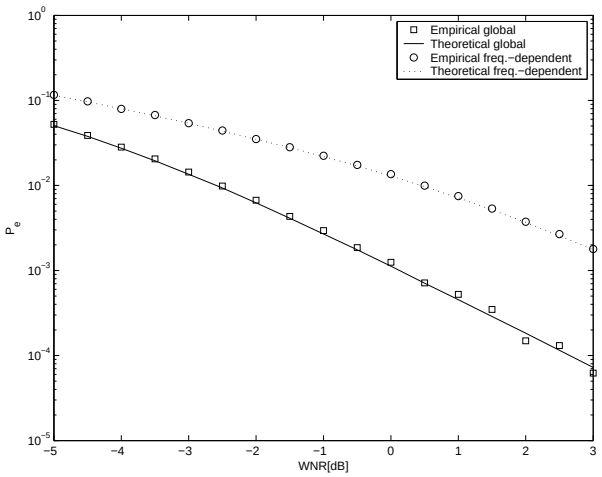


Fig. 2. Empirical and theoretical performance obtained with global vs. frequency-dependent pseudorandom partitions, using DC-DM on the DCT domain with optimally weighted Euclidean distance decoding. $L = 20$, $\nu = 0.4$, uniform noise, host image Lena 256×256, payload = 1126 bits.

This last strategy resembles the one used by Ramkumar and Akansu in [19] as well as the parallel channels studied by Moulin [20] applied to the DCT domain. In the former work, the data hiding capacity of compressed images is analyzed by decomposing an image into M subbands using transform blocks, thus giving rise to M parallel subchannels. Then, each symbol is only transmitted through a specific subchannel. With that strategy, all the coefficients devoted to conveying a certain symbol can be assumed to have the same noise statistics, differently to what happens when the indices are chosen pseudorandomly.

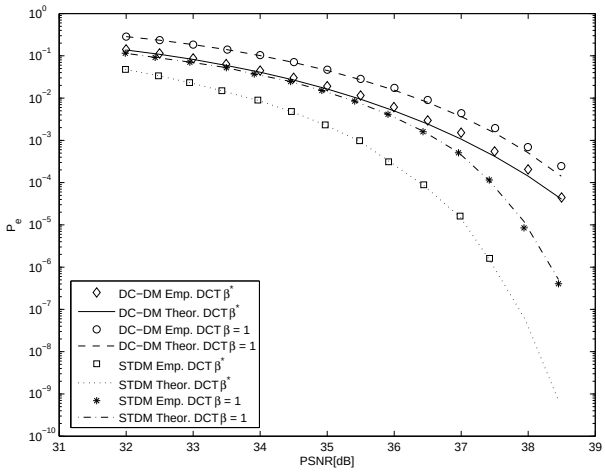
In Fig. 2 the improvement due to the use of global pseudorandom partitions is shown, choosing the mid-frequencies as in [21] and using the same perceptual mask and attack as in Section VI-B. The theoretical results were obtained using the DFT method. It is important however to note that a fixed subvector length has been assumed in this comparison, which clearly puts the frequency-dependent scheme at disadvantage, because each subchannel will have different host and noise statistics and, thus, different SNR's. A solution to this is to use subvector lengths that are also frequency-dependent, at the price of needing additional knowledge about the channel at both embedder and decoder, something that is not required when global partitions are used. Additionally, global pseudorandom partitions increase the entropy of the watermark and hence the security of the system.

E. Comparison with STDM

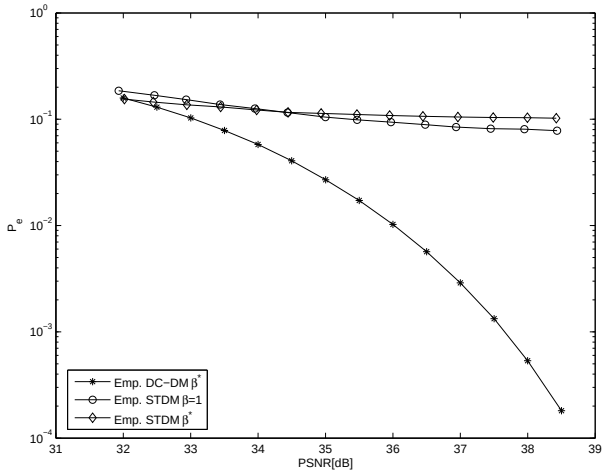
Although Spread Transform-Dither Modulation (STDM) [3], Spread Transform-Scalar Costa Scheme (ST-SCS) [4], and Quantized Projection (QP) [9] techniques do not constitute at all the main issue of this paper, a comparison with DC-DM with repetition coding is pertinent here for the sake of completeness. As shown in [4] and later confirmed by the authors in [9], STDM-like methods show superior performance than DC-DM in AWGN channels as the repetition L (and, equivalently, the spreading factor) increases. This is experimentally confirmed in Fig. 3(a) using real images as host data. The watermark is embedded in the mid-frequency coefficients of the 8×8 block-DCT domain [21] with a fixed Peak Signal to Noise Ratio (PSNR) of 40 dB, and uniform noise is added with standard deviation proportional to the corresponding JPEG quantization step in each dimension (quality factor QF=80). The figure also shows theoretical results, obtained using the CLT method in Section III-C for DC-DM and [9] for STDM. We observe a large gap between both methods for high PSNR's, but it is necessary to take into account that $\nu = 0.4$ used in the plot is not the optimal one when the PSNR of the attacked signal is close to 40 dB (large WNR). The optimal projection parameters $\boldsymbol{\beta}^*$ for STDM in Fig 3(a) are the ones derived in [22], even though other optimization strategies are available (see for instance [23]). A further advantage of STDM-like strategies, pointed out by an anonymous reviewer, is that they are quite independent of the particular statistical distribution of each sample, because in most circumstances the projected samples will look Gaussian (see [23], [9]). Moreover, it is more feasible to design attacks which render the attacked samples close to the decision boundary for DC-DM than for STDM.

This said, there are other simple attacks which can be much more detrimental for STDM-like methods than for DC-DM, as for instance cropping. Whereas for DC-DM the cropping attack is simply equivalent to decreasing the repetition factor L, which implies a smooth performance degradation, for STDM it can be seen as adding noise with variance equivalent to that of the removed samples. As this variance is usually much larger than the watermark variance, STDM performance is severely degraded by cropping[2]. The effect of cropping may be seen in Fig. 3(b), that compares the performance of both methods after removing an 8-pixels-wide outer frame. These results suggest that a combination of DC-DM and STDM is a good choice towards a truly robust moderate-rate scheme.

[2]The influence of the cropping in the BER could be reduced by taking block-wise partitions. Nevertheless, this would significantly reduce the security of the system.



(a)



(b)

Fig. 3. Performance of DC-DM ($\nu = 0.4$) vs. STDM, watermarking the DCT domain of real images; results averaged over twenty-two 256×256 images, with $L = 20$ (payload $= 1126$ bits). (a): Empirical and theoretical results with additive uniform noise. (b): Experimental results with additive uniform noise after cropping an external 8-pixels-wide frame.

F. Performance under Unforeseen Attacks

An interesting problem is posed by the performance analysis of DC-DM when the attack is different than the one expected by both the embedder and the decoder. We remind that the available information about the attack is exploited by them to compute, respectively, the optimal distortion compensation parameter and the optimal decoding weights. The general problem should be addressed from a game-theoretic approach, trying to find the optimal attack and the optimal encoding/decoding strategies, using

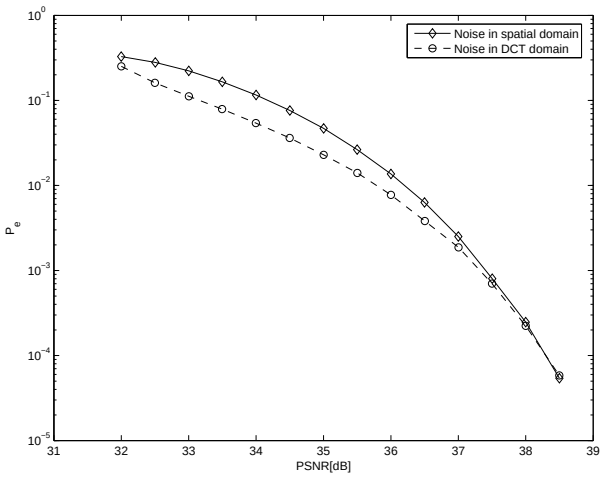


Fig. 4. Experimental performance of DC-DM in the spatial under uniform additive noise applied in the spatial and DCT domain, with $\nu = 0.4$, $L = 20$ and optimal decoding weight computed taking into account the noise in the spatial domain (payload $= 1126$ bits); results averaged over twenty-two 256×256 images.

a bit error rate payoff in our case. Unfortunately, we have not been able to obtain a solution due to the cumbersome expressions for the P_e.

In any case, it is interesting to observe the performance degradation when there is a mismatch between the actual attack and the one considered when optimizing the method. In Fig. 4 experimental results for this case are shown for a particular case in which DC-DM is applied to the Lena image in the spatial domain, and the embedder and decoder expect uniform noise in the same domain. However, the noise is added in the DCT domain in both cases. In order to set realistic conditions, the uniform noise in the DCT domain has, at each coefficient, variances proportional to a squared perceptual mask computed following Watson [24]. Although it can be verified that the energy distribution of the corresponding inverse transformed noise in the spatial domain differs considerably from the spatial perceptual mask, we may see that there is only a small performance difference (in fact a gain) with respect to the ideal case where the noise follows the expected distribution.

V. DC-DM Performance under Coarse Quantization

In this section we will analyze the performance of DC-DM when the watermarked signal $\mathbf{Y}$ undergoes coarse quantization, which is quite a common unintentional attack. Notice that we cannot deal with this particular attack using the generic methods presented up to this point, as in this case we cannot assume

the independence of the channel noise (actually the coarse quantization error). Furthermore, our analysis will serve to show how to improve the performance of DC-DM under this particular attack.

We assume next that a coarse quantizer with centroids given by the lattice $\delta_i\mathbb{Z}$ is applied to y_i for all $1 \leq i \leq L$. The computation of the probability of decoding error relies on knowing the probability mass function (pmf) of Z_i. Notice that this pmf will not only depend on the pdf of the host image, but also on that of the watermark, which in turn depends on the transmitted bit b and on the dither d_i. In order to obtain the desired probability we need the upper and lower limits of the k-th coarse-quantization bin, which will be denoted by $\theta_{i_k}^+ \triangleq k\delta_i + \delta_i/2$ and $\theta_{i_k}^- \triangleq k\delta_i - \delta_i/2$, respectively. So, the probability that Z_i is equal to the k-th coarse-quantization centroid conditioned to the transmission of b is

$$\begin{aligned} \Pr\{Z_i = k\delta_i \mid b\} &= \Pr\{Y_i \in (\theta_{i_k}^-, \theta_{i_k}^+] \mid b\} \\ &= \int_{\theta_{i_k}^-}^{\theta_{i_k}^+} f_{Y_i}(y_i|b)dy_i. \end{aligned} \tag{27}$$

We are interested in reformulating this integral in terms of X_i, what requires a change of variable affecting the integration limits of the expression. This change of variable is not evident, but it can be obtained in a straightforward manner. First, notice that the DC-DM centroid corresponding to the symbol b and closest to the upper limit $\theta_{i_k}^+$ of the integral (27) is just $Q_b(\theta_{i_k}^+)$, with $Q_b(\cdot)$ defined in (3). Then, considering the offset $\rho_y(\theta_{i_k}^+, b) \triangleq \theta_{i_k}^+ - Q_b(\theta_{i_k}^+)$, it can be shown that the corresponding offset with respect to $Q_b(\theta_{i_k}^+)$ from the point of view of X_i is $\rho_x(\theta_{i_k}^+, b) \triangleq \frac{\min\{\max[\rho_y(\theta_{i_k}^+, b), -(1-\nu)\Delta_i], (1-\nu)\Delta_i\}}{(1-\nu)}$. Therefore, the upper limit when the integral in (27) is evaluated using $f_{X_i}(x_i)$ is just $\gamma_{i_k}^+(b) \triangleq Q_b(\theta_{i_k}^+) + \rho_x(\theta_{i_k}^+, b)$. The lower limit $\gamma_{i_k}^-$ can be obtained similarly, and then the desired probability can be put as

$$\Pr\{Z_i = k\delta_i \mid b\} = \int_{\gamma_{i_k}^-(b)}^{\gamma_{i_k}^+(b)} f_{X_i}(x_i)dx_i. \tag{28}$$

This pmf plays a similar role as the pdf $f_{T_i}(\cdot)$ in (8). Hence, the probability of decoding error under coarse quantization can be obtained by applying to this pmf the same modular strategy used in Section III. Unfortunately, the resulting expression is quite involved and it has to be computed numerically in practice.

Notice that the probability of error thus obtained will be in general dependent on b. A side-effect of this dependence is that the weights optimization in Section IV-B is not valid for coarse quantization in general. Actually, the improved decoding weights β_i^* will only be valid for symmetric settings. In section VI-D we will compare the performance under coarse quantization using two kinds of dithers. For the first one we choose $d_i \in \{\pm\Delta_i/2\}$, for all $i = 1, \cdots, L$. Due to symmetry, in this case the statistics for each dimension are independent of the embedded bit, and the procedure to compute the decoding weights can still be used. For the second one, $d_i \in \{0, \Delta_i\}$ for all $i = 1, \cdots, L$, which does not give

a symmetric setting. With this choice, the statistics in each dimension do depend on the embedded bit, thus making it impossible to derive the aforementioned weights.

1) JPEG Compression: We may particularize the expression (28) for a real coarse quantization case such as the one induced by the popular JPEG standard for image compression. Accordingly, let us assume throughout this subsection that the host signal is given in the 8×8 block-DCT domain where JPEG works. As discussed in [21] the AC coefficients of the DCT can be reasonably modeled by zero-mean generalized Gaussian pdfs, given by the expression

$$f_X(x) = A\mathrm{e}^{-|\eta x|^c}. \tag{29}$$

The parameters A and η can be expressed as a function of the shape parameter c and the standard deviation σ_X. We refer the reader to [21] for the details on how to tackle in practice the issue of their estimation. Taking into account the model (29), and assuming that its parameters are estimated adaptively for each dimension, we may rewrite (28) as $\Pr\{Z_i = k\delta_i \mid b\} = \Pr\{X_i \leq \gamma_{i_k}^+(b)\} - \Pr\{X_i \leq \gamma_{i_k}^-(b)\}$, with

$$\Pr(X_i \leq \tau) = \begin{cases} \frac{A_i}{\eta_i c_i}\Gamma(1/c_i, |\eta_i\tau|^{c_i}), & \text{if } \tau \leq 0 \\ 1 - \frac{A_i}{\eta_i c_i}\Gamma(1/c_i, |\eta_i\tau|^{c_i}), & \text{if } \tau > 0 \end{cases},$$

where $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function[3].

VI. Empirical Results

In this section we will check the validity of our theoretical developments, comparing the analytical results with empirical ones. First, we retake the discussion in Section II about the optimal channel coding for DC-DM with uniform scalar quantizers. In Figure 5 the performance of two coding settings using DC-DM is depicted. The concatenation of DC-DM with repetition coding ($L = 6$) with a simple outer turbo code rate $1/3$ —to yield an overall rate $1/18$— is compared to a rate $1/17$ turbo code over DC-DM. We observe that, for these similar rates, the concatenation only loses about 1 dB with respect to a turbo code with a much more complex decoding. Notice that the channel model used to decode the turbo code concatenated with repetition is the CLT approximation described in Section III-C, which is detrimental for the concatenation for such a low L. This plot is in agreement with the results shown in [4] for this type of concatenation, and it supports the practical utility of repetition coding for DC-DM.

[3] $\Gamma(a,z) \triangleq \int_z^\infty t^{a-1}\mathrm{e}^{-t}dt$.

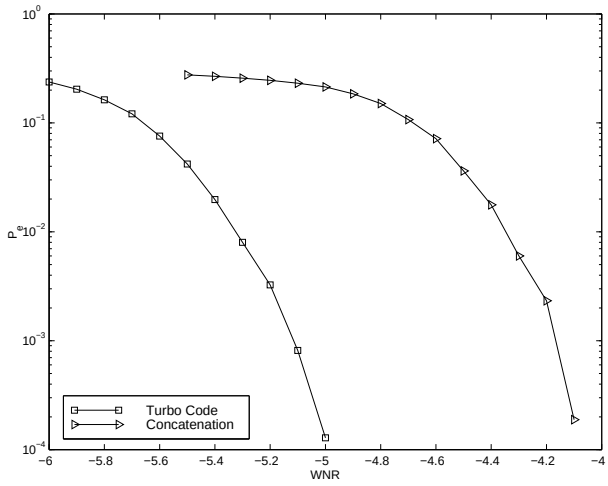


Fig. 5. Performance comparison of DC-DM with a turbo code $R = 1/17$ ($\nu = 0.30$) vs. DC-DM with repetition $L = 6$ concatenated with a $R = 1/3$ turbo code ($\nu = 0.35$), interleaver size 1000 symbols. Synthetic host data.

A. Comparison of the Approximations and Bounds

Figure 6 shows the approximations and bounds in Section III versus the outcomes of i.i.d. Montecarlo simulations. In this plot channel noise is additive zero-mean Gaussian, the components of $\mathbf{X}$ and $\mathbf{N}$ are i.i.d., $L = 10$ and ν is optimized following Costa's formula, i.e., $\nu = \nu_c \triangleq D_w/(D_w + D_c)$. We may verify that the accuracy of the approximations given in Sections III-A and III-B is remarkable. The CLT-based approximation is excellent for low values of the WNR, but, as the WNR is increased, it gets away from the true probability of error. As it was explained in Section III-C, this is due to the support of $f_{U_i^+}(u_i^+)$ being only positive, to the small value of L used in the experiment, and to the increase in the skew-effect of the resulting pdf for large values of the WNR. Since this approach underrates the importance of the tails of $f_R(r)$, the approximation produces overly optimistic results.

On the other hand, the union bound gets closer to the empirical results when the WNR increases. This is a consequence of the reduction of the probability corresponding to the overlapped decision regions when the WNR grows. We also plotted the results of applying the CLT to compute the probability of error with only one neighbor and then using the union bound, as described in Section III-D.2. In this case the pdf involved in the computation is symmetric about the origin, so convergence to the Gaussian distribution is unaffected when the WNR is increased. Note that both bounds approach the true probability of bit error asymptotically as the WNR increases. The values predicted by the approximation of Chen and Wornell are obviously parallel to those obtained when both the union bound and the CLT are used

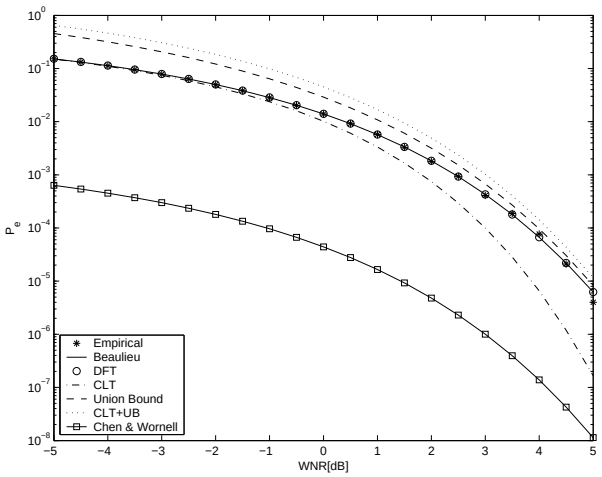


Fig. 6. Comparison of the empirical BER vs. the different analytical and numerical approximations and bounds for DC-DM under Gaussian noise. $L = 10$, $\nu = \nu_c$. Synthetic host data.

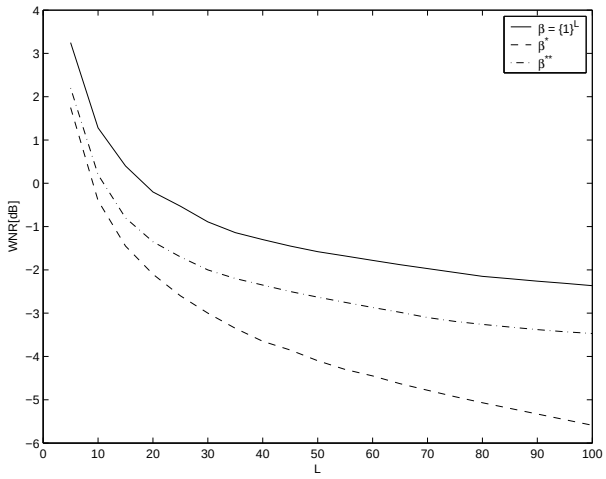
(see Section III-D.2). As it should be expected, those values are clearly lower than the empirical results, since only the probability of mistaking two neighbors is taken into account.

Finally, the bound by Erez and Zamir is not shown in Figure 6 because its value is around 10^3. It is pertinent to remark here that even though this bound is valid for any pair of nested lattices, it was designed to show the capacity-achieving property of lattice decoding. Nevertheless, for that purpose, it is necessary that the pair of nested lattices verify certain properties which fall short of being true for the lattices used by DC-DM. This explains why such large values arise and demonstrates how information-theoretic results cannot always be effortlessly extrapolated to practical schemes.

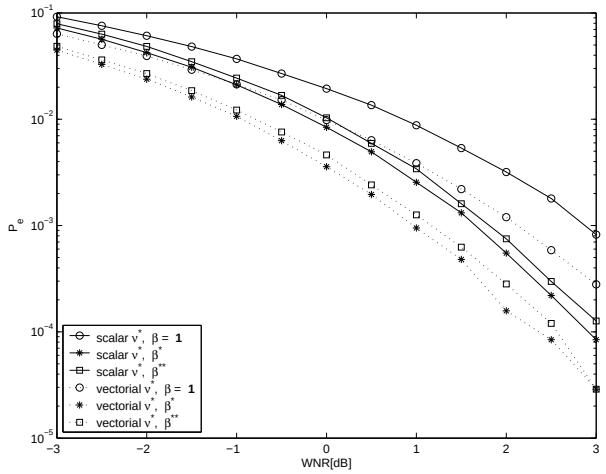
B. Optimized Distortion Compensation Parameter and Improved Decoding Weights

The next set of experiments were carried out by watermarking the image *Lena* 256×256 in the DCT domain, using a perceptual mask proportional to the perceptual thresholds proposed in [24] and [25]. The attack is uniformly distributed with amplitude proportional in each dimension to the corresponding JPEG quantization step for QF=80.

Figure 7(a) shows the performance improvements due to the use of the weights $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^{**}$ in the Euclidean distance decoder. The plot depicts the WNR needed to achieve $P_e = 0.01$ with L ranging from 5 to 100, and clearly shows the improvement obtained when $\boldsymbol{\beta}^*$ is used. The performance gain is already large at $L = 100$, but the gap keeps increasing with L. Nevertheless, the improvement is not so large



(a)



(b)

Fig. 7. DC-DM watermarking of the Lena 256×256 host signal in the DCT domain with uniformly distributed additive attack. (a): WNR needed to achieve $P_e = 0.01$ vs. L for $\nu = 0.7$, with different weightings on the Euclidean distance decoder. (b): Comparison of the empirical BER obtained when ν^* or $\boldsymbol{\nu}^*$ are used in conjunction with $\boldsymbol{\beta} = \mathbf{1}$, $\boldsymbol{\beta}^*$, and $\boldsymbol{\beta}^{**}$. $L = 10$, payload $= 2252$ bits.

when $\boldsymbol{\beta}^{**}$ is used. In order to explain this effect, consider that the WNR's studied are rather negative, and therefore that the CLT-based approximation used for the computation of $\boldsymbol{\beta}^*$ is clearly better than the union bound plus CLT expression used for the computation of $\boldsymbol{\beta}^{**}$ (see Section III-D.2 and cf. Figure 6).

Figure 7(b) shows the results obtained when ν^* and $\boldsymbol{\nu}^*$ are used in conjunction with $\boldsymbol{\beta} = \mathbf{1}$ (i.e., no weighting), $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^{**}$, for the case $L = 10$. A considerable gain is achieved by using a vectorial distortion compensation parameter $\boldsymbol{\nu}^*$ instead of a scalar one, ν^*. The improvement due to using $\boldsymbol{\beta}^*$ and

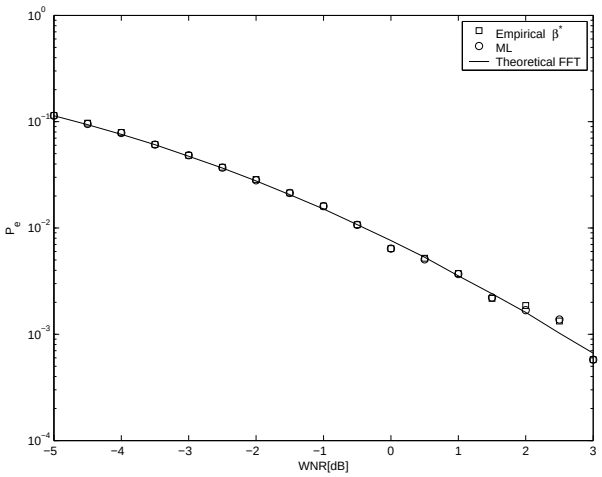


Fig. 8. Performance of DC-DM using ML lattice decoding vs. using Euclidean distance decoding using the optimal weights $\boldsymbol{\beta}^*$. Gaussian noise with variance proportional to the squared JPEG quantization step (QF = 80), $\nu = 0.5$, $L = 10$, payload $= 2252$ bits, host image Lena 256×256.

$\boldsymbol{\beta}^{**}$ compared to no weighting is also apparent. Note also that the weighting strategy $\boldsymbol{\beta}^*$ yields the best results for the whole range considered in this case. Finally, as we have pointed out in Section IV-A, the use of a distortion compensation vector is compatible with the improved decoding weights, so the combination offers improvements of about 2 dB over the standard embedding/decoding strategy.

In Figure 8 we compare ML lattice decoding versus Euclidean distance decoding weighted by $\boldsymbol{\beta}^*$. The theoretical results for $\boldsymbol{\beta}^*$ in that figure were computed employing the DFT method. This plot clearly shows the near-optimality of performing Euclidean distance decoding with our optimal weighting strategy, since the results obtained are virtually the same than those obtained with ML lattice decoding. This result can be explained (at least for small values of WNR, where Forney's approximation is valid) in view of the resemblance between the decision regions used by these two decoders (see Section IV-C). It is interesting to remark that, as the variance of the host signal is much larger than that of the watermark, adjacent DC-DM centroids have similar probabilities, and then ML lattice decoding approaches ML decoding.

C. Comparison with Miller et al.'s Trellis-based Embedding

We compare next DC-DM to the side-informed algorithm based on trellis quantization presented in [26]. In order to undertake the comparison, we encoded DC-DM using the cascade of an outer code, given by two serially-concatenated codes [27] with global rate $1/4$, with an inner $1/3$ repetition code, obtaining the same overall coding rate $1/12$ used in [26]. The use of channel coding is necessary in order to make

a fair comparison, since the method in [26] inherently includes an involved (source) code. Admittedly, the comparison will be dependent on the particular codes used in each case, but we may get in this way an acceptable perspective of the relative performance of both methods.

In order to set the same test conditions, DC-DM embedding is performed with the same image and using the same DCT coefficients as in [26], and hence the payload is also 1380 bits. Similarly, the same Watson-based perceptual constraints [24] are taken into account, and the Watson measure due to the DC-DM watermark is fixed to 27.20 as in [26]. Our experiments show that $P_e \approx 10^{-3}$ for DC-DM when the standard deviation of the additive noise is 8.5, marking the region of the turbo-cliff in the iteratively decoded DC-DM scheme. For the same noise power, Miller et al.'s method yields $P_e \approx 3.3 \times 10^{-3}$. Thus, both techniques exhibit similar performance under this very specific scenario. Testing under other circumstances is left open for future research.

D. Coarse Quantization: JPEG Compression

We compare next in Fig. 9 the performance of DC-DM under the coarse quantization attack given by JPEG compression, using the symmetric and asymmetric dithers discussed in Section V. In the plot, the probability of bit error is plotted versus the quality factor QF used to compress the watermarked signal *Lena* using JPEG. Embedding takes place in the 8×8 block-DCT domain. In order to obtain the theoretical results we have used the CLT-based approximation and assumed a Laplacian distribution for the host signal, which corresponds to $c = 1$ in (29). This approximation explains the small discrepancies between the theoretical and empirical results, which are more evident for $\boldsymbol{\beta}^*$ as the convergence of the decision statistic to a Gaussian is slower with weighting. As it can be seen, the use of an asymmetric dither yields superior performance, even considering that it is not possible to use the optimal weights in this case.

VII. CONCLUSIONS

Quantization-based methods have opened the gate to high-rate data-hiding and watermarking applications. Distortion-Compensated Dither Modulation with uniform scalar quantizers and repetition coding is probably the simplest algorithm for robust informed embedding. For this reason, it is likely to become an increasingly popular method in the near future. Moreover, as we have shown, repetition coding is a reasonable choice for concatenation with more powerful coding schemes, such as turbo codes.

This paper comes to fill an existing gap in the theory of DC-DM with repetition coding. Firstly, we have shown how to modify the basic method to take into account perceptual shaping. Secondly, we

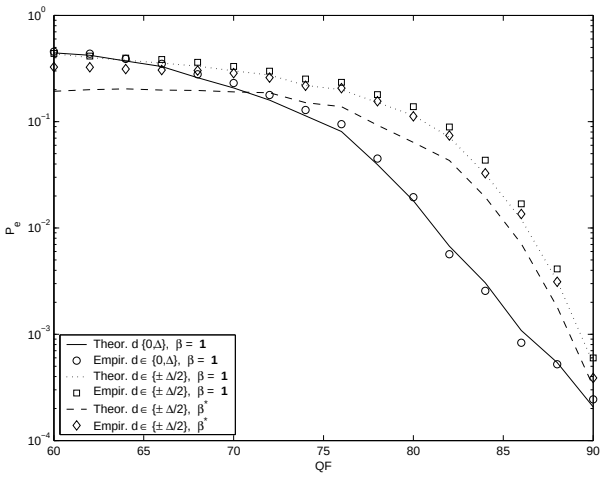


Fig. 9. Empirical and theoretical performance when *Lena* is watermarked with DC-DM in the DCT domain and JPEG-compressed with quality factor QF, for symmetric and asymmetric dithers. $L = 10$, payload $= 2252$ bits, $\nu = 0.5$.

have provided a complete discussion of the several decoding alternatives, including ML lattice decoding and Euclidean distance decoding. A thorough analysis of the bit error rate has been also presented. This analysis, based on the key idea of projecting all the random variables onto the positive orthant, includes two procedures for the *exact* computation of the BER as well as several approximations and bounds with theoretical value. Furthermore, we have proposed two important enhancements: the use of decoding weights, which can be approximately optimized thanks to our novel theoretical analysis, and the application of a vectorial distortion compensation parameter. Together, they produce significant improvements as it has been shown both analytically and empirically. A comparison with STDM-like methods revealed that, even though the latter perform better under additive noise, DC-DM is much more robust against cropping attacks. Finally, we have extended our methodology to the case of coarse-quantization attacks such as JPEG, and discussed the advantages brought about by asymmetric dither vectors. Results averaging error probabilities over real images have been reported as well, showing the accuracy of our theoretical analyses. A comparison with Miller et al.'s trellis-based embedding has been carried out, evidencing that DC-DM with repetition and an outer turbo code achieves similar performance.

Some of the proposals made in this paper, such as the employment of decoding weights and varying distortion compensation parameters, can be easily extended to other quantization-based methods with perceptual constraints. Nevertheless, in order to avoid costly numerical optimizations, it is of paramount importance to have good analytical approximations to the desired performance measures. We expect that

the guidelines here proposed be extended to more sophisticated lattices and/or coding schemes.

APPENDIX

We derive next the characteristic function $F_{U_i^+}(u_i^+)$ required for computing P_e in front of Gaussian noise following Beaulieu's method. Let $\sigma_{G_i} \triangleq \frac{\sigma_{N_i}}{\Delta_i}$ be the standard deviation of the Gaussian attack after the normalization by Δ_i. Taking into account (7), (8) and (16), the pdf of U_i^+ can be written as

$$f_{U_i^+}(u_i^+) = \begin{cases} \sum_{k=-\infty}^{\infty} \frac{1}{\mu_i} \left[\mathcal{Q}\left(\frac{u_i^+ - (1-\nu) - 2k}{\sigma_{G_i}}\right) \right. \\ \quad \left. - \mathcal{Q}\left(\frac{u_i^+ + (1-\nu) - 2k}{\sigma_{G_i}}\right) \right], \quad \text{if } 0 \le u_i^+ \le 1, \ , \\ 0, \quad \text{otherwise,} \end{cases}$$

with $\mu_i \triangleq \frac{1-\nu}{\sigma_{G_i}}$. For the sake of simplicity we define $M_i \triangleq \frac{U_i^+}{\sigma_{G_i}}$, whose characteristic function is

$$\begin{aligned} F_{M_i}(\omega) &= \int_0^{\delta_i} e^{j\omega m_i} \sum_{k=-\infty}^{\infty} \frac{1}{\mu_i} \left[\mathcal{Q}\left(m_i - \mu_i - 2k\delta_i\right) \right. \\ &- \left. \mathcal{Q}\left(m_i + \mu_i - 2k\delta_i\right) \right] dm_i = \\ &= \frac{j}{2\mu_i \omega} \left\{ - \operatorname{erf}\left(\frac{-\mu_i - 2k\delta_i}{\sqrt{2}}\right) \right. \\ &+ \operatorname{erf}\left(\frac{\mu_i - 2k\delta_i}{\sqrt{2}}\right) \\ &+ e^{j\delta_i\omega} \left[\operatorname{erf}\left(\frac{\delta_i - 2k\delta_i - \mu_i}{\sqrt{2}}\right) \right. \\ &- \left. \operatorname{erf}\left(\frac{\delta_i - 2k\delta_i + \mu_i}{\sqrt{2}}\right) \right] \\ &+ e^{-\omega\left(\frac{\omega}{2} - j(\mu_i + 2k\delta_i)\right)} \left[- \operatorname{erf}\left(\frac{\mu_i + 2k\delta_i + j\omega}{\sqrt{2}}\right) \right. \\ &+ \left. \operatorname{erf}\left(\frac{-\delta_i + \mu_i + 2k\delta_i + j\omega}{\sqrt{2}}\right) \right] \\ &+ e^{-\omega\left(\frac{\omega}{2} - j(-\mu_i + 2k\delta_i)\right)} \left[- \operatorname{erf}\left(\frac{-\mu_i + 2k\delta_i + j\omega}{\sqrt{2}}\right) \right. \\ &+ \left. \left. \operatorname{erf}\left(\frac{-\delta_i - \mu_i + 2k\delta_i + j\omega}{\sqrt{2}}\right) \right] \right\} \end{aligned} \tag{30}$$

with $\delta_i \triangleq \frac{1}{\sigma_{G_i}}$. It is straightforward to see that $F_{U_i^+}(\omega) = F_{M_i}(\omega \cdot \sigma_{G_i})$. The $\operatorname{erf}(\cdot)$ function is defined as

$$\begin{aligned} \operatorname{erf}(z) &= \frac{2}{\sqrt{\pi}} \int_0^z \mathrm{e}^{-t^2/2} dt = \frac{2z}{\sqrt{\pi}} M\left(\frac{1}{2}, \frac{3}{2}, -z^2\right), \\ &\quad \text{with } z \in \mathbb{C}, \end{aligned} \tag{31}$$

with $M(\cdot,\cdot,\cdot)$ the Kummer confluent hypergeometric function of the first kind. The evaluation of (30) presents numerical problems due to the evaluation of (31), which is computed as

$$\begin{aligned}\mathrm{erf}(x+iy) &\approx \mathrm{erf}(x)+\frac{\mathrm{e}^{-x^2}}{2\pi x}[(1-\cos(2xy)+i\sin(2xy)] \\ &+ \frac{2}{\pi}\mathrm{e}^{-x^2}\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-n^2/4}}{n^2+4x^2}[f_n(x,y)+ig_n(x,y)],\end{aligned}$$

where

$$\begin{aligned}f_n(x,y) &= 2x-2x\cosh(ny)\cos(2xy)+n\sinh(ny)\sin(2xy), \\ g_n(x,y) &= 2x\cosh(ny)\sin(2xy)+n\sinh(ny)\cos(2xy).\end{aligned}$$

REFERENCES

[1] B. Chen and G. W. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking," in *Proc. of SPIE*, ser. Security and Watermarking of Multimedia Contents II, San José, USA, January 2000, pp. 48–59.

[2] M. H. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[3] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[4] J. Eggers and B. Girod, *Informed Watermarking*. Kluwer Academic Publishers, 2002.

[5] A. Levy and N. Merhav, "An image watermarking scheme based on information theoretic principles," HP Labs, Tech. Rep., 2001, available at *http://www.hpl.hp.com/techreports/2001/HPL-2001-13.html*.

[6] M. Ramkumar and A. N. Akansu, "Signaling methods for multimedia steganography," *IEEE Trans. on Signal Processing*, vol. 52, no. 4, pp. 1100–1111, April 2004.

[7] A. Piva, F. Bartolini, I. Coppini, A. D. Rosa, and E. Tamurini, "Analysis of data hiding technologies for medical images," in *Proc. of SPIE*, ser. Security and Watermarking of Multimedia Contents V, Santa Clara, USA, January 2003, pp. 379–390.

[8] J. Chou, K. Ramchandran, and A. Ortega, "Next generation techniques for robust and imperceptible audio data hiding," in *IEEE Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3. IEEE, May 2001, pp. 1349–1352.

[9] F. Pérez-González, F. Balado, and J. R. Hernández, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 960–980, April 2003, special Issue "Signal Processing for Data Hiding in Digital Media & Secure Content Delivery".

[10] F. Pérez-González and F. Balado, "Nothing but a kiss: A novel and accurate approach to assessing the performance of multidimensional distortion-compensated dither modulation," in *Proc. of the 5th International Workshop on Information Hiding*, ser. Lecture Notes in Computer Science. Noorwijkerhout, The Netherlands: Springer-Verlag, October 2002.

[11] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed., ser. Comprehensive Studies in Mathematics. Springer, 1999, vol. 290.

[12] M. Barni, F. Bartolini, and A. D. Rosa, "Advantages and drawbacks of multiplicative spread spectrum watermarking," in *Proc. of SPIE*, ser. Security and Watermarking of Multimedia Contents V, Santa Clara, USA, January 2003, pp. 290–299.

[13] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\text{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. on Information Theory*, vol. 50, no. 10, pp. 2293–2314, October 2004.

[14] G. D. Forney, M. D. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. on Information Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[15] N. C. Beaulieu, "An infinite series for the computation of the complementary probability distribution function of a sum of independent random variables and its application to the sum of Rayleigh random variables," *IEEE Trans. Commun.*, vol. 38, no. 9, pp. 1463–1474, September 1990.

[16] M. Barni, F. Bartolini, and A. Piva, "Performance analysis of ST-DM watermarking in presence of non-additive attacks," *IEEE Trans. on Signal Processing*, vol. 52, no. 10, pp. 2965–2974, October 2004.

[17] J. G. David Forney, "On the role of MMSE estimation in approaching the information-theoretic limits of linear gaussian channels: Shannon meets Wiener." 41st Annual Allerton Conference on Communications, Control, and Computing, October 2003.

[18] W. Yu, A. Sutivong, D. Julian, T. M. Cover, and M. Chiang, "Writing on colored paper," available at *http://www.comm.toronto.edu/ weiyu/publications.html*.

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

[20] P. Moulin and M. K. Mihçak, "The parallel-gaussian watermarking game," *IEEE Trans. on Information Theory*, vol. 50, no. 2, pp. 272–289, February 2004.

[21] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Trans. on Image Processing*, vol. 9, no. 1, pp. 55–68, January 2000, special Issue on Image and Video Processing for Digital Libraries. [Online]. Available: http://www.ieee.org/organizations/pubs/pub_preview/IP/09ip01_toc.html

[22] P. Comesaña, F. Pérez-González, and F. Balado, "Optimal strategies for spread-spectrum and quantized-projection image data hiding games with BER payoffs," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*, vol. 3, September 2003, pp. 479–482.

[23] G. L. Guelvouit, S. Pateux, and C. Guillemot, "Perceptual watermarking of non i.i.d. signals based on wide spread spectrum using side information," in *Proc. of the IEEE International Conference on Image Processing (ICIP)*, vol. 3, June 2002, pp. 477–480.

[24] A. B. Watson, "DCT quantization matrices visually optimized for individual images," in *Proc. of SPIE*, ser. Human Vision, Visual Processing and Digital Display IV, 1993, pp. 202–216.

[25] A. J. Ahumada Jr. and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Proc. of SPIE*, ser. Human Vision, Visual Processing and Digital Display III, 1992, pp. 365–374.

[26] M. L. Miller, G. J. Doërr, and I. J. Cox, "Applying informed coding and embedding to design a robust high-capacity watermarking," *IEEE Trans. on Image Processing*, vol. 13, no. 6, pp. 792–807, June 2004.

[27] S. Benedetto, D. Divsalar, G. Montorsi, and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *IEEE Trans. on Information Theory*, vol. 44, no. 3, pp. 909–926, May 1998.