Performance Analysis of Existing and New Methods for Data Hiding With Known-Host Information in Additive Channels

Fernando Pérez-González, *Member, IEEE*, Félix Balado, and Juan R. Hernández Martin, *Member, IEEE*

***Abstract*—A considerable amount of attention has been paid lately to a number of data hiding methods based in quantization, seeking to achieve in practice the results predicted by Costa for a channel with side information at the encoder. With the objective of filling a gap in the literature, this paper supplies a fair comparison between significant representatives of both this family of methods and the former spread-spectrum approaches that make use of near-optimal ML decoding; the comparison is based on measuring their probabilities of decoding error in the presence of channel distortions. Accurate analytical expressions and tight bounds for the probability of decoding error are given and validated by means of Monte Carlo simulations. For dithered modulation (DM), a novel technique that allows us to obtain tighter bounds to the probability of error is presented. Within the new framework, the strong points and weaknesses of both methods are distinctly displayed. This comparative study allows us to propose a new technique named "quantized projection" (QP), which, by adequately combining elements of those previous approaches, produces gains in performance.**

***Index Terms*—Data hiding, dither modulation, quantization index modulation, quantized projection, spread-spectrum.**

I. Introduction

DATA hiding is the generic name given to a number of techniques having the common characteristic of inserting a certain set of data into a regular signal without noticeably modifying it. The growing amount of exchanged information since the expansion of the Internet during the last decade has motivated active and prolific research in this field. It has also focused the attention of the investigations on multimedia signals (i.e., image, video and audio) as typical host signals for carrying the hidden data. The kind of multimedia applications of data hiding ranges from steganography, where the sheer act of the embedding tries to be concealed, to copyright enforcement and/or fingerprinting, where the hidden information has to reliably identify the host signal and/or its legal owner even under tampering efforts. The requirements for these applications also vary considerably. One of the most elusive of these requirements has been, since the first attempts, robustness, i.e., the ability to survive intentional or unintentional attacks aiming at removing or modifying the embedded payload.

Although a young discipline, digital data hiding has already gone through two different phases. During the first one, algorithms flooded the literature, in most cases, with weak theoretical grounds and with little concern about robustness or performance [1], [2]. The second phase started with the recognition that the data hiding problem was in fact a particular case of communications [3], clearing the way for using well-known techniques, such as spread-spectrum, and facilitating subsequent performance analyses. In addition, researchers started to look at data hiding under the light of Information Theory, thus helping to establish fundamental limits on performance. Within this framework the problem of embedding can be assimilated to optimally encoding the information to be hidden by shaping the host signal under certain perceptual restrictions; next, this signal goes through a statistical channel representing the eventual attacks before arriving to the decoder, which tries to recover the embedded information from the channel output as reliably as possible.

A third and current phase started when Cox *et al.* [4] identified that the data hiding problem could in addition be seen as one of communications with side information at the encoder. Soon afterwards, an important result by Costa [5] was rescued by Chen and Wornell in [6]. In his paper, Costa gave a solution for obtaining null host signal interference for a known realization of a Gaussian host signal and a random Gaussian channel. Then, the problem with side information only at the encoder—usually termed as "blind" data hiding—could be seen as equivalent to having full side information at the decoder. A number of down-to-earth implementations aiming at practically approaching Costa's construction have sprung up since then [6]–[8], consisting, in most cases, of quantization procedures. This family of approaches, that from now on will be referred to as *known-host-state methods*, has been claimed to improve the performance of those usually called "spread spectrum" [3], [9], which use different forms of diversity. In principle, the latter do not make use of host signal state information, although they are not completely unaware of it as most of them utilize the host state for determining the maximum allowed perceptual distortion. Moreover, in their best performing versions, they employ statistical information about the host signal to construct optimal or near-optimal detectors. This usage of host signal statistics

Manuscript received February 4, 2002; revised November 21, 2002. This work was supported in part by the *Xunta de Galicia* under Projects PGIDT01 PX132204PM and PGIDT02 PXIC32205PN, the European project Certimark (Certification of Watermarking Technologies), IST-1999-10987, and the CYCIT project AMULET under Reference TIC2001-3697-C03-01. The associate editor coordinating the review of this paper and approving it for publication was Dr. Ali N. Akansu.

F. Pérez-González and F. Balado are with the Departamento Teoría de la Señal y Comunicaciones, ETSI Telecomunicacion, Universidad de Vigo, 36200 Vigo, Spain (e-mail: {fperez@tsc.uvigo.es; fiz@tsc.uvigo.es).

J. R. Hernández Martín is with Nagravision S. A., Lausanne, Switzerland (e-mail: juan-ramon.hernandez@nagra.com).

Digital Object Identifier 10.1109/TSP.2003.809368

in their detection algorithms permits us to denominate them *known-host-statistics methods*.

The first objective of this paper is to provide a fair and rigorous comparison of the performance yielded by some of the most important representatives of these two families of methods, which is lacking in the literature. Some previous analyses rely on questionable assumptions and/or simplifications, which we will try to overcome. A further motivation for revisiting some recently proposed methods leans on the fact that Costa's scheme is only optimal for a certain class of channels and host signals. In fact, we will show that as frequently happens with real-world communications systems, channel capacity, important as it is, becomes secondary to other performance measurements, such as the bit error probability. We will see that since known-host-state methods do not use host signal information in their detection procedures—i.e., detectors act in a "deterministic" fashion—in certain cases, they happen to be too sensitive to certain power-limited channel distortions. On the other hand, even though known-host-statistics methods present the so-called *host signal interference* (meaning that a zero probability of decoding error is not attainable), we will verify that in some cases, they prove to be robust enough to better withstand such channel perturbations. All considered, this suggests that there may exist better practical approaches that encompass both watermarking philosophies, i.e., the use of channel state information together with host signal statistics information. This door was already opened by Chen and Wornell, who proposed the idea of "spreading" known-host-state schemes (spread transforms), much in the same way as additive spread spectrum schemes improve its operating signal-to-noise ratio (SNR) [10]. In this paper, we will propose a variant called quantized projection, which will be analyzed from the perspective of its probability of bit error, providing accurate formulas for its computation and showing how computationally attractive choices of the parameters offer excellent performance and robustness in front of channel distortions.

The paper is organized as follows. In Section II, the basic definitions and conventions are formulated. Next, in Sections III and IV, we analyze the behavior of some of the most representative known-host-state and known-host-statistics methods under channel distortions.The conclusions from thses sections serve to propose the quantized projection scheme in Section V, which we demonstrate improves the former methods. Last, we compare the empirical and predicted probabilities of decoding error of all of these methods in Section VI, and we draw the final conclusions in Section VII.

II. Problem Formulation

We will restrict our analysis to data hiding in still images, as most of the prevalent algorithms have taken this kind of host signals for benchmarking purposes. The model that we will follow is summarized in Fig. 1. Let $\mathbf{x}$ be a vector containing the samples of the host signal that will convey the hidden information; these vector samples are taken in a certain domain of interest that we will discuss later. We will make the hypothesis that $x[k]$ is a zero-mean random variable (r.v.); should this not be true, the exposition would remain valid after subtracting the nonzero mean from $x[k]$. Before the encoding stage, a perceptual mask vector $\boldsymbol{\alpha}$ is computed from $\mathbf{x}$ in the appropriate domain, taking into account the characteristics of the human visual system (HVS) and indicating the maximum allowed energy that produces the least noticeable modification of the corresponding sample from the host signal. Next, using a cryptographic key K, a watermark $\mathbf{w}$ is produced from the desired binary information vector $\mathbf{b}$ using a certain function $\mathbf{w} = g_K(\mathbf{x}, \boldsymbol{\alpha}, \mathbf{b})$. Without loss of generality, we will write the watermarked signal as the addition $\mathbf{y} = \mathbf{x}+\mathbf{w}$.

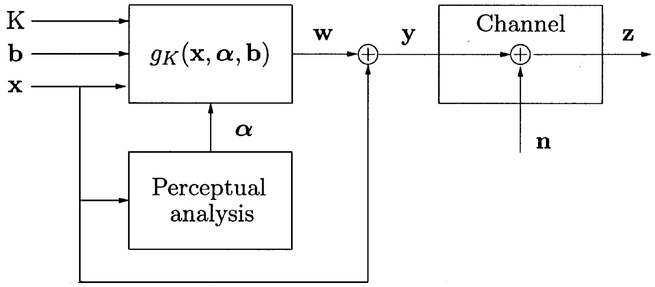

Fig. 1. Generic data hiding procedure.

In the sequel, we will concentrate only on the problem of hidden information decoding rather than watermark detection, and we will use the decoding bit error probability (P_e) as the final performance measurement. Other performance indices, such as the signal-to-noise ratio (SNR) [6], [10] or the mutual information [7] between the sent message and the channel output, have also been proposed. Nevertheless, we feel that in the data hiding problem, the probability of decoding error is the most reasonable measurement for comparing two methods. First, as it will be further developed throughout this paper, different SNRs might eventually lead to the same value of P_e under different underlying statistical models for the host signal. Second, the mutual information measurement establishes an upper bound on the admissible rate that can usually be achieved only under certain ideal conditions.

We will also assume that we want to hide only one binary digit of information that we consider to be mapped to an antipodal symbol $b \in \{\pm 1\}$. Two cases will be addressed:

- Unidimensional case, where only one host signal sample $x[k]$ is used to convey the information bit. Although not very practical by itself due to the associated high P_e, this case is interesting because it reveals the underlying mechanisms that appear when many dimensions are used.
- Multidimensional case, in which the information bit is embedded using a set $\mathcal{S} = \{k_1, \ldots, k_{|\mathcal{S}|}\}$ of key-dependent pseudorandomly chosen indices selecting $L = |\mathcal{S}|$ samples from $\mathbf{x}$. Obviously, in this case, the obtained results will depend on the partition $\mathcal{S}$ and should be averaged over all possible partitions for a given host signal (see [11]). In this paper this averaging is not undertaken, because for comparison purposes it is sufficient to consider a single partition.

It is also important to observe that in the multidimensional case, no form of coding other than repetition will be considered. As this paper targets the comparison of several families of methods, it is out of its scope the consideration of better coding strategies, which are of course possible, as it has been shown previously [11]. Thus, repetition coding can be regarded as the simplest way of achieving the necessary gain for the data hiding problem (for which the SNR is very low) and serves as a baseline for the comparison of "raw" algorithms. Moreover, repetition (or, equivalently, spreading) is very well-suited against attacks such as cropping. We will see the effect of this kind of coding with the appearance of a factor $\sqrt{L}$ (repetition coding gain) governing the asymptotic performance of the multidimensional analyses. Since increasing the size L of the embedding set will *always* reduce the probability of error, the actual choice of L will be only limited by the ratio between the number of available samples and the amount of payload to be hidden.

A. Embedding Distortion

A crucial aspect when performing a rigorous analysis lies in the election of proper distortion measures. Let us consider first the issue of *embedding distortion*. We will adhere to the often used global mean-squared error (MSE) distortion [12], [10], which for the single bit case is defined as

$$D_w = \frac{1}{L} \sum_{k \in \mathcal{S}} \mathrm{E}\{w^2[k]\}$$

where $w[k]$ is a random process representing the watermark. A straightforward generalization of this definition is possible for the multibit case.

The MSE, being adequate for measuring the total power devoted to the watermark, should be handled with care when dealing with visibility constraints. Although just constraining D_w to remain below a certain level is very convenient from an analytical point of view (e.g., Costa's result was enunciated for this type of restriction), it is at the same time questionable for the purpose of data hiding in images because it is not well-matched to the characteristics of the HVS, as explicitly mentioned in [12]. It is widely recognized that the masking phenomena affecting the HVS and exploited for the invisible embedding of information respond to local effects. All existing approaches to modeling distortions that are unnoticeable to the HVS take into account this fact, be it the just noticeable distortion function (JND) [13] in the spatial domain, the noise visibility function (NVF) [14] applicable to different domains, or equivalent models in other domains of interest like the DCT.

The main drawback encountered is that unacceptably high local distortions (from the perceptual perspective) could be globally compensated to meet the established restriction. An alternative consisting of a weighted MSE is discussed in [12]. This type of distortion measurement is appropriate when the weighting refers to local neighborhoods of the signal, e.g., sub-bands in a frequential transform or vicinities in spatial coordinates. In the extreme case, these neighborhoods consist only of one element, and the weighted MSE criterion is formally equivalent to the samplewise one, which is discussed below.

In view of the discussion above, it seems reasonable to constrain the local variance of the watermark in such a way that global compensations are not possible, and at the same time, perceptual weighting is taken into account. This is achieved by means of the following set of constraints:

$$\mathrm{E}\{w^2[k]\} \leq \alpha^2[k], \qquad \forall\, k \in \mathcal{S} \tag{1}$$

where the $\alpha[k]$ account for the perceptual characteristics of the HVS and ideally are determined to produce the least visible impact for a certain watermark *energy*. Note that if the pseudorandom samples $w[k]$ in (1) are chosen so that their mean-squared values take their extremal values, it is immediate to write

$$D_w = \frac{1}{L} \sum_{k \in \mathcal{S}} \alpha^2[k].$$

Therefore, it is clear that simultaneously meeting the set of constraints in (1) automatically leads to an MSE bound having fewer degrees of freedom for allocating the admissible distortion, which also rules out the global compensation phenomenon discussed above.

Having treated the embedding distortion D_w, let us introduce the squared root ratio of the host signal variance and the embedding distortion, i.e.,

$$\lambda \triangleq \sqrt{\sigma_x^2 / D_w}$$

that allows us to define the *document-to-watermark ratio* as $\mathrm{DWR} = 20 \log_{10} \lambda$.

B. Channel Distortion

Before arriving at the receiver, the watermarked signal undergoes an additive probabilistic channel with zero-mean noise $\mathbf{n}$ independent of $\mathbf{x}$, yielding a received signal $\mathbf{z} = \mathbf{y} + \mathbf{n}$. This channel models certain attacking operations. By virtue of the pseudorandom choice of the indices in $\mathcal{S}$, we may assume for the multidimensional case that the samples in $\mathbf{n}$ are also mutually independent, with diagonal covariance matrix $\mathbf{\Gamma} = \mathrm{diag}(\sigma_n^2[k_1], \ldots, \sigma_n^2[k_L])$.

The *channel distortion* is defined similarly to the embedding distortion, that is

$$D_c = \frac{1}{L} \sum_{k \in \mathcal{S}} \sigma_n^2[k].$$

As with the DWR, it will be useful to introduce the following square-root ratio:

$$\xi \triangleq \sqrt{D_w / D_c}$$

that relates the power of embedding and channel distortions. In addition, we will call WNR$= 20 \log_{10} \xi$ the *watermark-to-noise ratio*.

As discussed in the previous section, if some sort of perceptual shaping is introduced in the noise distribution (which will be likely the case, even for certain unintentional attacks), a simple constraint on D_c will not be enough, and rather, a set of conditions similar to (1) will be required. Typically, this shaping will imply that $\sigma_n^2[k]$ will be approximately proportional to $\alpha^2[k]$, for all $k \in \mathcal{S}$. This consideration will be useful when deriving optimal decoding strategies.

It is difficult to answer the question of what choice of the probability distribution function (pdf) of the distortion, under a restriction on its variance, causes the worst P_e on the receiver. In general, it will heavily depend on both the embedding and detection methods employed. The Gaussian channel has been commonly used in previous analyses of known-host-state data hiding schemes [7], [10] as a reference for measuring system performance; one argument supporting this choice is that Gaussian channel models can be expected to be good for a variety of applications in which robustness against unintentional attacks is required [10]. In addition, in [15], it is remarked that the additive white Gaussian noise channel is of interest because it can be easily applied to any watermarking method, giving upper capacity bounds to a more general scenario. Also, it has been shown in [16] that under the assumption of a Gaussian host signal, the Gaussian channel is the optimal attack under MSE distortion restrictions. For all these reasons, we will consider this type of distortion in our analysis.

However, we will also show how a simple uniform pdf can sometimes cause more harm to the performance of certain methods than Gaussian noise with the same variance. There are three reasons for also choosing this pdf; first, it leads in many cases to tractable analytical expressions of performance that permit gain insight into the behavior of the algorithms; second, we will show that it can be an especially harmful attack to some of the algorithms analyzed, as many of the existing known-host-state methods are based on uniform quantizers. Last, this kind of noise is unintentionally present whenever the watermarked signal is finely quantized (or requantized) [17].

C. Modeling

In Section IV, we will assume that the host signal $\mathbf{x}$ is defined in the DCT domain since a statistical model will be necessary. The reason for the election of this domain is twofold; first, since good statistical models are available for the elements of $\mathbf{x}$, it is possible to arrive at a complete theoretical formulation of the problem that will allow a fair comparison between the different classes of algorithms; second, the DCT has important practical implications (for instance, it is used in the JPEG standard). Consequently, the perceptual mask will be computed in this domain. The details into the calculation of such mask were given in [18].

III. Known-Host-State Methods

As previously said, we denote with this name those methods using side information but no host signal statistics.

A. Unidimensional Case

We will first briefly review the principles of some basic approaches for the unidimensional case. In binary quantization index modulation (QIM) [19], the basic procedure is to quantize x using one of two quantizers, depending on the binary value b to be embedded. We will concentrate our attention on the special case of QIM known as dither modulation (DM) [20], as it is the one that has been commonly analyzed in the literature; however, the same kind of analysis is valid for other quantization approaches. In this case, the centroids $Q_{-1}(\cdot)$ and $Q_1(\cdot)$ of the dithered quantizers, depicted in Fig. 2, respectively, belong to the unidimensional lattices

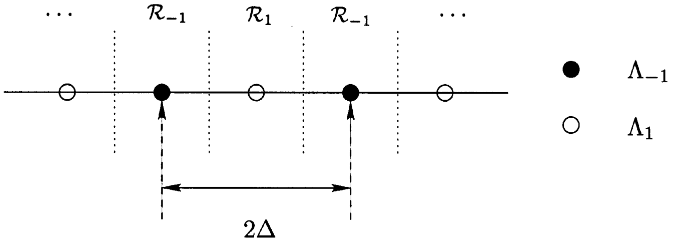

Fig. 2. Uniform dithered quantizer.

$$\Lambda_{-1} = 2\Delta\mathbb{Z} + d \tag{2}$$

$$\Lambda_1 = 2\Delta\mathbb{Z} + d + \Delta \tag{3}$$

where d is an arbitrary value that may be made key-dependent. Since this offset d is known to the decoder, its value is unimportant for the performance analysis; we choose $d = 0$ for the remainder of the section. In this way , the watermarked signal turns out to be the quantization centroid closest to x, i.e.,

$$y = Q_b(x) = x + w$$

the watermark being just the quantization error:

$$w = e = Q_b(x) - x.$$

Assuming that the quantization bins are small and $f_x(x)$ smooth enough, leading to the validity of Schuchman's condition [17], the watermark can be considered to have a pdf that is roughly constant within each individual bin.

Taking for instance the lattice Λ_{-1}, this means that $f_x(x) \approx f_i$, $x \in (2\Delta i + d, 2\Delta(i+1) + d)$, and $i \in \mathbb{Z}$. In this way, the quantization error has a uniform pdf on $(-\Delta, \Delta)$, and the distortion for a given cell is the variance of this uniformly distributed random variable, i.e., $\Delta^2/3$. As both quantizers are uniform and dithered, the average embedding distortion will also be $D_w = \Delta^2/3$.

Decoding is simply performed by quantization of $z = y + n$, which amounts to a minimum Euclidean distance decoder

$$\hat{b} = \arg\min_{-1,1} \|z - Q_b(z)\|^2. \tag{4}$$

Distortion-compensated QIM (DC-QIM) [6], which is also called the scalar Costa scheme (SCS) in [7] for the DM case, follows a principle similar to that of QIM, but it tries to practically approach the results of Costa's optimum codebook. For this purpose, it takes as the watermark the quantization error scaled by a certain constant ν, i.e.,

$$w = \nu \cdot e = \nu(Q_b(x) - x)$$

which is alleged to work in the same way as the optimizable constant α used in [5]. Observe that for $\nu = 1$, the method reduces to QIM. Once again, and as we will do throughout the paper, we consider only in our analysis the dither modulation implementation, i.e., DC-DM. Assuming again uniformity in the quantization bin, we have that the quantization error e is also uniformly distributed in the interval $(-\nu\Delta, \nu\Delta)$, and the watermarked signal

$$y = x + w = Q_b(x) - (1-\nu)\, e$$

follows now a uniform distribution $U(Q_b(x) - (1-\nu)\Delta, Q_b(x) + (1-\nu)\Delta)$. Therefore, $D_w = \nu^2\Delta^2/3$, and decoding is also made, as in (4).

We should remark that there is a difference between this procedure and the assumptions of Costa's result, where the host signal interference rejection is achieved specifically for Gaussian sources. By contrast, for the DC-DM method, host signal statistics are unimportant because in any case, uniformity is assumed to hold inside the quantization bins.

In addition, DC-DM can be modified without transgressing orthogonality in the sense specified in [7]. Clearly, we can use the uniform quantization error to generate the watermark following any arbitrary law $w = T(e)$ and not just $T(e) = \nu \cdot e$, as in DC-DM. Notice that the variable obtained by this transformation is also orthogonal to x. In fact, and just to illustrate the possible advantages brought about by this formulation, we propose as an academic example a variant of DC-DM named GDC-DM (Gaussian DC-DM) in which

$$y = x + T(e) = Q_b(x) + G(e)$$

where $G(\cdot)$ is the transformation of a uniform variable into a zero-mean Gaussian variable with variance σ_ν^2. For this end, the normalized error e is fed into the inverse complementary cumulative Gaussian distribution function $Q^{-1}(\cdot)$, with $Q(x) \triangleq (1/\sqrt{2\pi}) \int_x^\infty e^{-(\tau^2/2)}\, d\tau$. This transformation just maps the uniform error e to a Gaussian distribution. Since this new pdf is unbounded, it might well occur that 1) the watermark becomes more perceptible, and 2) it could induce more decoding errors. Concerning visibility, as long as σ_ν^2 is small enough, the local MSE constraints enunciated in Section II-A are not violated, but it is clear that these constraints do not capture all perceptual phenomena; therefore, applicability of GDC-DM would require prior extensive perceptual testing, constituting an open line of research. As for the decoding performance, the shape of the new pdf will prove to be advantageous when the attacking distortion is uniform noise, as in this case, the pdf of z will yield higher probabilities inside the correct detection bin (see Section III-A1).

As now $w = e + G(e)$, and noting that e and $G(e)$ are not independent random variables, the embedding distortion in this case is

$$D_w = \int_{-\Delta}^{\Delta} \left(e + \sigma_\nu\, Q^{-1}\left(\frac{e+\Delta}{2\Delta}\right)\right)^2 \frac{1}{2\Delta}\, de$$

that can be rewritten as

$$D_w = \frac{\Delta^2}{2}\, H\left(\frac{\sigma_\nu}{\Delta}\right) \tag{5}$$

where

$$H(\tau) \triangleq \int_{-1}^{1} \left(e + \tau\, Q^{-1}\left(\frac{e+1}{2}\right)\right)^2 de. \tag{6}$$

Once again, we must note that neither DM nor DC-DM nor GDC-DM use statistical knowledge about $\mathbf{x}$ to extract the embedded information.

1) Performance Analysis: Let $\mathcal{R}_{-1}$ and $\mathcal{R}_1$ denote the decision regions associated with $\hat{b} = -1$ and $\hat{b} = 1$, respectively. In addition, let us assume that the watermarked sample is corrupted by an additive random distortion n with pdf $f_n(n)$, yielding $z = y + n = x + w + n$. In this case, it is straightforward to determine P_e since by symmetry

$$\begin{aligned} P_e &= P\left\{\|z - Q_1(z)\|^2 < \|z - Q_{-1}(z)\|^2 | b = -1\right\} \\ &= P\left\{z \in \mathcal{R}_1 | b = -1\right\}. \end{aligned}$$

In order to obtain an integral expression for P_e, it is useful to take into account that this probability is independent of the quantization bin of $Q_{-1}(\cdot)$ in which x lies. This independence with the quantization bin is a consequence of the periodicity in $\mathcal{R}_1$ and the assumption of $f_x(x)$ constant within any bin. Thus, it is sufficient to compute P_e by conditioning x to lie in, say, the 0th bin of $Q_{-1}(\cdot)$ so that $Q_{-1}(x) = 0$.

Now, let $t \triangleq z - Q_b(x)$. Then, we can write

$$P_e = \int_{\mathcal{R}_1} f_z(z | b = -1;\, Q_{-1}(x) = 0) = \int_{\mathcal{R}_1} f_t(t)\, dt. \tag{7}$$

It is easy to see that for the uncompensated DM $t = n$; therefore, in this case, $f_t(t) = f_n(t)$. On the other hand, for DC-DM and GDC-DM, $t = n - (1-\nu)e$ and $t = n + G(e)$, respectively; consequently, $f_t(t)$ will be the convolution of $f_n(n)$ with a uniform and a normal pdf, respectively.

QIM (and so, DM) has been considered to a large extent as a "provably good" method, due to its property of presenting zero probability of decoding error for certain amplitude-bounded attacks. Indeed, for DM as long as $f_t(t) = 0$ for $|n| > \Delta/2$, we have $P_e = 0$. The same can be said of DC-DM, whereas GDC-DM presents a nonzero P_e, even with no attacks. For this kind of bounded attacks, we can write (7) as

$$P_e = \int_{\mathcal{R}_1 \cap \mathcal{E}} f_t(t)\, dt \tag{8}$$

where $\mathcal{E}$ the interval centered at $Q_{-1}(x)$, where $f_t(t) > 0$.

a) Uniform Noise: Let us assume a uniform pdf between $-\eta$ and η as the additive random attack. It can be argued that for a uniform quantizer and under the ignorance of the position of the quantization centroids, this is a worst-case attack. Then, the distortion introduced by this channel (which is just the noise power) is $D_c = \eta^2/3$.

For DM, we have that $f_t(t) = 1/2\eta$ and $\mathcal{E} = (Q_{-1}(x) - \eta, Q_{-1}(x) + \eta)$. Then, following (8), one finds that the probability of error is

$$P_e = \begin{cases} 0, & \xi \geq 2 \\ 1 - \dfrac{\xi}{2}, & 2/3 \leq \xi < 2 \end{cases}$$

with $\xi = \Delta/\eta$. As previously commented, we see that as long as the noise is bounded in a way such that $\xi \geq 2$, we have that P_e is zero, which provides a certain degree of provable robustness. Unfortunately, when $\xi < 2$, the bit error probability starts to grow rapidly. It is interesting to note that when $\xi = 1$, $P_e = 0.5$;

therefore, the channel is useless. In fact, the bit error probability can take a maximum value of $2/3$ when $\xi = 2/3$. From these results, we can conclude that for uniform channels with a distortion comparable with that introduced by the watermark, the DM modulation scheme in the unidimensional case produces poor results.

For DC-DM, we have that for $\eta \geq (1-\nu)\Delta$, the resulting pdf is

$$f_t(t) = \begin{cases} \dfrac{1}{2\eta}, & |t| \leq \eta - (1-\nu)\Delta \\ \dfrac{\eta + (1-\nu)\Delta - |t|}{4(1-\nu)\Delta\eta}, & \eta - (1-\nu)\Delta \\ & < |t| \leq \eta + (1-\nu)\Delta. \end{cases}$$

For $\eta < (1-\nu)\Delta$, the same formula applies swapping η and $(1-\nu)\Delta$. Now, $\mathcal{E} = [Q_{-1}(x) - \eta - (1-\nu)\Delta,\ Q_{-1}(x) + \eta + (1-\nu)\Delta)$; therefore, the probability of error for $\nu \geq 1/2$ is

$$P_e = \begin{cases} 0, & \xi \geq \dfrac{\nu}{\nu - 1/2} \\ \dfrac{(\nu - (\nu - 1/2)\,\xi)^2}{4\nu(1-\nu)\,\xi}, & \dfrac{\nu}{3/2 - \nu} \leq \xi < \dfrac{\nu}{\nu - 1/2} \\ \dfrac{5/2 - 2\nu}{3/2 - \nu} - \dfrac{3/2 - \nu}{\nu}\,\xi, & \xi < \dfrac{\nu}{3/2 - \nu} \end{cases} \tag{9}$$

where $\xi = \nu \cdot \Delta/\eta$. Note that it is still possible to have a zero probability of error for $\nu \geq 1/2$ whenever the first condition in (9) applies. Nonetheless, if $\nu < 1/2$, in any case, we will have $P_e \neq 0$. We observe that for values of $\nu < 1$, we get probabilities lower than 0.5. For instance, for the case $\xi = 1$, (9) is minimized at $\nu = 0.5$ with a $P_e = 0.25$; therefore, this method clearly outperforms DM in this scenario.

Last, we consider GDC-DM, in which the watermark has a Gaussian pdf that makes $f_t(t)$ of infinite length. In this case

$$f_t(t) = \frac{1}{2\eta}\left(Q\left(\frac{t-\eta}{\sigma_\nu}\right) - Q\left(\frac{t+\eta}{\sigma_\nu}\right)\right). \tag{10}$$

Now, P_e is computed using (7) by means of numerical integration, and $\xi = \sqrt{3D_w}/\eta$. Substituting (5) into (10) and using ξ, we can write, after some algebraic manipulations, an approximation to P_e (actually an upper bound) as

$$P_e \approx \sqrt{2/3}\,\frac{\xi\,\gamma}{H^{1/2}(\gamma)} \int_{1/2\gamma}^{\infty} \left\{ Q\left(t - \sqrt{3/2}\,\frac{H^{1/2}(\gamma)}{\xi\,\gamma}\right) - Q\left(t + \sqrt{3/2}\,\frac{H^{1/2}(\gamma)}{\xi\,\gamma}\right)\right\} dt \tag{11}$$

with $\gamma \triangleq \sigma_\nu/\Delta$ and $H(\cdot)$ defined in (6). For $\xi = 1$, this probability presents a minimum lower than 0.25 for $\gamma \approx 0.3$, which permits a performance improvement over DC-DM. This is a simple proof that DC-DM is not optimal in the sense of minimizing the probability of decoding error. In fact, even better alternatives to GDC-DM could have been chosen, for instance, taking $G(e)$ to be a truncated (amplitude limited) Gaussian pdf instead of a pure Gaussian.

b) Gaussian Noise: When the pdf of the attack is Gaussian with variance σ_g^2, the distortion introduced by this channel is simply $D_c = \sigma_g^2$. In this case, it is possible to analytically determine the value of P_e for DM since using (7), we have

$$P_e = 2\sum_{k=0}^{\infty} \int_{(4k+1)\Delta/2}^{(4k+3)\Delta/2} \frac{1}{\sqrt{2\pi}\sigma_g}\, e^{-t^2/2\sigma_g{}^2}\, dt$$

which results in

$$P_e = 2\sum_{k=0}^{\infty} \left\{ Q\left(\frac{(4k+1)}{2}\sqrt{3}\,\xi\right) - Q\left(\frac{(4k+3)}{2}\sqrt{3}\,\xi\right)\right\} \tag{12}$$

where $\xi = \Delta/(\sqrt{3}\,\sigma_g)$.

For DC-DM, we have once again the convolution of a uniform distribution with a Gaussian; thus, $f_t(t)$ has the same expression as (10) after replacing η and σ_ν by $(1-\nu)\Delta$ and σ_g, respectively. In this case, P_e is found through numerical integration, and $\xi = \nu\Delta/(\sqrt{3}\,\sigma_g)$.

As for GDC-DM, $f_t(t)$ is just the convolution of two Gaussian pdfs and, therefore, a Gaussian pdf with variance $\sigma_g^2 + \sigma_\nu^2$. A formula similar to (12) applies for P_e, and the parameter ξ takes the value $\sqrt{D_w}/\sigma_g$. Following the same kind of normalization performed in (11), we can write

$$P_e \approx 2\,Q\left(\frac{1}{2\sqrt{\frac{H(\gamma)}{2\xi^2} + \gamma^2}}\right) \tag{13}$$

with $\gamma = \sigma_\nu/\Delta$. It can be shown that (13) achieves its minimum at $\gamma \approx 0.3$ when $\xi = 1$.

Full discussion and comparisons of these results are made in Section VI.

B. Multidimensional Case

In multidimensional DM, one binary information symbol b is hidden by using a L-dimensional uniform quantizer $\mathbf{Q}_b(\cdot)$ on the host image, resulting in

$$\mathbf{y} = \mathbf{Q}_b(\mathbf{x}).$$

The uniform quantizers $\mathbf{Q}_{-1}(\cdot)$ and $\mathbf{Q}_1(\cdot)$ are such that the corresponding centroids are the points in the lattices

$$\Lambda_{-1} = 2\Delta\mathbb{Z}^L + \mathbf{d} \tag{14}$$

$$\Lambda_1 = 2\Delta\mathbb{Z}^L + \mathbf{d} + \Delta(1, \ldots, 1)^T \tag{15}$$

where $\mathbf{d}$ is an arbitrary vector that may be key-dependent to introduce an additional degree of uncertainty. In arriving at (14) and (15), we have assumed that the allowable perceptual distortion at every sample is identical.

Since the presence of an offset $\mathbf{d}$ in the above description of the lattices does not alter the final results, we will assume from now on that $\mathbf{d} = (0, \ldots, 0)^T$.

Now, let $\mathbf{z} = \mathbf{y} + \mathbf{n}$ be the watermarked image that has been corrupted by a noise vector $\mathbf{n}$. As in (4), given the channel output $\mathbf{z}$, the minimum Euclidean distance decoder simply decides

$$\hat{b} = \arg\min_{-1,1} \|\mathbf{z} - \mathbf{Q}_b(\mathbf{z})\|^2.$$

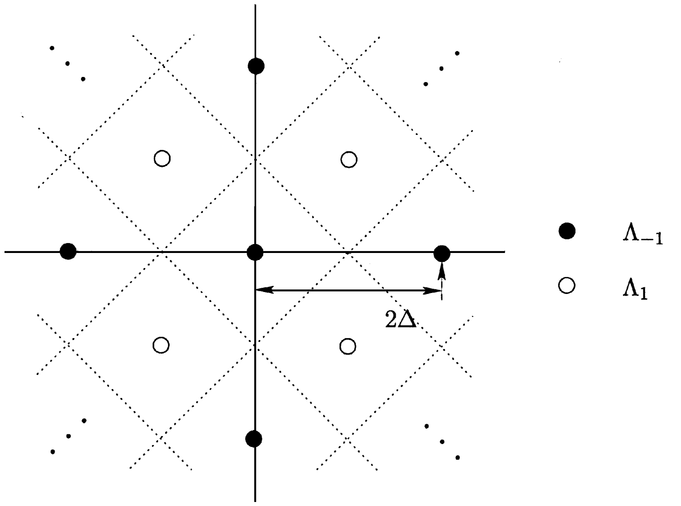

Fig. 3. Centroids and decision regions $(L = 2)$.

The decision regions associated with $\hat{b} = -1$ and $\hat{b} = 1$ are denoted by, respectively, $\mathcal{R}_{-1}$ and $\mathcal{R}_1$. In the sequel, we will find it useful to identify the decision regions associated with each of the centroids in the lattices Λ_{-1} and Λ_1. To that end, let $\mathbf{c} \in \Lambda_{-1} \bigcup \Lambda_1$ be any centroid; then, we will denote by $\mathcal{S}_\mathbf{c}$ the Voronoi cell associated with $\mathbf{c}$, i.e.,

$$\mathcal{S}_\mathbf{c} \triangleq \left\{ \mathbf{z} \in \mathbb{R}^L | \, ||\mathbf{z} - \mathbf{c}||^2 \leq ||\mathbf{z} - \mathbf{c}'||^2, \ \forall \mathbf{c}' \in \Lambda_{-1} \bigcup \Lambda_1 \right\}. \tag{16}$$

It follows immediately from (16) that

$$\mathcal{R}_b = \bigcup_{\mathbf{c} \in \Lambda_b} \mathcal{S}_\mathbf{c}, \qquad b = \{-1, 1\}.$$

The centroids and decision regions $\mathcal{R}_{-1}$ and $\mathcal{R}_1$ for the case $L = 2$ are depicted in Fig. 3. The Voronoi cells $\mathcal{S}_\mathbf{c}$ are generalized truncated octahedra [21]. We will find useful to denote by $\mathcal{S}_\mathbf{0}^*$ the generalized octahedron that contains the origin and is limited by all the hyperplanes having the form:

$$\mathbf{v}^T \left(\mathbf{z} - \frac{\Delta}{2} \mathbf{v} \right) = 0$$

where $\mathbf{v}$ is any vector such that $v[k] \in \{\pm 1\}$, $k = 1, \ldots, L$, and $\mathbf{z} \in \mathbb{R}^L$. These hyperplanes simply bisect the segments that connect the origin $\mathbf{0}$ and its nearest neighbors in Λ_1. It is straightforward to see that the set $\mathcal{S}_\mathbf{0}^*$ is also the convex hull of all the points $\mathbf{z} \in \mathbb{R}^L$ having only one nonzero component with value $\pm L\Delta/2$. Obviously, $\mathcal{S}_\mathbf{0} \subseteq \mathcal{S}_\mathbf{0}^*$, with equality only when $L = 1, 2$. Both regions are depicted in Fig. 4 for $L = 3$.

Several geometrical properties will later allow us to obtain upper bounds to the bit error probability:

P.1 [21]*:* Let $\mathcal{S}_\mathbf{0}$ be the Voronoi cell associated with the zero centroid, i.e., $\mathbf{0} = (0, \ldots, 0)^T$. Then, for any other codeword $\mathbf{c} \in \Lambda_{-1} \bigcup \Lambda_1$, its decision region is such that

$$\mathcal{S}_\mathbf{c} = \mathcal{S}_\mathbf{0} + \mathbf{c}. \qquad \square$$

P.2: By construction, it follows that the set $\mathcal{S}_\mathbf{0}^*$ is symmetric with respect to the coordinate planes. □

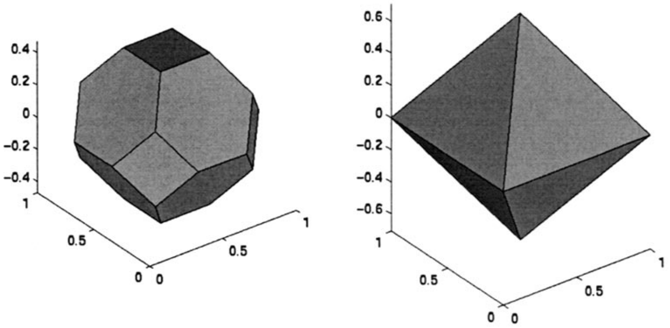

Fig. 4. Regions $\mathcal{S}_\mathbf{0}$ and $\mathcal{S}_\mathbf{0}^*$ $(L = 3)$.

P.3: $\mathcal{S}_\mathbf{0}^* \subset \mathcal{R}_{-1}$.

Proof: Without loss of generality, assume that $\mathbf{x} \in \mathcal{S}_\mathbf{0}^*$ is located in the positive orthant $\mathcal{O}$. Let $\mathbf{c}_1$ be its closest centroid in Λ_1. Then, it is easy to show that $\mathbf{c}_1$ must belong to $\mathcal{O}$ and then must be of the form $\mathbf{c}_1 = \Delta \cdot (l_1, l_2, \ldots, l_L)^T$, where l_i is an odd number such that $l_i \geq 1$ for all $i = 1, \ldots, L$. Now, consider the centroid $\mathbf{c}_{-1} = \Delta \cdot (l_1 - 1, l_2 - 1, \ldots, l_L - 1)^T$. It is obvious that $\mathbf{c}_{-1} \in \Lambda_{-1}$ and that $\mathbf{c}_{-1}$ has non-negative components. Defining the vector $\mathbf{v}$ as $\mathbf{v}^T = (1, 1, \ldots, 1)$, we can write

$$\begin{aligned} ||\mathbf{x} - \mathbf{c}_1||^2 &= ||\mathbf{x} - \mathbf{c}_{-1}||^2 + L\Delta^2 - 2\Delta \mathbf{v}^T \mathbf{x} + 2\Delta \mathbf{v}^T \mathbf{c}_{-1} \\ &\geq ||\mathbf{x} - \mathbf{c}_{-1}||^2 + 2\Delta \mathbf{v}^T \mathbf{c}_{-1} \geq ||\mathbf{x} - \mathbf{c}_{-1}||^2 \end{aligned}$$

where the first inequality follows from the fact that $2\mathbf{v}^T \mathbf{x} \leq L\Delta$ for any $\mathbf{x} \in \mathcal{S}_0^*$ and the second from the non-negativeness of the components of $\mathbf{c}_{-1}$.

Thus, for any $\mathbf{x} \in \mathcal{S}_0^*$, there is a centroid in Λ_{-1} at minimum distance among all centroids. Consequently, $\mathbf{x} \in \mathcal{R}_{-1}$. □

If the perceptual mask is not constant, it is easy to see that the quantization step in each dimension $\Delta[k]$ should be now proportional to $\alpha[k]$. This has the effect of stretching the regions $\mathcal{R}_{\pm 1}$ and, consequently, the octahedron $\mathcal{S}_\mathbf{0}^*$. If the noise variance at each sample is proportional to $\alpha^2[k]$ (perceptually-shaped noise), then it is possible to rescale both the octahedron and the noise by dividing the kth sample by $\alpha[k]$ to recover the original setup with a constant perceptual mask and i.i.d. noise. A more general setup has been considered in [22].

1) Performance Analysis: In order to obtain the bit error probability of multidimensional DM, we assume without loss of generality that the vector $\mathbf{y} = (0, \ldots, 0)^T$, which corresponds to $b = -1$, is sent. Then, we have that

$$P_e = P\{\mathbf{z} \in \mathcal{R}_1\}.$$

For the determination of P_e, one might be tempted to resort to the well-known union bound with the 2^L nearest neighbors of $\mathbf{0}$ belonging to Λ_1; unfortunately, for moderate values of D_c/D_w, the results become impractical (the bound for P_e is greater than 1) due to the overlap between the different decision regions that result when only two centroids (i.e., $\mathbf{c} = 0$ and its nearest neighbors $\mathbf{c} \in \Lambda_1$) are taken into account. On the other hand, consideration of a single nearest neighbor, as done in [10], produces overly optimistic results.

For obtaining a useful upper bound, we will follow a different strategy. Recalling properties P.1 and P.3, it is possible to conclude that

$$P_e \le P_s = P\left\{\mathbf{z} \in \overline{\mathcal{S}_{\mathbf{0}}^*}\right\}$$

where $\overline{\mathcal{S}}$ denotes the complement of $\mathcal{S}$ in $\mathbb{R}^L$.

If the pdf of $\mathbf{n}$ is symmetric with respect to the coordinate planes, from property P.2, the above evaluation of the bound P_s can be reduced to

$$\begin{aligned} P_s &= P\left\{\|\mathbf{n}\|^2 > \|\mathbf{n} - \Delta\cdot(1, \ldots, 1)^T\|^2 | \mathbf{n} \in \mathcal{O}\right\} \\ &= P\left\{\|\mathbf{n}'\|^2 > \|\mathbf{n}' - \Delta\cdot(1, \ldots, 1)^T\|^2\right\} \\ &= P\left\{\sum_{k\in\mathcal{S}} n'[k] > L\Delta/2\right\} \end{aligned} \tag{17}$$

where $\mathcal{O}$ is the positive orthant, and $\mathbf{n}'$ is an auxiliary random vector with i.i.d. components whose pdf is

$$f_{n'}(n'[k]) \triangleq \begin{cases} 2f_n(n'[k]), & n'[k] > 0 \\ 0, & \text{otherwise} \end{cases} \qquad k \in \mathcal{S}. \tag{18}$$

Now, let $r \triangleq \sum_{k\in\mathcal{S}} n'[k]$. Then, the random variable r is the convolution of L independent random variables with pdf $f_{n'}(n')$, and P_s is the integral of its tail from $L\Delta/2$ to infinity. This arrangement allows to transform the L-dimensional problem into a unidimensional one. By the central limit theorem (CLT), as $L \to \infty$, $f_r(r)$ tends to a normal curve. Then, for L very large, $f_r(r)$ can be approximated by a Gaussian pdf whose mean and variance would suffice to compute the desired probability as

$$P_s \approx Q\left(\frac{L\Delta/2 - E\{r\}}{\sqrt{\text{Var}\{r\}}}\right). \tag{19}$$

Furthermore, since the components of $\mathbf{n}$ and, therefore, those of $\mathbf{n}'$ are i.i.d., then $E\{r\} = L \cdot E\{n'\}$, and $\text{Var}\{r\} = L \cdot \text{Var}\{n'\}$.

However, a word of caution is needed here because the process of building the one-sided distribution $n'[k]$ may produce highly skewed pdfs, whose sum converges very slowly to a Gaussian distribution as L increases [23]. If this is the case, the Gaussian approximation to P_s may underestimate the importance of the tails of $f_r(r)$ and give results that are no longer an upper bound to the true P_e.

a) Uniform Noise: Here, for i.i.d. noise components, we have that $n[k] \sim U(-\eta, \eta)$, $k \in \mathcal{S}$. It is interesting to point out that although for this case it is possible to derive an analytical expression for P_e, the exact result becomes quite involved and has little interest. For this reason, we will analyze it using the Gaussian approximation described above.

Second, we remark that on the range $\eta \le \Delta$ (or, equivalently, $D_c \le D_w$), the upper bound (19) becomes in this case a good approximation of P_e. This is due to the fact that the one-sided uniform distribution is symmetric with respect to its mean and highly localized [23].

Then, following the established notation, we have that $n'[k] \sim U(0, \eta)$, $k \in \mathcal{S}$; therefore, we can compute the mean and variance of r as

$$E\{r\} = L\,\frac{\eta}{2}, \qquad \text{Var}\{r\} = L\,\frac{\eta^2}{12}$$

and P_e can be approximated by

$$P_e \begin{cases} \approx Q\left(\sqrt{3}(\xi - 1)\sqrt{L}\right), & 1 \le \xi < 2 \\ =0, & \xi \ge 2 \end{cases}$$

where $\xi = \Delta/\eta$. Again, for the special case $\xi = 1$, we have that $P_e = 0.5$ for any L. Note that for $\xi < 1$, the CLT approximation holds only as an upper bound.

b) Gaussian Noise: In this case, $\mathbf{n}$ is a random vector with L i.i.d. components with zero mean and variance σ_g^2, that is

$$f_n(\mathbf{n}) = \frac{1}{(2\pi\sigma_g^2)^{L/2}} \exp\left[\frac{-\mathbf{n}\Gamma^{-1}\mathbf{n}^T}{2}\right]$$

where Γ is the noise covariance matrix that takes the form $\Gamma = \sigma_g^2\mathbf{I}$, where $\mathbf{I}$ is the $L \times L$ identity matrix.

Now, the i.i.d. components of the auxiliary random vector $\mathbf{n}'$ given by (18) have a one-sided Gaussian distribution that may be regarded as a particular case of a Nakagami-m pdf with parameter $m = 1/2$ [24]. Since the one-sided Gaussian distribution is highly skewed, the CLT approximation to the bound P_s is only valid for very large L. On the other hand, this approximation becomes very simple to compute; for this reason, we give here the result, noting that its practical utility is limited by the actual values of L and ξ. Even worse, as discussed above, when ξ is large, the CLT approximation to P_s fails to give an upper bound to P_e (see Fig. 11).

For calculating (19) in this case, we follow the procedure outlined above by introducing the auxiliary random variable r and computing its mean and variance as

$$\begin{aligned} E\{r\} &= L\,\frac{2}{\sqrt{2\pi}\,\sigma_g}\int_0^\infty me^{-m^2/2\sigma_g^2}\,dm \\ &= \frac{\sqrt{2}\,L\sigma_g}{\sqrt{\pi}} \\ \text{Var}\{r\} &= L\left(\frac{2}{\sqrt{2\pi}\,\sigma_g}\int_0^\infty m^2e^{-m^2/2\sigma_g^2}\,dm - E^2\{n'\}\right) \\ &= L\sigma_g^2\left(1 - \frac{2}{\pi}\right). \end{aligned}$$

Therefore, the CLT approximation for the upper bound P_s to the bit error probability is

$$P_s \approx Q\left(\frac{\xi\sqrt{3\pi} - 2^{3/2}}{2\sqrt{\pi - 2}}\,\sqrt{L}\right)$$

where $\xi = \Delta/(\sqrt{3}\sigma_g)$ and is valid for L large.

In any case, we have succeeded in accurately computing P_s for any range of L and ξ by adapting the procedure first given by Beaulieu in [24] and [25] for analyzing the performance of equal gain diversity communications receivers in fading channels. The steps needed to solve our problem are outlined in Appendix A.

We must mention that the bound given by P_s is asymptotically tight as $\xi \to \infty$. In the limit, the probability that $\mathbf{n}$ falls in $\mathcal{R}_{-1}$ but not in $\mathcal{S}_0^*$ becomes negligible.

C. Connections With Costa's Result

Even though the random code proposed by Costa is impractical because of its exponential complexity, we can draw some interesting conclusions if we carefully analyze how the random codewords are generated. Under the conditions of [5], each codebook contains approximately $2^{NR_C(\nu)}$ codewords, with N the host signal length and $R_C(\nu) = I(U; X)$, that can be written as

$$R_C(\nu) = \frac{1}{2}\log_2\left(1 + \nu^2 \frac{\sigma_x^2}{D_w}\right).$$

Given a certain $\mathbf{x}$ and a certain message M, the encoder searches the codebook $\mathcal{U}_M$ for a codeword $\mathbf{u}_0$ such that $\mathbf{u}_0 - \nu\mathbf{x}$ is nearly orthogonal to $\mathbf{x}$. Every codebook divides $\mathbb{R}^N$ into as many regions as codewords it contains, and in each of these regions, the encoding consists in applying a certain linear mapping to $\mathbf{x}$. When $\nu = 1$, regions are mapped into single points; in other words, the encoder is equivalent to a set of vector quantizers (one for each possible message M). In this case, the number of codewords per message is $2^{NR_C(1)}$, but since $R_C(\nu)$ is increasing with ν, when ν diminishes, the number of codewords per codebook also decreases. In the limit, $R_C(0) = 0$, and we have only one codeword; for this case, the encoder is equivalent to an additive watermarking scheme that is similar to spread-spectrum watermarking techniques extensively studied in the literature, in which knowledge of the host signal is not used by the encoder.

Hence, the value of ν determines how close the encoder is to a pure vector quantizer. When $D_c = 0$, a zero probability of error can be achieved with $\nu = 1$, i.e., with a set of quantizers. However, when $D_c \neq 0$, it turns out that capacity is not achieved with a set of quantizers but with a value of ν less than one. In conclusion, as the noise power D_c increases, it is better to reduce the number of codewords and let a portion of the host signal pass through the encoder and be superimposed onto the quantization centroid. By doing this, even though part of the host signal appears at the output of the decoder as what might seem to be a host interference, the probability of error can be reduced, and thus, a higher information rate can be achieved.

IV. Known-Host-Statistics Methods

These methods rely on a probabilistic characterization of the host data that is employed in order to develop an optimal [in the maximum likelihood (ML) sense] information decoder. This statistical characterization is available in some domains such as the DCT, the discrete wavelet transform (DWT), or the discrete Fourier transform (DFT). For the sake of simplicity, we will concentrate in the case where $\mathbf{x}$ is modeled by a Laplacian pdf, which is adequate when information is hidden in the DCT domain. Thus, the pdf of each $x[k]$ has the form

$$f_{x[k]}(x) = \frac{\beta[k]}{2}\, e^{-\beta[k]|x|} \tag{20}$$

where $\beta[k] = \sqrt{2}/\sigma_x[k]$. We allow the Laplacian distribution to vary in each sample of $\mathbf{x}$ to better fit the host signal statistics in the multidimensional case.

One of the most important advantages of these type of methods happens to be the difficulty to severely modify the host signal's underlying pdf for a power-constrained channel. This relative invariance of the statistical properties of $\mathbf{x}$ is naturally translated into a graceful degradation of performance in the presence of distortion as the statistical characterization of the host signal is the basis for the optimization at the decoding stage. Thus, the P_e yielded at the decoder in a noise-free scenario will not vary too fast when applying growing channel distortions. Alternatively, disregarding the channel state introduces an inherent probability of error; those channels uses that are known beforehand to provoke an error at the decoder (due to the eventual presence of host signal samples with large amplitudes) are nonetheless employed for modulating information.

A. Unidimensional Case

As in Section III, we will consider first the unidimensional case. As a representative of the known-host-statistics family, we choose an amplitude modulation method, in which the watermarked sample y is obtained after adding the watermark w to the host image x. In this case, the watermark is computed from the information symbol as $w = s\alpha \cdot b$, with s a key-dependent pseudorandom variable for which $E\{s^2\} = 1$ and $\alpha > 0$ a quantity chosen so that the perceptual constraint is met. Then, for this method, $D_w = \alpha^2$.

1) Performance Analysis: It can be shown [18] that the optimal ML decoder for this scheme decides $\hat{b} = +1$ if

$$|y - s\alpha| < |y + s\alpha|$$

or, equivalently, if $ys > 0$. Let $z = y + n$ be the watermarked sample that has been corrupted by additive noise with pdf $f_n(n)$. Assuming that s takes the values ±1 with probability 1/2,[1] by symmetry, the bit error probability will be

$$P_e = P\{z < 0 | b = 1\} = \int_{-\infty}^{0} f_t(t - \alpha)\, dt \tag{21}$$

where $f_t(t) = f_n(t) * f_x(t)$. It is interesting to note that in the case when there is no distortion present in the channel, i.e., $n = 0$, we have that $f_t(t) = f_x(t)$; therefore

$$P_e = \tfrac{1}{2}\, e^{-(\sqrt{2}/\lambda)} \tag{22}$$

with $\lambda = \sigma_x/\alpha$. Hence, even in the absence of channel distortion, this statistical extraction method is not "provably robust": P_e is not zero even for null distortions. Nevertheless, we want to know how a nonzero distortion would modify this probability

[1] This choice minimizes the bit error probability while meeting the constraint $E\{s^2\} = 1$. The proof of this fact follows from the convexity of the integral of the tail of $f_t(t)$.

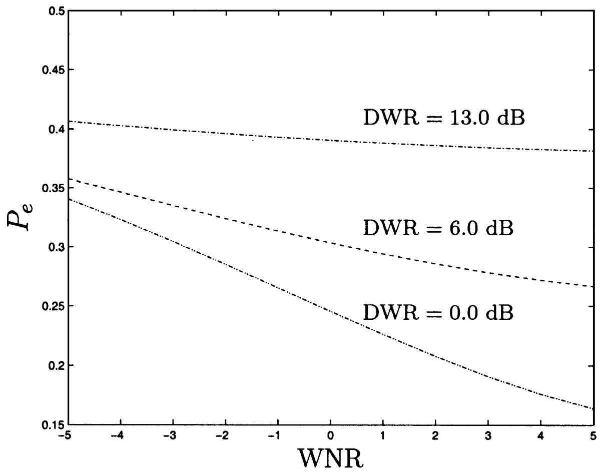

Fig. 5. Example of a known-host-statistics method dependence with the DWR (theoretical values).

by evaluating P_e for the same kind of distortions presented in Section III-A1.

a) Uniform Noise: When n follows a uniform distribution between $-\eta$ and η, we have that

$$f_t(t) = \frac{\beta}{4\eta} \int_{t-\eta}^{t+\eta} e^{-\beta|u|}\, du.$$

This integral can be split into two parts, yielding

$$f_t(t) = \begin{cases} \left(1 - e^{-\beta\eta}\cosh(\beta t)\right)/2\eta, & |t| < \eta \\ e^{-\beta|t|}\sinh(\beta\eta)/2\eta, & |t| \geq \eta. \end{cases} \tag{23}$$

Now, the derivation of P_e becomes simple by using (21)

$$P_e = \begin{cases} \lambda\xi\, \dfrac{e^{-\sqrt{2}/\lambda}}{2\sqrt{6}} \sinh\left(\dfrac{\sqrt{6}}{\lambda\xi}\right), & \xi > \sqrt{3} \\ \dfrac{1}{2} - \dfrac{\xi}{2\sqrt{3}} + \lambda\xi\, \dfrac{e^{-\sqrt{6}/(\lambda\xi)}}{2\sqrt{6}} \sinh\left(\sqrt{2}/\lambda\right) & \\ & \xi \leq \sqrt{3} \end{cases}$$

where now, $\xi = \sqrt{3}\alpha/\eta$, and $\lambda = \sigma_x/\alpha$. It is quite important to note that just as in (22), this probability of error depends on the DWR (i.e., λ). We can see in Fig. 5 its dependence with this parameter. It becomes evident that the lower the DWR, the lower the attainable probabilities of decoding error, reflecting that the statistical knowledge of $\mathbf{x}$ is successfully exploited by the decoder. At the same time, the P_e plot becomes less flat when channel distortions are progressively comparable with or greater than D_w because the noise pdf has an increasing relative importance when compared with the pdf of x.

b) Gaussian Noise: For the case in which n is a zero-mean Gaussian random variable with variance σ_g^2, we have that

$$\begin{aligned} f_t(t) &= \frac{\beta}{2\sigma_g\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\beta|x|}\, e^{-(x-t)^2/2\sigma_g^2}\, dx \\ &= \frac{\beta}{2}\, e^{\beta^2\sigma_g^2/2} \left\{ e^{-\beta t} Q\left(\frac{-t+\beta\sigma_g^2}{\sigma_g}\right) \right. \\ &\qquad \left. + e^{\beta t}\, Q\left(\frac{t+\beta\sigma_g^2}{\sigma_g}\right) \right\}. \end{aligned} \tag{24}$$

For this distortion, $\xi = \alpha/\sigma_g$, and numerical integration is required to compute P_e after inserting (24) into (21). As in the previous case, the calculated P_e will be dependent on λ, and the previous discussion on the effects of the DWR applies similarly.

B. Multidimensional Case

For the indices selected by $\mathcal{S}$, $\mathbf{y}$ is obtained as the addition of the watermark $\mathbf{w}$ to the host image $\mathbf{x}$, where $\mathbf{w}$ is computed as $w[k] = s[k]\alpha[k] \cdot b$, $k \in \mathcal{S}$. Note that this choice implicitly uses repetition coding, allowing for a fair comparison with the results obtained in Section III-B. We must recall again that better encoding alternatives have been explored to enhance these methods [11].

1) Performance Analysis: Assuming a Laplacian distribution for $\mathbf{x}$, the ML sufficient statistic [18] can be written as

$$r \triangleq \sum_{k \in \mathcal{S}} \beta[k] \left\{ \left|y[k] + s[k]\alpha[k]\right| - \left|y[k] - s[k]\alpha[k]\right| \right\}.$$

If, without loss of generality, we suppose that $b = 1$ is sent, then r can be written as

$$r = \sum_{k \in \mathcal{S}} \beta[k] \left\{ \left|x[k] + 2s[k]\alpha[k]\right| - \left|x[k]\right| \right\} \triangleq \sum_{k \in \mathcal{S}} \beta[k] P[k]. \tag{25}$$

Using this statistic, the decision made is $\hat{b} = \text{sgn}\{r\}$. For a large value of $|\mathcal{S}|$, we may assume a Gaussian approximation for r, and we compute its mean and variance as

$$\begin{aligned} E\{r\} &= \sum_{k \in \mathcal{S}} \beta[k]\, E\{P[k]\} \\ &= \sum_{k \in \mathcal{S}} \left(e^{-2\sqrt{2}/\lambda[k]} + \frac{2\sqrt{2}}{\lambda[k]} - 1 \right) \end{aligned} \tag{26}$$

$$\begin{aligned} \text{Var}\{r\} &= \sum_{k \in \mathcal{S}} \beta^2[k]\, \text{Var}\{P[k]\} \\ &= \sum_{k \in \mathcal{S}} \left(3 - e^{-4\sqrt{2}/\lambda[k]} - 2\, e^{-2\sqrt{2}/\lambda[k]} \right. \\ &\qquad \left. \cdot \left(1 + 4\, \frac{\sqrt{2}}{\lambda[k]} \right) \right) \end{aligned} \tag{27}$$

where $\lambda[k] = \sigma_x[k]/\alpha[k]$; therefore, P_e depends again on the DWR. When $b = -1$, the expectation in (26) has the opposite sign. The details of the calculation of $E\{P[k]\}$ and $\text{Var}\{P[k]\}$ can be found in Appendix B. We have to remark that in the derivation of the first- and second-order moments of r, we are taking a further step compared to what was done in [18], where the host signal $\mathbf{x}$ was considered to be deterministic.

Thus, we have that the probability of bit error is given by

$$P_e = Q\left(\frac{|E\{r\}|}{\sqrt{\text{Var}\{r\}}} \right). \tag{28}$$

As in Section IV-A1, we get $P_e \neq 0$ even without considering channel distortions. In the case where $\mathbf{y}$ is corrupted by additive noise $\mathbf{n}$, (26) and (27) have to be recalculated for the pdf of $\mathbf{u} = \mathbf{x} + \mathbf{n}$, assuming that the optimal detector for the Laplacian case is used regardless of $\mathbf{n}$. In general, the calculated probabilities will show the observed dependence on the DWR.

Interestingly, for the special case in which $\lambda[k] = \lambda$ and the watermark-to-noise ratio is constant at each sample, a factor $\sqrt{L}$ appears at the numerator inside $Q(\cdot)$ in (28), thus showing the effect of repetition coding mentioned in Section II.

a) Uniform Noise: In this case, $\mathbf{n}$ is the uniform noise vector added to the watermarked signal, with $n[k] \sim U(-\eta[k], \eta[k])$, and $k \in \mathcal{S}$. The mean and variance of $P[k]$ are derived in Appendix B.

b) Gaussian Noise: When the distortion is Gaussian noise, we have again the pdf (24) with particular parameters for each sample. For this formula, closed-form expressions of $E\{P[k]\}$ and Var$\{P[k]\}$ cannot be obtained, and we have to resort to numerical integration.

V. Quantized Projection Methods

As it was discussed in the previous section, the performance of known-host-statistics methods becomes quite independent of the distortion level that is present in the additive channel. This is attributable to the fact that for high DWRs, the statistics of the host image dominate those of the additive noise in such a way that the ML detector remains the same despite of channel distortions. Unfortunately, we will show in Section VI how known-host-state schemes usually have better performance, especially as the degree of diversity L increases. One can explain this poorer behavior of known-host-statistics methods by noticing that they are equivalent to using an infinite quantization step on the decision variable r. Thus, when the host vector $\mathbf{x}$ is such that the decision variable r for the unwatermarked case falls far apart from the decision threshold (here located at the origin), there is no possibility of effectively conveying an information bit by modifying $\mathbf{x}$ so that r falls in the desired side of the threshold because of the perceptual limit in the achievable embedding distortion.

This consideration opens the door for other watermarking schemes that effectively combine the advantages of both known-host-statistics and host-state methods. The idea is to employ a projection function that produces a scalar decision variable (playing the same role as r in known-host-statistics methods) but then uniformly quantizing this variable in a similar way as DM methods. For this reason, we will call the resulting scheme the quantized projection (QP) method. In implementing this idea, one faces the problem of finding the watermark such that when adding it to the host image, the subsequent projection results in the desired centroid. The fact that r in (25) is nonlinear in the watermark hinders the search for a solution. As a compromise, we have decided to use a standard cross-correlation for the projection function because it is linear, and it would be the optimal ML decoding function were the statistics of the host image Gaussian, as we discussed in [9].

QP methods have their roots in the influential proposal by Chen and Wornell of the spread-transform dither modulation (STDM) [10] that effectively combines quantization-based schemes with the diversity afforded by spread-spectrum methods. Specifically, we will stick to the particular case of STDM, where only one transformed component is quantized since, as we will show, it yields the minimum probability of error. However, we will depart from this form of STDM to show that for large spreading gains, the hypothesis of host data uniform inside the Voronoi cells has to be abandoned, with the consequence that the size of the cells can be larger, and thus, the performance can be better than expected. Unfortunately, as we will see, this perspective complicates the corresponding theoretical analysis. Another improvement over the original STDM method is that here, we will take into account not only that the embedding power is in general different at each sample but that this also occurs to channel distortion (see Section II-A).

Soon after the STDM method was proposed, Ramkumar *et al.* put forward the idea of improving it by means of distortion compensation, in what was called *Type III* schemes [26], computing the achievable rate for the special case in which the embedding distortion is hard-limited at a certain value. On the other hand, Eggers *et al.* considered in [7] the so-called spread-transform scalar Costa scheme (STSCS), which is equivalent to compensating the embedding distortion in the STDM method, and determined its achievable rate. Taking into account that none of these works has dealt with theoretical error probabilities, in Section V-B, we will also give rise to the concept of distortion-compensated QP (DC-QP) and analyze its performance.

A. Watermarking With Correlation-Based Uniform QP

In the basic QP case, the projection function consists of computing a weighted cross-correlation between the watermarked image and the watermark; therefore, for a single transmitted bit, the projection r is such that

$$r \triangleq \sum_{k \in \mathcal{S}} \frac{y[k]s[k]}{\alpha[k]}. \tag{29}$$

In a second stage, r is quantized with a uniform scalar quantizer with step 2Δ so that the centroids of the decision cells associated with $\hat{b} = 1$ and $\hat{b} = -1$ are given by the unidimensional lattices in (2) and (3) with $d = -\Delta/2$, due to symmetry considerations on the pdf of the host signal projection that we will see next.

It can be shown that dividing by $\alpha[k]$ in (29) becomes the optimal decoding strategy whenever the noise is perceptually shaped, i.e., when its variance is proportional to the perceptual mask. Although it is possible to determine the optimal decoding strategy for any other noise joint distribution, this path will not be taken here. In this regard, it is important to note that the STDM proposal in [10] was carried out with the same spreading vector for both embedding and decoding stages, which is not our case. Of course, the improvement given by the structure proposed in (29) will become important when the set of perceptual masks $\alpha[k]$, $k \in \mathcal{S}$ has a significant “spectrum,” and it will be null for constant perceptual masks.

If $w[k]$ denotes the kth sample of the watermark, then it is possible to rewrite (29) as

$$r = r_x + r_w$$

where r_w is the projected watermark

$$r_w = \sum_{k \in \mathcal{S}} \frac{w[k]s[k]}{\alpha[k]} \tag{30}$$

and with a similar definition for the projected host image r_x.

The projected watermark r_w is selected in such a way that when added to the projected image $r_x = \sum_{k\in\mathcal{S}} x[k]s[k]/\alpha[k]$, the result is a member of the desired lattice. Thus, in order to transmit $b=1$, the embedder finds r_w with the smallest magnitude such that $r_x + r_w \in \Lambda_1$, which can be immediately adapted to account for the case $b=-1$.

Now, the problem that faces the embedder is to select the watermark samples $w[k]$, $k\in\mathcal{S}$ so that (30) is satisfied. Since there are infinitely many solutions to this problem, it would be possible to exploit this fact to provide additional robustness against attacks. In fact, this problem resembles the so-called *knapsack* problem [27] (an NP-complete one) that was the base for some cryptographic algorithms, although it was later abandoned for not providing enough security. For our purposes, we will content ourselves with choosing $w[k]$, $k\in\mathcal{S}$, to become proportional to $\alpha[k]$. It can be shown that under our perceptual constraints, this choice minimizes the probability of error. Then, the watermark samples are perceptually weighted by their respective masks. Then

$$w[k] = \rho\alpha[k]s[k] \tag{31}$$

with ρ a real number that will be determined next. To this end, simply substitute (31) into (30) so that $\rho = r_w/L$, and finally

$$w[k] = \frac{r_w\alpha[k]s[k]}{L}.$$

Noticing that r_w and the pseudorandom sequence $s[k]$ are statistically independent and that $E\{s^2\}=1$, it is immediate to write

$$D_w = \frac{1}{L}\sum_{k\in\mathcal{S}} E\{w^2[k]\} = \frac{E\{r_w^2\}\sum\limits_{k\in\mathcal{S}}\alpha^2[k]}{L^3} \tag{32}$$

where $L = |\mathcal{S}|$.

It is interesting to note that while the random variables $w[k]$, $k\in\mathcal{S}$ are mutually uncorrelated, this is far from being the case if one considers the random variables $w[k]\cdot s[k]/\alpha[k]$, $k\in\mathcal{S}$.

In order to simplify the performance analysis of this data hiding method, let us assume a constant perceptual mask, i.e., $\alpha[k]=\alpha$, for all $k\in\mathcal{S}$. Although an analysis for varying perceptual masks is still possible by following the lines here presented, our assumption leads to more compact results. With this assumption, (32) becomes

$$D_w = \frac{E\{r_w^2\}\alpha^2}{L^2}. \tag{33}$$

Now, we are left with the problem of evaluating $E\{r_w^2\}$. To meet this objective, it is necessary to statistically characterize the random variable r_x. Since the $s[k]$, $k\in\mathcal{S}$ are statistically independent, it is possible to resort to the CLT to show that for large L,[2] r_x can be accurately modeled by a Gaussian pdf with zero mean and variance given by

$$\sigma_{r_x}^2 = \frac{L\sigma_x^2}{\alpha^2} \tag{34}$$

where $\sigma_x^2 = E\{x^2[k]\}$. Of course, a Gaussian pdf also results for any L if the host image $\mathbf{x}$ is normally distributed. In any case, assuming an equiprobable information bit b, we have

$$E\{r_w^2\} = \frac{E\{r_w^2|b=1\} + E\{r_w^2|b=-1\}}{2} \tag{35}$$

where

$$\begin{aligned}
&E\{r_w^2|b=1\}\\
&\quad= \sum_{i=-\infty}^{\infty}\int_{2i\Delta-\Delta/2}^{2(i+1)\Delta-\Delta/2} f_{r_x}(r_x)(2i\Delta+\Delta/2-r_x)^2\,dr_x\\
&\quad= \sum_{i=-\infty}^{\infty}\int_{-\Delta/2}^{3\Delta/2} f_{r_x}(r_x+2i\Delta)(\Delta/2-r_x)^2\,dr_x\\
&\quad= \sum_{i=-\infty}^{\infty}\int_{-\Delta/2}^{\Delta/2} f_{r_x}(r_x+2i\Delta)(\Delta/2-r_x)^2\,dr_x\\
&\qquad+ \sum_{i=-\infty}^{\infty}\int_{-\Delta/2}^{\Delta/2} f_{r_x}(r_x+(2i+1)\Delta)(\Delta/2+r_x)^2\,dr_x
\end{aligned} \tag{36}$$

and where $f_{r_x}(r_x)$ is the pdf of r_x. Similarly, we can write

$$\begin{aligned}
&E\{r_w^2|b=-1\}\\
&\quad= \sum_{i=-\infty}^{\infty}\int_{-\Delta/2}^{\Delta/2} f_{r_x}(r_x+2i\Delta)(\Delta/2+r_x)^2\,dr_x\\
&\qquad+ \sum_{i=-\infty}^{\infty}\int_{-\Delta/2}^{\Delta/2} f_{r_x}(r_x+(2i-1)\Delta)(\Delta/2-r_x)^2\,dr_x.
\end{aligned} \tag{37}$$

Substituting (36) and (37) into (35) and operating, we have

$$\begin{aligned}
E\{r_w^2\} &= \frac{\Delta^2}{4} + \sum_{i=-\infty}^{\infty}\int_{-\Delta/2}^{\Delta/2} f_{r_x}(r_x+i\Delta)\,r_x^2\,dr_x\\
&= \Delta^2\left(\tfrac{1}{4} + I(\sigma_{r_x}/\Delta)\right)
\end{aligned} \tag{38}$$

where

$$I(\sigma) \triangleq \frac{1}{\sqrt{2\pi}\sigma}\sum_{i=-\infty}^{\infty}\int_{-1/2}^{1/2} e^{-(r_x+i)^2/2\sigma^2}\,r_x^2\,dr_x. \tag{39}$$

Having obtained the distortion D_w for arbitrary Δ, σ_x and α, we will determine the bit error probability. As before, let $z[k]=y[k]+n[k]$, $k\in\mathcal{S}$, where $\mathbf{n}$ has zero-mean i.i.d. components with an arbitrary pdf and variance σ_n^2 so that $D_c=\sigma_n^2$. Now, the projection r becomes

$$r = r_x + r_w + r_n$$

where the projected watermark r_w was defined in (30), and the projected host image r_x and noise r_n are similarly defined using $x[k]$ and $n[k]$, respectively, instead of $w[k]$.

For large L, we can again apply the CLT to state that for a wide class of distributions in $n[k]$, the pdf of r_n can be approximated by a zero-mean Gaussian pdf with variance $\sigma_{r_n}^2$ equal to

$$\sigma_{r_n}^2 = \frac{L\sigma_n^2}{\alpha^2} = \frac{LD_c}{\alpha^2}. \tag{40}$$

[2] The validity of the large L assumption is supported by the results of Section V-C.

The bit error probability P_e can be determined by taking into account the symmetry in the problem; therefore, it is enough to consider the errors made when decoding a transmitted $b = 1$. Similarly to (12), this probability can be written as

$$P_e = 2\sum_{k=0}^{\infty}\left\{Q\left(\frac{(4k+1)\Delta}{2\sigma_{r_n}}\right) - Q\left(\frac{(4k+3)\Delta}{2\sigma_{r_n}}\right)\right\}. \quad (41)$$

In order to rewrite P_e in terms of the desired parameters, we may use (33) and (38) to obtain

$$\Delta = \frac{\sqrt{D_w}L\tau}{\alpha} \quad (42)$$

with τ such that

$$\tau \triangleq (1/4 + I(\sigma_{r_x}/\Delta))^{-1/2}. \quad (43)$$

Now, using (40) and (42), it is immediate to write

$$\frac{\Delta}{\sigma_{r_n}} = \xi\sqrt{L}\tau \quad (44)$$

which can be plugged into (41) to yield the desired expression for P_e. For a further interpretation of this result, let us write τ in (43) in terms of the DWR, the WNR, and L. Thus, by making use of (34) and (42), and after some trivial algebra, we obtain

$$\begin{aligned}\frac{1}{\lambda} = \frac{\sqrt{D_w}}{\sigma_x} &= \frac{\Delta}{\sqrt{L}\sigma_{r_x}}\left(1/4 + I\left(\frac{\sigma_{r_x}}{\Delta}\right)\right)^{1/2}\\ &= \frac{1}{\sqrt{L}}F\left(\frac{\Delta}{\sigma_{r_x}}\right) \quad (45)\end{aligned}$$

with the function $F(\cdot)$ defined as

$$F(x) \triangleq x\sqrt{1/4 + I(1/x)}.$$

It can be shown that $F(x)$ is one to one and monotonically increasing for $x > 0$. Then, inversion of (45) yields

$$\frac{\Delta}{\sigma_{r_x}} = F^{-1}\left(\frac{\sqrt{L}}{\lambda}\right). \quad (46)$$

Finally, τ can be written as

$$\tau = \left[1/4 + I\left(\frac{1}{F^{-1}\left(\frac{\sqrt{L}}{\lambda}\right)}\right)\right]^{-1/2}.$$

It is easy to see that τ monotonically increases from $\sqrt{3}$ to $\sqrt{4}$ as the ratio $\sqrt{L}/\lambda$ goes from 0 to ∞. Thus, the parameter τ can be interpreted as a measure of the uniformity of the projected host signal r_x within the quantization bins of size Δ. If $f_{r_x}(r_x)$ is constant within any bin, $\tau = \sqrt{3}$. However, it is interesting to see that as $\sqrt{L}/\lambda$ increases, a significant portion of $f_{r_x}(r_x)$ (which will have a Gaussian shape) will be contained in the two bins closest to the origin, and then, τ approaches $\sqrt{4}$. Consequently, by increasing L, one may obtain an additional gain on the SNR of at most $4/3$ (1.25 dB), besides the expected diversity gain of L. This additional gain depends not only on L but on the document-to-watermark ratio λ as well; therefore, the smaller the DWR, the larger this extra gain will be if L is kept constant. From this latter observation, we may conclude that P_e actually depends on the DWR as it happened with known-host-statistics methods although now in a weaker fashion. Finally, we must note that since in practice very large values of L are required for this gain to become significant, this desirable feature will show up in robust data hiding applications (i.e., for negative WNRs), where it is also most welcome.

Last, when Δ/σ_{r_n} is large, the sum (41) is dominated by the term corresponding to $k = 0$; therefore, P_e can be approximated by

$$P_e \approx 2Q\left(\frac{\Delta}{2\sigma_{r_n}}\right) = 2Q\left(\frac{\xi\sqrt{L}\tau}{2}\right)$$

where again, the right-hand side is actually an upper bound to P_e.

B. Distortion-Compensated Quantized Projection

The projection function and quantization centroids are the same as in the previous section, as given by (2) and (3) as well as (29). Given the projected host image r_x, the projected watermark r_w is selected in such a way that

$$r_w = \nu(Q_b(r_x) - r_x) \quad (47)$$

where $Q_b(x)$, $b = -1, 1$ denotes the closest centroid to x in the lattices Λ_b, $b = -1, 1$ given in (2) and (3), with $d = -\Delta/2$.

The watermark samples are selected according to the method presented in Section V-A, which implies choosing the scaling factor ρ as in the QP method, where r_w is now given by (47). Compared with the previous method, this Costa-based variant of the QP method, which we will term distortion-compensated QP (DC-QP) in the sequel, scales the quantization error in the projected domain by ν. This in turn implies that the development in Section V-A for determining $E\{r_w^2\}$ can be easily adapted to the present case to yield

$$E\{r_w^2\} = \nu^2\Delta^2\left(\tfrac{1}{4} + I(\sigma_{r_x}/\Delta)\right) \quad (48)$$

where $I(\sigma)$ was given in (39).

Regarding the bit error probability P_e for a zero-mean additive i.i.d. noise channel with distortion $D_c = \sigma_n^2$, the problem becomes more involved due to the presence of the residual error. We will assume without loss of generality that the symbol $b = 1$ is sent. Then, the probability of bit error can be written as

$$P_e = \sum_i \int_{2i\Delta-\Delta/2}^{2(i+1)\Delta-\Delta/2} f_{r_x}(r_x)P(\text{ error}|r_x)\,dr_x \quad (49)$$

where $f_{r_x}(r_x)$ is the pdf of r_x, and $P(\text{error}|r_x)$ is the probability of error for a given value of r_x. Recall that $f_{r_x}(r_x)$ is a zero-mean Gaussian pdf with variance $\sigma_{r_x}^2 = L\sigma_x^2/\alpha^2$. In order to determine $P(\text{error}|r_x)$, note that if

$$r_x \in [2i_*\Delta - \Delta/2,\ 2(i_*+1)\Delta - \Delta/2) \quad (50)$$

for some integer i_*, then following (47), the undistorted projected watermarked image $r = r_x + r_w$ becomes

$$r = \nu Q_1(r_x) + (1-\nu)r_x = \nu(2i_*\Delta + \Delta/2) + (1-\nu)r_x.$$

When Δ/σ_{r_x} is large, it is possible to simplify the analysis by considering that whenever there exists a decoding error, it is due to $r_x + r_w + r_n$ lying in the Voronoi cells associated to one of the neighboring centroids to $r_x + r_w$ in Λ_{-1}, namely,

$2(i_* + 1)\Delta - \Delta/2$ and $2i_*\Delta - \Delta/2$. From here, it is possible to conclude that

$$\begin{aligned} P(\text{error}|r_x) & \\ \approx Q & \left(\frac{(2i_* + 1)\Delta - \nu(2i_*\Delta + \Delta/2) - (1-\nu)r_x}{\sigma_{r_n}} \right) \\ + Q & \left(\frac{\nu(2i_*\Delta + \Delta/2) + (1-\nu)r_x - 2i_*\Delta}{\sigma_{r_n}} \right) \end{aligned} \quad (51)$$

where i_* is such that (50) holds.

After substituting (51) into (49) and after some tedious but straightforward algebraic manipulations, it is possible to arrive at the following result:

$$\begin{aligned} P_e \approx \sum_i \frac{\Delta}{\sqrt{2\pi}\,\sigma_{r_x}} \int_{-1/2}^{3/2} & e^{-\Delta^2 (r_x + 2i)^2 / 2\sigma_{r_x}^2} \\ \cdot \Bigg[Q & \left(\frac{(1 + (1-\nu)(1-2r_x))}{2\nu} \xi\tau\sqrt{L} \right) \\ + Q & \left(\frac{(1 + (1-\nu)(2r_x - 1))}{2\nu} \xi\tau\sqrt{L} \right) \Bigg] \, dr_x \end{aligned} \quad (52)$$

where τ was defined in (43). The integral in (52) must be evaluated numerically, but it is interesting to see that in any case, it depends only on ξ, λ, L, and ν, i.e., the WNR, the DWR, the number of dimensions used, and the size of the residual error, respectively. Bear in mind that the ratio $\Delta^2/\sigma_{r_x}^2$ in (52) may be written as a function of L and the DWR [cf. (46)].

It is interesting to optimize P_e in (52) in terms of ν. In Fig. 6, we plot P_e as a function of ν for $\xi = 1$, $L = 20$ and for two values of DWR. Some important remarks can be drawn:

• The optimal value of ν that results in each case is smaller than one; hence, note that the DC-QP method offers to the designer an improvement over the QP method by choosing an appropriate value of ν. Note that as QP is equivalent to DC-QP with $\nu = 1$, it is clear that DC-QP for the *optimal* ν will never perform worse than QP. However, if the operating WNR is not known at the embedder's side (recall that the WNR also depends on the attacking power), it may be safer to set $\nu = 1$ since too small a ν will increase the bit error probability (cf. Fig. 6).

• Since the DC-QP scheme resembles Costa's method, one might expect that the optimal parameter ν, there derived for maximizing capacity, should become similar to the one here obtained. In Costa's paper, the optimal value is $\nu = 1/(1 + \text{NSR})$ where NSR is the noise-to-signal ratio. In the QP method, the NSR after projection becomes $\sigma_{r_n}^2/E\{r_w^2\}$, which can be rewritten as

$$\frac{\sigma_{r_n}^2}{E\{r_w^2\}} = \frac{D_c}{LD_w}. \quad (53)$$

Consequently, the NSR for the QP method approaches 0 asymptotically when $L \to \infty$, and then, the optimal value of ν achieved with Costa's procedure approaches 1 asymptotically for $L \to \infty$. Numerical optimization experiments show that this is in fact the case when P_e is minimized in terms of ν so that DC-QP asymptotically approaches QP.

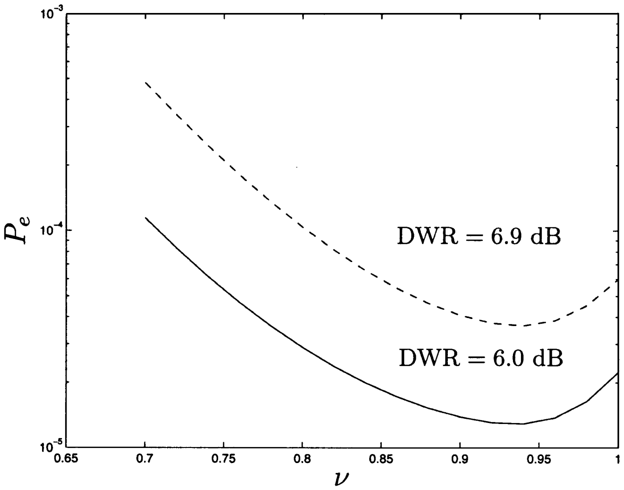

Fig. 6. Theoretical bit error probability versus ν for different values of the DWR, $D_c = D_w$, and $L = 20$.

• For moderate L, experimentation shows that the optimal value of ν depends on the ratio D_w/D_c, as is to be expected from (53). Nevertheless, optimization experiments (cf. Fig. 6) show that the optimal ν also depends on the DWR in contrast with Costa's result in which the host image (there the "channel state") does not show up in the optimal value of ν. The reason for this difference must be found in the fact that Costa's method has unlimited complexity, which is obviously not our case (unless L is made unpractically large). Interestingly, knowledge of the image second-order statistics becomes useful when trying to optimize performance. Moreover, for larger DWRs, there is more room for improvement by choosing the proper value of ν.

C. Generalizing the DC-QP Method

The DC-QP method described in Sections V-A and B can be generalized in such a way that the quantization operations take place in a vector subspace, introducing dimensionality as an additional degree of freedom. This idea is borrowed from the general formulation of STDM given by Chen and Wornell [10] and has the advantage of filling the gamut between the basic DC-QP method and a multidimensional distortion-compensated dither modulation (DC-DM) method (itself a generalization of the multidimensional DM scheme), for which the dimensionality of the projected subspace equals that of the set $\mathcal{S}$. Formally, let M be the dimensionality of the projected subspace; then, for the DC-QP method, $M = 1$, whereas for the DC-DM method, $M = L$, with $L = |\mathcal{S}|$.

Then, (29) now reads

$$\mathbf{r} = \mathbf{P}^T\mathbf{y} \quad (54)$$

where $\mathbf{P} = (\mathbf{v}_1, \ldots, \mathbf{v}_M)$ is a projection matrix with orthogonal columns. Several ways of constructing $\mathbf{P}$ are possible, all giving essentially the same results; therefore, here, we will adopt the simplest, also used for example in [26], which consists of dividing the set of indices $\mathcal{S}$ in M nonoverlapping subsets $\mathcal{S}_i$,

$i = 1, \ldots, M$, each with cardinality L/M.[3] Then, for the ith vector $\mathbf{v}_i$, we have that

$$v_i[k] = \begin{cases} s[k]/\alpha[k], & k \in \mathcal{S}_i \\ 0, & \text{otherwise.} \end{cases}$$

As before, we can write the projected watermarked image in terms of the projected host image and the projected watermark, i.e., $\mathbf{r} = \mathbf{r}_x + \mathbf{r}_w$. Then, in a generalized QP scheme (that is, with no distortion compensation), the watermark $\mathbf{w}$ should be selected as the minimum Euclidean norm vector such that $\mathbf{r}$ belongs to the desired information-dependent lattice, as in (14) and (15), with M dimensions instead of L.

In a generalized DC-QP scheme, $\mathbf{w}$ is chosen as the minimum norm vector such that its projection $\mathbf{r}_w$ satisfies $\mathbf{r}_w = \nu \mathbf{e}$, where ν is a selectable real parameter, and $\mathbf{e} = \mathbf{Q}_b(\mathbf{r}_x) - \mathbf{r}_x$, with $\mathbf{Q}_b(\cdot)$ an M-dimensional quantizer as in (14) and (15).

We are interested in determining the value of M that minimizes P_e for a given WNR and for the optimal value of ν. A similar setup is considered in [26] and [7] to obtain the value of M that maximizes the achievable rate of a certain data hiding method. In order to keep the discussion simple, we will assume constant perceptual masks, i.e., $\alpha[k] = \alpha$, $k \in \mathcal{S}$, and i.i.d. Gaussian noise. First, note that the analysis of Section V-A can be repeated to derive a relation between the projected-noise variance σ_{r_n} and the quantization step Δ so that (44) transforms into

$$\frac{\Delta}{\sigma_{r_n}} = \frac{\xi \sqrt{L/M} \tau}{\nu}.$$

Second, the procedure given in Section II-B can be adapted to the present case by considering that now $n[k]$ should be replaced by the sum of a Gaussian r.v. with variance σ_{r_n} and a r.v. uniform in the interval $(-(1-\nu)\Delta, (1-\nu)\Delta)$ and noting that, instead of L dimensions, we have M. As for the DC-DM case, an exact analysis of P_e is not possible, but the method detailed in Appendix A can also be used to yield an accurate upper bound. Unfortunately, the fact that the pdf of projected noise is now more involved makes the computations rather lengthy; therefore, they are not included here and will be published in preliminary form in [22]. With this theoretical upper bound, one can see that the optimal value of M is 1, meaning that the basic DC-QP gives the best performance in the class of generalized DC-QP schemes.

To illustrate this, in Fig. 7, we depict the probability of error P_e versus the parameter M for different WNRs, showing both the analytical bound (which is asymptotically tight) and the outcome of 8 × 10^6 Monte Carlo simulations. For each WNR and M pair, the distortion parameter ν is chosen to minimize P_e, and it is found by exhaustive search. The same tendency (increase in P_e with M) was perceived for many more WNRs tested by the authors and not shown here. A further justification for this behavior is afforded by working with the union bound and making the simplification that the projected-noise variance follows a Gaussian distribution (as is for instance done in [7]).The advantage of this formulation is that it allows us to analytically obtain the value of ν that minimizes the union bound (and that happens to be the same as that given by Costa) and then differentiate this bound in terms of the parameter M. Due to a lack of space, we do not show the calculations here, but note that for any fixed WNR, the union bound to P_e monotonically increases with M for $M > 0.65$, which, given the discrete nature of M, provides additional support to the hypothesis that $M = 1$ produces the minimum P_e.

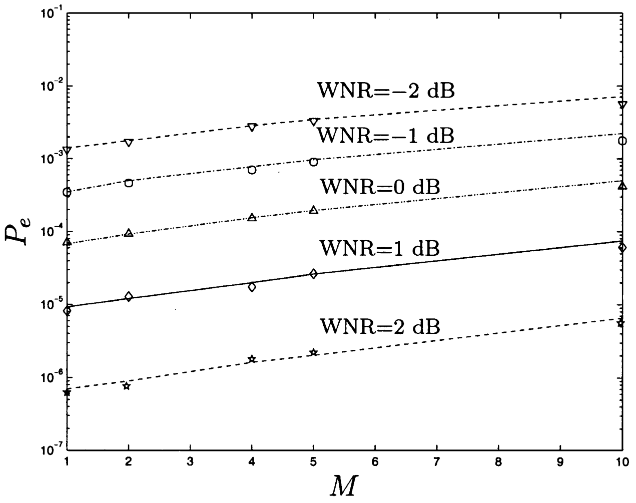

Fig. 7. Performance of generalized QP, using multidimensional DC-DM at the projection ($L = 20$). Lines and symbols stand for theoretical values and empirical data, respectively.

Therefore, we conclude that working with just one projected dimension is optimal for this family of data hiding methods and, as a corollary, that the basic DC-QP method *always* performs better than the multidimensional DC-DM scheme. These conclusions are in good agreement with the experimental results reported by Eggers and Girod in [28], who compared the SCS (equivalently, DC-DM) method with repetition coding, and the spread-transform SCS method, both of which hold close likeness with our two extremal cases (DC-DM and DC-QP, respectively).

It is also interesting to relate our results to the achievable rates computed in [28] for spread-transform-like methods. As shown there, for Costa's scheme, spreading can only reduce the achievable rate (the "critical slope" is never attained). The same happens with diversity in digital communications for Gaussian channels; after all, capacity formulas are identical in both cases. However, this does not mean that increasing the spreading factor L will ever increase P_e; on the contrary, as readily seen in (52), L should be made as large as possible, limited only by the payload to be hidden. Of course, with better coding schemes and/or constellations, it would be possible to fall closer to capacity than what is achieved through just spreading. Nevertheless, if coding is to be used, spreading may be useful for pushing up the operating WNR to values for which good codes are known. Interestingly, Eggers and Girod have shown in [28] that for negative WNRs, the spread-transform SCS method may increase the achievable rate by augmenting L, thus recovering part of the rate that is lost due to the suboptimality of this data hiding scheme. Since for a given WNR there is an optimal L beyond which the achievable rate decreases, this suggests concatenating diversity (to improve the WNR) and channel coding.

[3] Here, we assume that L/M takes an integer value, neglecting possible border effects. If the M projection vectors are allowed to overlap, as is done in [10], this problem is completely overcome, even though the analysis turns out to be more complex. In both cases, the results are equivalent.

D. Connections of QP With Costa's Result

We have seen in Section III-C that in Costa's encoding scheme, the ν parameter lets a part of the host signal $\mathbf{x}$ traverse the encoder and appear at the input. In fact, when $\nu = 0$, the whole host signal gets to the output. In the generalized QP, instead of quantizing $\mathbf{x}$, we alternatively quantize the projection of $\mathbf{x}$ into a subspace and leave the part of $\mathbf{x}$ orthogonal to that subspace intact.

Let us assume that the projection matrix $\mathbf{P}$ given in (54) has M orthonormal columns (note that this normalization does not have any impact on performance, although Δ has to be scaled accordingly). The watermark is computed as

$$\mathbf{w} = \mathbf{P}\left(\mathbf{Q}_b(\mathbf{P}^T\mathbf{x}) - \mathbf{P}^T\mathbf{x}\right)$$

where $\mathbf{Q}_b(\cdot)$ is an M-dimensional quantizer. The watermark can alternatively be expressed as

$$\mathbf{w} = \mathbf{u}_0 - \mathbf{P}\mathbf{P}^T\mathbf{x}$$

where $\mathbf{u}_0 = \mathbf{P}\,\mathbf{Q}_b(\mathbf{P}^T\mathbf{x})$ is an L-dimensional codeword, lying in the subspace spanned by $\{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$. Therefore, the signal received at the decoder is

$$\mathbf{z} = \mathbf{u}_0 + (\mathbf{I}_L - \mathbf{P}\mathbf{P}^T)\mathbf{x} + \mathbf{n}.$$

Now, it is easy to see that the role played by ν in Costa's random coding scheme is now played by the matrix $\mathbf{P}\mathbf{P}^T$. The role is similar in the sense that it performs a linear transformation on $\mathbf{x}$ with an energy reduction. The rest of the energy of $\mathbf{x}$, i.e., $(\mathbf{I}_L - \mathbf{P}\mathbf{P}^T)\mathbf{x}$, or the projection of $\mathbf{x}$ onto the subspace orthogonal to the subspace spanned by $\{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$, can be employed to compensate for distortions introduced by the quantizer. Recall that the $(1-\nu)\mathbf{x}$ term in Costa's scheme was used in a similar way.

In fact, for a given maximum average watermark distortion D_w, projecting into a subspace allows us to increase the size of the quantization cells. As a consequence, the minimum distance between quantization centroids associated with different messages can be increased. In other words, the performance against noise can be improved. One may compute the capacity of the QP method, assuming that $\mathbf{P}$ is a random matrix, to show that it is lower than that given by Costa.

VI. Experimental Results and Comparisons

In this section, the theoretical probabilities of error obtained in Sections III–V are compared and validated with empirical data generated through Monte Carlo simulations. In the plots, theoretical values are represented by lines, whereas empirical data are represented by symbols.

First, in Fig. 8, the P_e's for the unidimensional cases under uniform noise are compared. Both DC-DM and GDC-DM are presented for the respective optimized values $\nu = 0.5$ and $\sigma_\nu/\Delta = 0.3$ at WNR $= 0$ dB. This choice is reasonable, as the level of attacking distortion is unknown beforehand by the encoder.

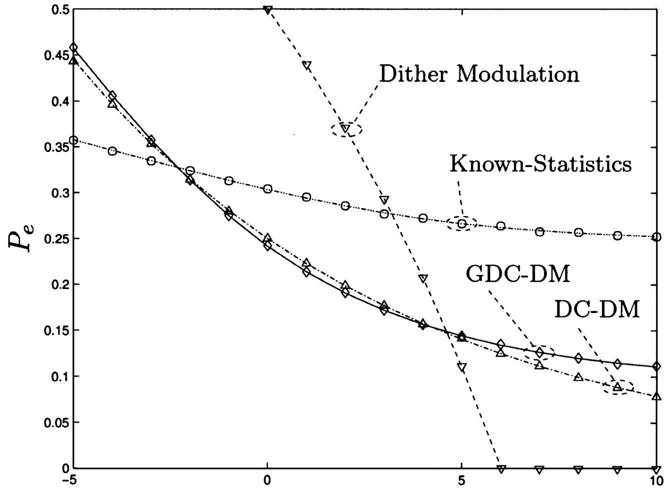

Fig. 8. Bit error probability versus WNR for the unidimensional case: uniform noise. DWR $=$ 6.0 dB. Lines and symbols stand for theoretical values and empirical data, respectively.

It is remarkable that the P_e of DM grows from 0 to 0.5 for values of WNR decreasing from 6 to 0 dB, and it worsens for negative vales. By contrast, DC-DM is much more well-behaved when presented with distortions of importance, even though it has a nonzero probability of error for ranges where DM is error-free. We may observe also that GDC-DM improves on the performance of DC-DM for large channel distortions, showing that the alleged optimality of the latter does not apply to this type of channel.

The statistical method in this case seems to perform worse than GDC-DM and DC-DM for a wide range of WNRs, but its robustness is noteworthy. The slope of its curve is the least steep of all the methods; the reason for this behavior was explained in Section IV. Noticeably, for WNR values lower than zero, the known-statistics method continues to improve its performance with respect to the known-state methods; for the depicted example, it performs better than all the other methods for WNR ≈ -2 dB and below.

It is also very important to recognize again that the plot for the known-host-statistics method is drawn for a DWR of 6 dB. Notably, lower values of DWR are advantageously exploited by this kind of procedures to diminish the probability of error, as illustrated for the particular case of Fig. 5; as we have shown in previous publications [29], one straightforward way of reducing the DWR consists in applying Wiener filtering to estimate the original host signal. Unfortunately, a rather small value of the DWR (0 dB) would be needed for this method to catch up with the best performing ones for WNR $= 0$ dB. By contrast, the P_e for the known-host-state methods remains constant for different DWR values as they are unaware of the statistical properties of the host signal used for the embedment. It must be stressed that further improvements of the known-host-statistics methods take place when a more accurate modeling of $\mathbf{x}$ is used. We recall that the Laplacian model is chosen here for representing a good compromise between accuracy and ease of analytical manipulation. More exact models such as the generalized Gaussian or other heavy tailed pdfs have been proposed to model DCT coefficients with outstanding results [18].

Fig. 9 shows the performance of the aforementioned methods for the Gaussian case. The P_e versus WNR curves corresponding to DC-DM and GDC-DM are depicted for their

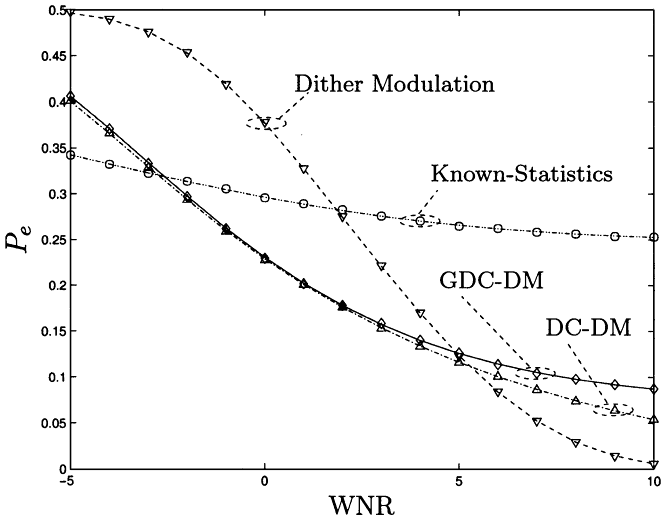

Fig. 9. Bit error probability versus WNR for the unidimensional case: Gaussian noise. DWR = 6.0 dB. Lines and symbols stand for theoretical values and empirical data, respectively.

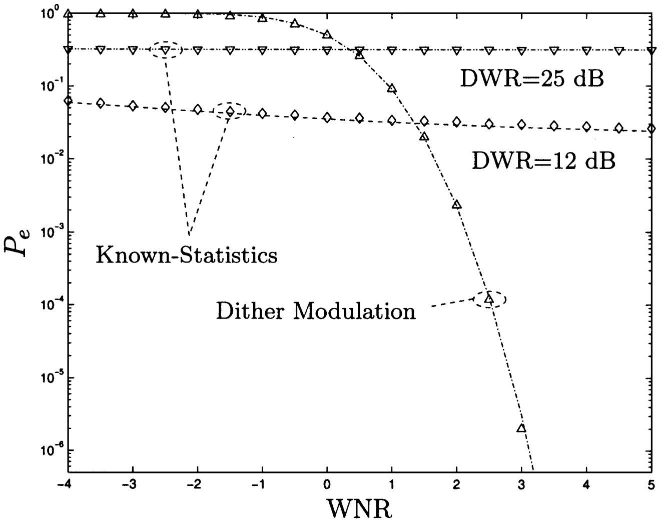

Fig. 10. Bit error probability versus WNR for the multidimensional case $(L = 40)$: uniform noise. Lines and symbols stand for theoretical values and empirical data, respectively.

respective optimum values $\nu = 0.53$ and $\sigma_\nu/\Delta = 0.275$ at WNR $= 0$ dB. In general, we can draw the same conclusions from these results as for the uniform noise case although, by comparison with Fig. 8, it becomes clear that Gaussian noise is not at all a worst-case attack for DM, as the largest distortion only causes now a P_e lower than 0.4. This reveals that analyses of DM under Gaussian channels may overstate its potential. On the other hand, now, the DC-DM method proves to perform better than GDC-DM, thus making evident that the Gaussian channel fits better to this method. The same reasoning above about the known-statistics method readily applies here.

Next, the multidimensional cases for DM and the known-host-statistics method, under uniform and Gaussian noise, respectively, are presented in Figs. 10 and 11. The performance of the known-host-statistics method has been evaluated for the particular case of $\lambda[k] = \lambda$ and a constant WNR to allow for a fair comparison with the known-host-state case that was analyzed under this assumption. The sensitivity to noise of the latter is evident from a glance at both plots; in addition, it becomes apparent, as in the unidimensional case, that the uniform channel is a worse attack than the Gaussian channel. Regarding the theoretical approximation used for predicting its performance, we can see that while the CLT approximation agrees almost perfectly with the empirical results in the uniform case, for the Gaussian channel, it cannot even upperbound the empirical values of P_e when the WNR grows beyond 3 dB. As explained in Section III-B, this effect is due to the slow convergence of the summation in (17) to a true Gaussian, considering that we are only using $L = 20$. Therefore, it is necessary to resort to the bound derived in the Appendix A, which we can see becomes a very tight approximation to P_e for large values of the WNR.

Last, note that the known-host-statistics method now remains almost invariable for the whole range of allowed distortions and, remarkably, for the two different noise sources applied. As in the unidimensional case, the method improves for lower DWRs by contrast with the invariability of the known-host-state one.

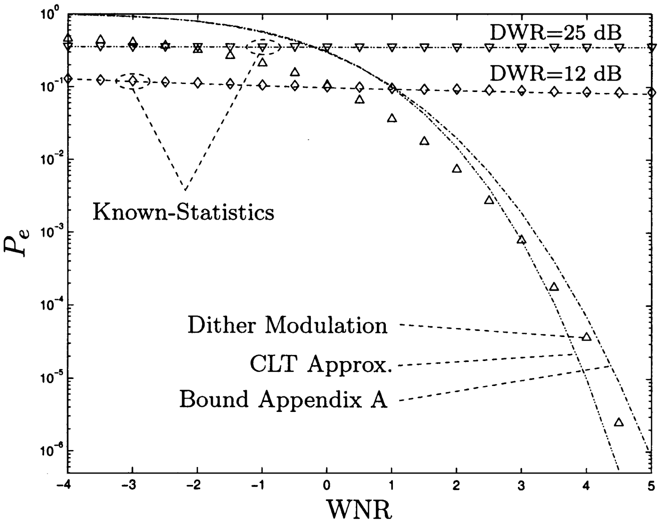

Fig. 11. Bit error probability versus WNR for the multidimensional case $(L = 20)$: Gaussian noise. Lines and symbols stand for theoretical values and empirical data, respectively.

In Fig. 12, the results for QP and DC-QP are presented. For both cases, the final performance is excellent, outdoing by several orders of magnitude in P_e the other multidimensional methods with the same L. Notice that although DC-QP offers an improvement over QP for every value of the WNR, the resulting gain is small. Furthermore, the performance curve for DC-QP is obtained under the assumption that the operating WNR is known at the embedder so that the optimal value of ν can be used. If, in a practical application, this knowledge is not available, then use of the uncompensated QP (i.e., $\nu = 1$) is recommended.

Next, in Fig. 13, the theoretical performance values of QP and DC-QP for different values of L are shown. The ν parameters are empirically obtained for optimizing DC-QP at each value of the WNR. We can see that impressively low probabilities of error are attainable. Observe that the predicted values are so low that its empirical simulation becomes difficult; the soundness of

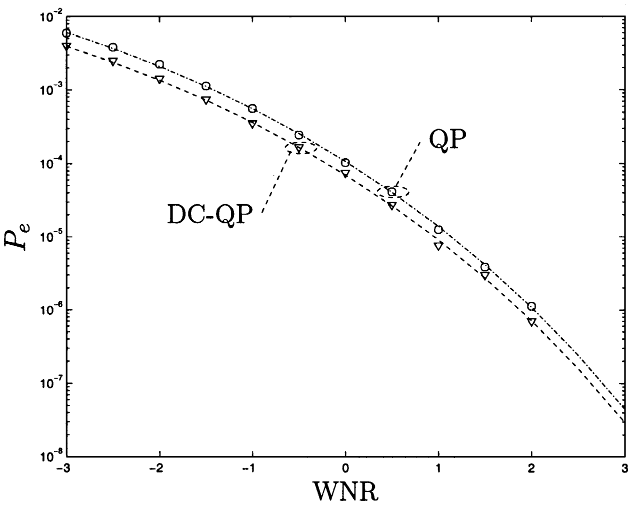

Fig. 12. Bit error probability versus WNR for QP $(L = 20)$: DWR $= 25.0$ dB. Lines and symbols stand for theoretical values and empirical data, respectively.

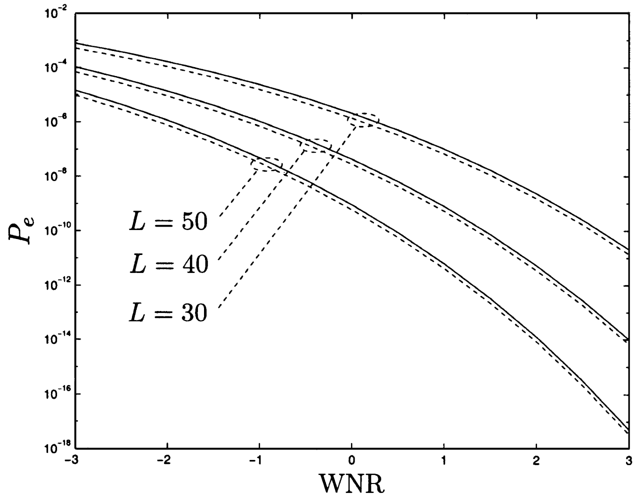

Fig. 13. Theoretical bit error probability versus WNR compared for QP (solid lines) and DC-QP (dashed lines) for increasing values of L: (DWR $= 25.0$ dB).

our results is supported by the empirical validation of Fig. 12. As a final observation, note that as depicted in Fig. 6, QP methods are also DWR-dependent, thus gathering the best properties exhibited by known-host-statistics and known-host-state methods.

Finally, in Fig. 14, we compare the performance of QP with that of STDM in a real case, using data from the well-known gray-scale *Lena* image (256 × 256). As QP does not need statistical modeling for the host signal, in this example, the data hiding process takes place in the spatial domain using a perceptual mask computed following [30] and applying no Wiener preprocessing, which leads to a very large DWR (approximately 40 dB). The distortion consists of independent additive Gaussian noise, with its variance shaped at each sample to match the perceptual mask. The orthonormal vector used for STDM spreading is just the perceptual mask multiplied by a zero-mean, unit-variance pseudorandom sequence and normalized in energy, i.e., $s[k]\alpha[k]/(\sum_{j\in\mathcal{S}}\alpha^2[j])$, $k \in \mathcal{S}$. The advantage of QP over STDM is attributable to the optimality of

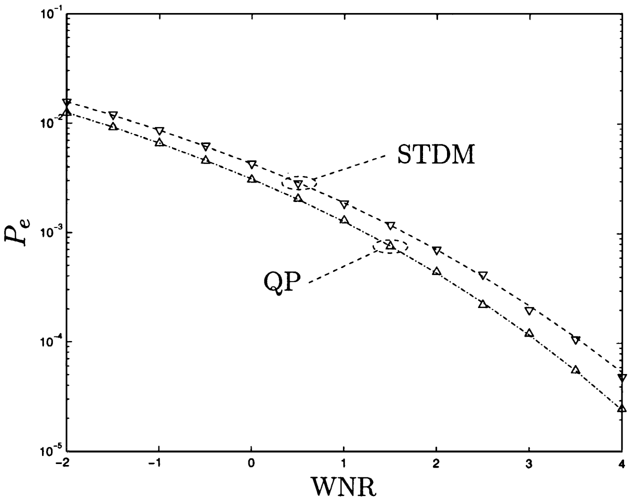

Fig. 14. Performance comparison between QP and STDM with real data (Lena image) and noise variance locally proportional to the perceptual mask $(L = 20)$. Lines and symbols stand for theoretical values and empirical data, respectively.

the correlation weights (i.e., $s[k]/\alpha[k]$) in (29). It is also worth seeing that the performance of QP in Fig. 14 is worse than in Fig. 12 for the same WNR, this being due to the larger DWR corresponding to the former.

VII. Conclusions

Throughout this paper, we have compared the performance of know-host-state and known-host-statistics methods. Even though the former offer strong host-signal rejection properties by disregarding its underlying features, we have seen that the host-signal statistical properties can play a significant role in the resulting performance when correctly taken into account. This fact is reflected in the improvement obtained by means of the QP method that we have proposed and analyzed. In all cases, we have given theoretical formulas that allow to assess and predict performance for different scenarios with a high degree of accuracy.

One important conclusion that must be drawn is that Gaussian channels and Gaussian host signal models do not offer in all cases upper bounds to performance measured in terms of the bit error probability. Apart from their excellent performance, one important advantage of the newly proposed QP techniques is that even for moderate values of L, they free us from the necessity of accurately modeling both the host signal and the channel. Of particular importance is the fact that QP becomes insensitive to the actual statistics of the channel noise, as long as the latter is independent from the host image.

Several extensions to the work presented here are currently being investigated. First, multidimensional QIM lattices other than the checkerboard one can be analyzed with the techniques given here; in this case, additional gains can be expected. Second, some more work is necessary in order to establish the performance of good channel codes within the data hiding framework, especially for known-host-state methods. Since the idea of distortion compensation already introduces significant gains, new coding schemes that are especially tailored to this scenario deserve further attention. Finally, improved linear and nonlinear projection functions for the QP method will be investigated; this relaxation already offers significant advantages whenever knowledge of the statistics of the host signal is available. Therefore, one might expect a similar behavior when dealing with QPs.

Appendix

A. Procedure for the Computation of the Upper Bound P_s in the Multidimensional DM Method With Gaussian Channel Noise

In this Appendix, we adapt and simplify the procedure presented in [24] for the sum of L Nakagami-m distributions. The numerical procedure given in [24] is valid for any integer m, whereas here, we need to deal with the sum of L Nakagami-$1/2$ distributions.

We want to compute

$$P_s = P\left\{\sum_{k\in\mathcal{S}} n'[k] > L\Delta/2\right\} \tag{55}$$

where $n'[k]$, $k \in \mathcal{S}$ is a one-sided Gaussian (Nakagami-$1/2$) distribution with pdf given by

$$f_{n'}(n'[k]) = \begin{cases} \dfrac{2}{\sqrt{2\pi}\,\sigma_g}\, e^{-(n'[k])^2/2\sigma_g^2}, & n'[k] > 0 \\ 0, & \text{otherwise.} \end{cases}$$

It is convenient to consider instead the normalized random variables $m[k] = n'[k]/\sigma_g$ so that (55) can be rewritten as

$$P_s = P\left\{\sum_{k\in\mathcal{S}} m[k] > L\xi\sqrt{3}/2\right\}$$

where now, $\xi = \Delta/(\sqrt{3}\sigma_g)$.

Let $M(\omega)$ denote the characteristic function of $m[k]$. Then, $M(\omega)$ turns out to be [31]

$$M(\omega) = \exp\left(-\frac{\omega^2}{2}\right) + j\omega\sqrt{\frac{2}{\pi}} \cdot \exp\left(-\frac{\omega^2}{2}\right)\Phi\left(\frac{1}{2}, \frac{3}{2}; \frac{\omega^2}{2}\right) \tag{56}$$

where $\Phi(\cdot, \cdot; \cdot)$ is the Kummer confluent hypergeometric function defined in [32].

Let $\omega_l = 2\pi l/T$ for any positive integer l and with T a sufficiently large real number. Then, following [25], P_s may be written as

$$P_s \approx \frac{1}{2} + \frac{2}{\pi}\sum_{\substack{l=1\\ l\ \text{odd}}}^{\infty} \frac{|M(\omega_l)|^L \sin(L\theta(\omega_l))}{l}$$

where $\theta(\omega)$ is defined as

$$\theta(\omega) \triangleq \arg\{M(\omega)\} - \frac{\omega\xi\sqrt{3}}{2}$$

and $\arg(x)$ denotes the four quadrant phase of the complex number x.

The series above is point-wise convergent with an accuracy that depends on the value of T. A greater accuracy is obtained for larger values of T, but this requires truncation to more terms in the series for a practical implementation. We have used a value of $T = 500$ and truncation to 1000 terms in our simulations. Regarding the evaluation of the confluent hypergeometric function that arises in (56), this can be done by means of another convergent series since [31]

$$\Phi\left(\frac{1}{2}, \frac{3}{2}; \omega^2\right) = 1 + \sum_{k=1}^{\infty} \frac{\omega^{2k}}{(2k+1)\cdot k!}.$$

B. Calculation of $E\{P[k]\}$ and Var$\{P[k]\}$ for the Known-Statistics Multidimensional Case

We assume that $x[k]$ and $s[k]$ are the only random variables in (25) and that the latter takes the values ±1 with equal probability. After integrating over $s[k]$, we get the following expressions for the summands:

$$\begin{aligned} E\{P[k]|x[k]\} &= \tfrac{1}{2}\Big(|x[k]| + 2\alpha[k] + \big||x[k]| - 2\alpha[k]\big|\Big) \\ &\quad - |x[k]| \\ \text{Var}\{P[k]|x[k]\} &= \tfrac{1}{4}\Big(|x[k]| + 2\alpha[k] - \big||x[k]| - 2\alpha[k]\big|\Big)^2. \end{aligned}$$

To ease the manipulation, these equations can be rewritten as

$$E\{P[k]|x[k]\} = \begin{cases} -|x[k]| + 2\alpha[k], & |x[k]| \le 2\alpha[k] \\ 0, & |x[k]| > 2\alpha[k], \end{cases} \tag{57}$$

$$\text{Var}\{P[k]|x[k]\} = \begin{cases} x^2[k], & |x[k]| \le 2\alpha[k] \\ 4\alpha^2[k], & |x[k]| > 2\alpha[k]. \end{cases} \tag{58}$$

Averaging over the pdf of $x[k]$, we have

$$E\{P[k]\} = E_{x[k]}\Big\{E\{P[k]|x[k]\}\Big\} \tag{59}$$

$$\begin{aligned} \text{Var}\{P[k]\} &= E_{x[k]}\Big\{\text{Var}\{P[k]|x[k]\}\Big\} \\ &\quad + \text{Var}_{x[k]}\Big\{E\{P[k]|x[k]\}\Big\} \\ &= E_{x[k]}\Big\{\text{Var}\{P[k]|x[k]\}\Big\} \\ &\quad + E_{x[k]}\Big\{E^2\{P[k]|x[k]\}\Big\} \\ &\quad - E^2_{x[k]}\Big\{E\{P[k]|x[k]\}\Big\}. \end{aligned} \tag{60}$$

Therefore, for the noiseless case, we have to compute the different expectations in (59) and (60) using the pdf of $x[k]$ given by (20) and the expressions (57) and (58), resulting in

$$\begin{aligned} E\{P[k]\} &= \frac{1}{\beta[k]}\left(e^{-2\beta[k]\alpha[k]} + 2\beta[k]\alpha[k] - 1\right) \\ \text{Var}\{P[k]\} &= \frac{1}{\beta^2[k]}\Big(3 - e^{-4\beta[k]\alpha[k]} \\ &\quad - 2\,e^{-2\beta[k]\alpha[k]}(1 + 4\beta[k]\alpha[k])\Big). \end{aligned}$$

For the calculation of the same expectations when uniform noise is added, we must use instead the pdf of $t[k] = x[k] + n[k]$ given by (23). As this pdf is symmetrical and has two parts, for the case $\eta[k] \le 2\,\alpha[k]$, we may divide now each integral in the interval $[0,\, 2\alpha[k])$ into two integrals in the intervals $[0,\, \eta[k])$ and $[\eta[k],\, 2\alpha[k])$, respectively. Thus, we can write $E\{P[k]\}$ as

$$\begin{aligned} E\{P[k]\} =& \, 2\int_0^{\eta[k]} (-t + 2\alpha[k])\, f_t(t|\, |t| < \eta[k])\, dt \\ &+ 2\int_{\eta[k]}^{2\alpha[k]} (-t + 2\alpha[k])\, f_t(t \,|\, |t| \geq \eta[k])\, dt \\ =& \, \frac{1}{2}(4\alpha[k] - \eta[k]) + \frac{1}{\beta^2[k]\eta[k]} \\ &\cdot \Big(\sinh(\beta[k]\eta[k])\, e^{-2\beta[k]\alpha[k]} + e^{-\beta[k]\eta[k]} - 1\Big). \end{aligned} \tag{61}$$

As for $\mathrm{Var}\{P[k]\}$, using also the same restriction on $\eta[k]$, we may summarize its calculation in the following expression:

$$\begin{aligned} \mathrm{Var}\{P[k]\} =& \, 2\int_0^{\eta[k]} t^2\, f_t(t \,|\, |t| < \eta[k])\, dt \\ &+ 2\int_{\eta[k]}^{2\alpha[k]} t^2\, f_t(t \,|\, |t| \geq \eta[k])\, dt \\ &+ 2\int_{2\alpha[k]}^{+\infty} 4\alpha^2[k]\, f_t(t \,|\, |t| \geq \eta[k])\, dt \\ &+ 2\int_0^{\eta[k]} (-t + 2\alpha[k])^2\, f_t(t \,|\, |t| < \eta[k])\, dt \\ &+ 2\int_{\eta[k]}^{2\alpha[k]} (-t + 2\alpha[k])^2\, f_t(t \,|\, |t| \geq \eta[k])\, dt \\ &- E^2\{P[k]\} \\ =& \, \frac{5\eta^2[k]}{12} + \frac{1}{\beta^2[k]}\, (e^{-\beta[k]\eta[k]} + 3) + \frac{1}{\beta^4[k]\eta^2[k]} \\ &\cdot \Big\{ e^{-2\alpha[k]\beta[k]} \Big(2\sinh(\beta[k]\eta[k]) + e^{-2\beta[k]\eta[k]} - 1\Big) \\ &\quad - 2\, e^{-\beta[k]\eta[k]} \Big(\cosh(\beta[k]\eta[k]) - 1\Big) \\ &\quad - \tfrac{1}{2}\, e^{-4\alpha[k]\beta[k]} \Big(\cosh(2\beta[k]\eta[k]) - 1\Big) \Big\} \\ &+ \left(\frac{\eta[k] - 8\alpha[k]}{\beta^2[k]\eta[k]} - \frac{4}{\beta^3[k]\eta[k]} \right) \\ &\cdot e^{-2\alpha[k]\beta[k]} \sinh(\beta[k]\eta[k]). \end{aligned} \tag{62}$$

The statistics in (61) and (62) can be similarly obtained for the case $\eta[k] > 2\,\alpha[k]$.

ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. J. J. Eggers for kindly reading and discussing the manuscript and providing valuable suggestions, P. Comesaña for helping with some simulations, and the anonymous reviewers for their constructive comments.

REFERENCES

[1] R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, “A digital watermark,” in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, 1994, pp. 86–89.

[2] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, “Techniques for data hiding,” *IBM Syst. J.*, vol. 35, no. 3/4, pp. 313–336, 1996.

[3] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, “Secure spread spectrum watermarking for multimedia,” *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.

[4] I. J. Cox, M. L. Miller, and A. L. McKellips, “Watermarking as communications with side information,” *Proc. IEEE*, vol. 87, pp. 1127–1141, July 1999.

[5] M. H. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. IT–29, pp. 439–441, May 1983.

[6] B. Chen and G. W. Wornell, “Preprocessed and postprocessed quantization index modulation methods for digital watermarking,” *Proc. SPIE, Security Watermarking Multimedia Contents*, pp. 48–59, Jan. 2000.

[7] J. J. Eggers, J. K. Su, and B. Girod, “A blind watermarking scheme based on structured codebooks,” in *Proc. Inst. Elect. Eng. Conf. Secure Images Image Authentication*, London, U.K., Apr. 2000.

[8] M. Ramkumar, “Data hiding in multimedia: Theory and applications,” Ph.D. dissertation, New Jersey Inst. Technol., Newark, NJ, Jan. 2000.

[9] J. R. Hernández and F. Pérez-González, “Statistical analysis of watermarking schemes for copyright protection of images,” *Proc. IEEE*, vol. 87, Special Issue on Identification and Protection of Multimedia Information, pp. 1142–1166, July 1999.

[10] B. Chen and G. W. Wornell, “Quantization index modulation: A class of provably good methods for digital watermarking and information embedding,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 1423–1443, May 2001.

[11] F. Pérez-González, J. R. Hernández, and F. Balado, “Approaching the capacity limit in image watermarking: A perspective on coding techniques for data hiding applications,” *Signal Process.*, vol. 81, Special Section on Information Theoretic Aspects of Digital Watermarking, pp. 1215–1238, June 2001.

[12] P. Moulin and M. Mıhçak, “A framework of evaluating the data-hiding capacity of image sources,” *IEEE Trans. Image Processing*, vol. 11, pp. 1029–1042, Sept. 2002.

[13] R. B. Wolfgang, C. I. Podilchuk, and E. J. Delp, “Perceptual watermarks for digital images and video,” *Proc. IEEE*, vol. 87, pp. 1108–1125, July 1999.

[14] S. Voloshynovskiy, A. Herrigel, N. Baumgärtner, and T. Pun, “A stochastic approach to content adaptive digital image watermarking,” in *Proc. 3rd Int. Workshop Inform. Hiding* Dresden, Germany, Oct. 1999.

[15] J. J. Eggers, R. Bäuml, and B. Girod, “Digital watermarking facing attacks by amplitude scaling and additive white noise,” in *Proc. 4th Int. ITG Conf. Source Channel Coding*, Berlin, Germany, Jan. 2002.

[16] P. Moulin and J. O’Sullivan, “Information-theoretic analysis of information hiding,” Univ. Illinois, Urbana, IL, Dec. 2001, to be published.

[17] L. Schuchman, “Dither signals and their effect on quantization noise,” *IEEE Trans. Commun. Technol.*, vol. CT–12, pp. 162–165, Dec. 1964.

[18] 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. Image Processing*, vol. 9, Special Issue on Image and Video Processing for Digital Libraries, pp. 55–68, Jan. 2000.

[19] B. Chen and G. W. Wornell, “Provably robust digital watermarking,” *Proc. SPIE*, vol. 3845 of Multimedia Syst. Applications II, pp. 43–54, 1999.

[20] ——, “Dither modulation: A new approach to digital watermarking and information embedding,” *Proc. SPIE*, vol. 3657 of Security Watermarking Multimedia Contents, pp. 342–353, 1999.

[21] J. Conway and N. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed. New York: Springer, 1999, vol. 290 of Comprehensive Studies Math..

[22] 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. 5th Int. Workshop Inform. Hiding*, ser. Lecture Notes in Computer Science Noorwijkerhout, The Netherlands, Oct. 2002.

[23] K. Bury, *Statistical Models in Applied Science*. Malabar, FL: Krieger, 1975.

[24] N. C. Beaulieu and A. A. Abu-Dayya, “Analysis of equal gain diversity on Nakagami fading channels,” *IEEE Trans. Commun.*, vol. 39, pp. 225–234, Feb. 1991.

[25] 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, pp. 1463–1474, Sept. 1990.

[26] M. Ramkumar, A. Akansu, and X. Cai, “Floating signal constellations for multimedia steganography,” in *Proc. IEEE Int. Commun. Conf.*, New Orleans, LA, June 2000, pp. 249–253.

[27] M. R. Schroeder, *Number Theory in Science and Communication*, 2nd ed. New York: Springer-Verlag, 1990, Springer Series Inform. Sci..

[28] J. J. Eggers and B. Girod, *Informed Watermarking*. Boston, MA: Kluwer, 2002.

[29] J. R. Hernández, F. Pérez-González, J. M. Rodríguez, and G. Nieto, “Performance analysis of a 2D-multipulse amplitude modulation scheme for data hiding and watermarking of still images,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 510–524, May 1998.

[30] J. R. Hernández, F. Pérez-González, and J. M. Rodríguez, “Coding and synchronization: A boost and a bottleneck for the development of image watermarking,” in *Proc. COST #254 Int. Workshop Intell. Commun.* L’Aquila, Italy, June 1998, pp. 77–82.

[31] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. San Diego, CA: Academic, 1994.

[32] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. Washington, DC: Nat. Bureau Stand., 1972.

Fernando Pérez-González (M’90) received the telecommunications engineer degree from the University of Santiago, Santiago, Spain, in 1990 and the Ph.D. degree from the University of Vigo, Vigo, Spain, in 1993, also in telecommunications engineering.

He joined the faculty of the School of Telecommunications Engineering, University of Vigo, as an Assistant Professor in 1990 and is currently a Professor with the same institution. He has visited the University of New Mexico, Albuquerque, for different periods spanning ten months. His research interests lie in the areas of digital communications, adaptive algorithms, and digital watermarking. He has been the manager of a number of projects concerned with digital television and radio, both for satellite and terrestrial broadcasting. He is coeditor of the book *Intelligent Methods in Signal Processing and Communications* (Boston, MA: Birkhauser, 1997), has been Guest Editor of two special sections of the EURASIP journal *Signal Processing* devoted to signal processing for communications, and has co-edited a Feature Topic of the IEEE COMMUNICATIONS MAGAZINE devoted to digital watermarking.

Dr. Pérez-González was the Chairman of the Fifth Baiona Workshop on Emerging Technologies in Telecommunications, held in Baiona, Spain, in 1999 and will chair the Sixth workshop, which will take place September 2003.

Félix Balado received the telecommunications engineering degree from the University of Vigo, Vigo, Spain, in 1996, where he is currently pursuing the Ph.D. degree in the field of watermarking and data hiding.

He is currently an Associate Researcher at the University of Vigo.

Juan R. Hernández Martín (M’99) received the Ingeniero de Telecomunicación degree from the University of Vigo, Vigo, Spain, in 1993, the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 1996, and the Ph.D. degree in telecommunications engineering from the University of Vigo, in 1998.

Until 1999, he was with the Department of Communication Technologies, University of Vigo, where he worked on digital communication, digital television, conditional access systems, and digital watermarking. Since 1999, he has worked for Ascom AG and Lysis SA, Lausanne, Switzerland on quality of service for VoIP applications and on digital television systems. He is currently working for Nagravision SA, Lausanne.

Dr. Hernández Martín’s Ph.D. thesis on watermarking techniques for the copyright protection of digital images received an extraordinary award from the University of Vigo and the best thesis on electronic commerce award by the Colegio Oficial de Ingenieros de Telecomunicación of Spain.