Rational Dither Modulation: A High-Rate Data-Hiding Method Invariant to Gain Attacks

Fernando Pérez-González, Carlos Mosquera, Mauro Barni, *Member, IEEE*, and Andrea Abrardo

***Abstract*—A novel quantization-based data-hiding method, called Rational Dither Modulation (RDM), is presented. This method retains most of the simplicity of the conventional dither modulation (DM) scheme, which is largely vulnerable to amplitude scalings and modifies it in such a way that the result becomes invariant to gain attacks. RDM is based on using a gain-invariant adaptive quantization step-size at both embedder and decoder. This causes the watermarked signal to be asymptotically stationary. Mathematical tools are used to determine the stationary probability density function, which is later used to assess the performance of RDM in Gaussian channels. It is also shown that by increasing the memory of the system, it is possible to asymptotically approach the performance of DM, still keeping invariance against gain attacks. RDM is compared with improved spread-spectrum methods, showing that the former can achieve much higher rates for the same bit error probability. Experimental results confirm the validity of the theoretical analyses given in the paper. Finally, a broader class of methods, that extends gain-invariance to quantization index modulation (QIM) methods, is also presented.**

***Index Terms*—Data hiding, dither modulation, quantization index modulation.**

I. Introduction

SINCE the pioneering papers by Cox *et al.* [1] and Chen and Wornell [2], [3] pointed out the close relationship between digital watermarking and communication over a channel with side information at the encoder [4], [5], watermarking research has been focused on the development of practical schemes, turning into reality the expectations raised by the theoretical analysis, that is, fulfilling the possibility of designing a watermarking system that completely rejects the interference between the host signal and the hidden message, while at the same time retaining the good robustness features of classical spread spectrum systems. In this framework, great attention has been given to the class of Quantization Index Modulation (QIM) algorithms [3], according to which watermark embedding is obtained by quantizing the host feature sequence with a quantizer chosen among a set of quantizers that are each associated with a different message. The core problem of QIM is the design of the codebooks of the quantizers used to embed the watermark. The simplest solution consists of the adoption of lattice-based quantizers [6] for which efficient embedding and decoding algorithms exist. For example, adopting a set of scalar and uniform quantizers results in the popular Dither Modulation (DM) scheme and its Distortion Compensated version (DC-DM) [3], [7]. As a matter of fact, no performance loss has to be expected by resorting to a lattice-based codebook since it has been recently shown that channel capacity can be achieved by means of QIM watermarking with lattice codebooks [8].

The main weakness of lattice-based watermarking is its vulnerability against the gain attack, consisting of the multiplication of the host feature sequence by a gain factor ρ, which is unknown to the decoder. If ρ is constant throughout the feature sequence, we obtain the so-called Fixed Gain Attack (FGA), which is the main focus of this paper. Weakness against FGA is a serious drawback, e.g., with respect to classical spread spectrum methods, since in many cases, multiplication by a constant factor does not introduce any annoying artifacts, yet it results in a dramatic performance degradation [9].

The solutions proposed so far to cope with the gain attack, in the framework of QIM watermarking, can be grouped into three main categories: i) embedding of an auxiliary pilot signal to be used by the decoder to recover from amplitude scaling [10][1]; ii) adoption of spherical codewords [6] together with correlation decoding [12], [13]; and iii) definition of an embedding domain that is invariant to value-metric scaling. Insertion of a pilot signal is the most popular solution; however, it poses several problems, since embedding this signal, which is deterministically known to both transmitting and receiving ends, reduces the available payload and introduces an additional source of weakness against malicious attacks because attackers can either decide to attack the watermark or the pilot signal. The use of spherical codes to cope with FGA relies on the observation that using a minimum distance decoder on a set of codewords lying on the surface of a sphere ensures that multiplication by a constant factor does not move a point from one decoding region to the other. The problem with spherical codes is that watermark embedding and recovery get very complicated, thus losing the simplicity of lattice-based watermarking. Recent attempts to develop a simple watermarking scheme relying on the properties of orthogonal spherical codes have given good results; however, such results have been obtained at the expense of watermark

Manuscript received June 30, 2004; revised October 2, 2004. This work was supported in part by *Xunta de Galicia* under projects PGIDT04 TIC322013PR (SEQURDOCS) and PGIDT04 PXIC32202PM; MEC project DIPSTICK, reference TEC2004-02551/TCM; FIS project IM3, reference G03/185, and European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The associate editor coordinating the review of this manuscript and approving it for publication was Guest Editor Dr. Ingemar J. Cox.

F. Pérez-González and C. Mosquera are with the Department Teoría de la Señal y Comunicaciones, ETSI Telecom., Universidad de Vigo, 36200 Vigo, Spain (e-mail: fperez@gts.tsc.uvigo.es; mosquera@gts.tsc.uvigo.es).

M. Barni and A. Abrardo are with the Department of Information Engineering, University of Siena, 53100, Siena, Italy (e-mail: barni@dii.unisi.it; abrardo@dii.unisi.it).

Digital Object Identifier 10.1109/TSP.2005.855407

[1]In principle ρ could also be estimated blindly, as is done, for instance, in [11]; however, in this case, the accuracy of the estimate gets worse.

payload [13]. The third possibility requires that a standard lattice-based algorithm is applied in a domain that is invariant with respect to the presence of the gain ρ. Watermarking in a gain-invariant domain seems to be an ideal solution, even if finding such a domain while at the same time retaining a high capacity turned out to be a difficult task. This is exactly the scope of this paper: to present a new watermarking method that is invariant against FGA (and still holds some robustness against varying gain attacks), which at the same time retains the effectiveness and simplicity of conventional DM schemes.

In the basic form of our new scheme, invariance against FGA is obtained by applying conventional DM to the ratio of the current host sample and the previously generated watermarked sample. This simple approach is enough to ensure invariance against FGA. As to robustness against additive white Gaussian noise (AWGN), we show how the performance of our basic method can be significantly improved by watermarking the ratio between the current feature sample and a proper function of the L previously watermarked samples. By recognizing that the main idea behind the new scheme consists of the choice of watermarking a ratio-like function instead of the features themselves, we refer to the new scheme as Rational Dither Modulation (RDM).

The bulk of this paper is devoted to the theoretical analysis of the performance of RDM in terms of embedding distortion and bit error rate (BER) in the presence of FGA and AWGN attacks of given strength. When needed, we will use numerical simulations to support the theoretical analysis. Despite some analytical difficulties that are overcome by resorting to techniques that are new to data-hiding theory, we are able to prove several important results, showing the great potential of RDM. Among them, the single most important result is that by making the denominator of the to-be-watermarked ratio depend on an increasing number of previous samples, i.e., by increasing the memory L of the system, RDM approaches the performance of conventional DM, with the great advantage of insensitivity to FGA, and with a complexity increase that is only linear with L.

In the last part of the paper, we introduce a further generalization of RDM: to propose the class of Gain Invariant QIM (GI-QIM) watermarking schemes. We expect that new important results will derive from this analysis. For sake of brevity, we only analyze a nondistortion-compensated (non-DC), uncoded version of RDM; however, it is expected that further important improvements can be obtained in this way.

The paper is organized as follows: In Section II, we formulate the fixed gain attack and introduce some important notation; the RDM method is presented in Section III, where some statistical properties of the watermarked signal are also discussed. In Section IV, we improve the basic RDM scheme by introducing memory and a function that controls the step size. Section V is focused on analyzing the performance of the proposed data-hiding methods, while Section VI is devoted to presenting the results of numerical simulations that validate and complete our analysis. Section VII presents a generalization of RDM that effectively extends the gain-invariance idea to QIM methods. Finally, our main conclusions are summarized in Section VIII.

II. Problem Formulation

We assume that the host coefficients are arranged in a one-dimensional (1-D) vector $\mathbf{x}$. For 2-D hosts, this can be, for instance, done by means of lexicographical or zig-zag ordering. Throughout the text, we will assume that $\mathbf{x}$ contains N elements, i.e., $\mathbf{x} = (x_1, x_2, \ldots, x_N)^T$, where x_k refers to the kth element, and T denotes transpose. However, in those cases, where an asymptotic analysis is pursued, we will let $N \to \infty$.

In data-hiding applications, a message m is embedded into $\mathbf{x}$ by modifying the latter in some way. We will call the result a *watermarked signal* and denote it by vector $\mathbf{y}$. Then, the difference vector $\mathbf{w} \triangleq \mathbf{y} - \mathbf{x}$ is called the *watermark*. In a similar fashion to the host signal, y_k and w_k denote the kth element of $\mathbf{y}$ and $\mathbf{w}$, respectively. In this paper, we will assume that $\mathbf{x}, \mathbf{y}, \mathbf{w}$ are real vectors, which is enough to describe most domains used in practice. In any case, the extension of our results to complex vectors requires minimum changes.

Although the feasibility of the methods that will be proposed does not depend on host-statistics, for analytical purposes, it will be convenient to regard the host samples x_k as being generated according to some random variable X_k with probability density function (pdf) $f_{X_k}(x_k)$. We will further assume that $X_k, k = 1, \ldots, N$, are independent and identically distributed (i.i.d.) with zero-mean and variance σ_x^2. Throughout the text, we will use uppercase letters to denote random variables and lowercase letters to denote specific values. Thus, Y_k and W_k will be random variables modeling the watermarked signal and the watermark samples, respectively, and whose distribution will be thoroughly analyzed for the method proposed in this paper. We will also let $\mathbf{X} \triangleq (X_1, \ldots, X_N)^T$ be the random vector modeling the host signal; similar definitions follow for $\mathbf{Y}$ and $\mathbf{W}$. Given a random variable X, M_{xp} will denote its pth absolute moment, i.e., $M_{xp} \triangleq E\{|X|^p\}$.

The *fixed-gain attack* (FGA) consists of a constant scaling of the amplitudes of the watermarked signal. Furthermore, we will assume that zero-mean additive white noise $\mathbf{N}$ with variance σ_n^2 and independent of Y is also added by the attacker. This noise will allow us to compare the proposed schemes with standard methods that are nonrobust against FGA. Let $\rho > 0$ denote the gain parameter; then, following our model, the attacked vector $\mathbf{Z}$ will be written as

$$\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N}). \tag{1}$$

Then, the observed vector $\mathbf{z} = (z_1, z_2, \ldots, z_N)^T$ will be such that $z_k = \rho(y_k + n_k), k = 1, \ldots, N$, where n_k contains the noise samples. See Fig. 1. Since most of the methods that will be introduced need some initial state to be agreed upon by the embedder and the decoder, we will henceforth assume, without loss of generality, that $y_k = z_k = 1$, for all $k < 1$.

Several definitions will be also needed. The *embedding distortion* D_w is defined as the average power of the watermark, i.e.,

$$D_w \triangleq \frac{1}{N}\mathrm{E}[\|\mathbf{Y} - \mathbf{X}\|^2] = \frac{1}{N}\mathrm{E}[\|\mathbf{W}\|^2]$$

where $\mathrm{E}[\cdot]$ denotes statistical expectation, and $\|\cdot\|$ stands for Euclidean (i.e., ℓ_2) norm.

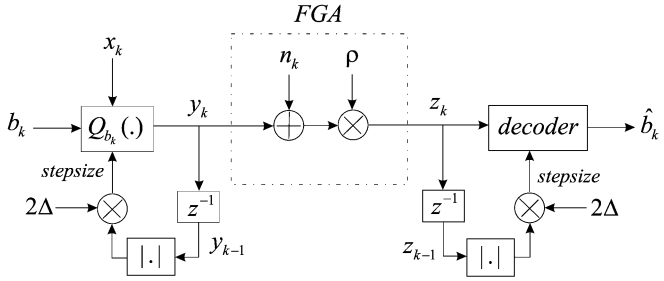

Fig. 1. Block-diagram of first-order RDM.

The *attacking distortion* D_c is defined as follows:

$$D_c \triangleq \frac{1}{N}\mathrm{E}[\|\rho^{-1}\mathbf{Z}-\mathbf{Y}\|^2] = \frac{1}{N}\mathrm{E}[\|\mathbf{N}\|^2] = \sigma_n^2. \quad (2)$$

The rationale behind premultiplying $\mathbf{Z}$ by ρ^{-1} in (2) is to make the attacking distortion independent of ρ. If gain attacks are considered here as being imperceptible, then it is reasonable that they be disregarded when measuring the attacking distortion. In other words, variations in ρ should leave D_c unchanged.

It is pertinent to remark here that it would be also possible to describe the FGA by the model $\mathbf{Z} = \rho\mathbf{Y} + \mathbf{N}$, as it has been done sometimes in the literature. In this case, in order to achieve an attacking distortion measure that is only noise-dependent, it would be reasonable to define D_c as $\mathrm{E}[\|\mathbf{Z}-\rho\mathbf{Y}\|^2/N]$, which would again yield $D_c = \sigma_n^2$. In any event, we point out that the model $\mathbf{Z} = \rho\mathbf{Y} + \mathbf{N}$ can be straightforwardly transformed into (1) and vice versa by properly scaling the noise variance. We will also find it useful to define some quantities that relate the powers of the host, the watermark, and the noise. The *Document-to-Watermark Ratio* (DWR) is given by σ_x^2/D_w; the *Watermark-to-Noise Ratio* (WNR) is D_w/D_c, and finally, the *Document-to-Noise Ratio* (DNR) is σ_x^2/D_c. Often, these quantities will be expressed in decibels.

Finally, in the foregoing sections, we will use the set $\mathcal{G}$ of functions $g : \mathbb{R}^L \to \mathbb{R}, L \geq 1$, which have the property that

$$g(\rho\mathbf{y}) = \rho g(\mathbf{y}), \quad \text{for all } \rho > 0, \quad \mathbf{y} \in \mathbb{R}^L. \quad (3)$$

III. First-Order RDM

In this section, we will introduce the simplest form of RDM, which makes use of only one past output sample. The objective is to show the basic properties of this new data-hiding scheme and reveal the difficulties of a rigorous statistical analysis. With this twofold aim, we will further assume that a binary modulation is employed; this means that one bit will be hidden into each host sample. A generalization with multilevel constellations is also possible but will not be pursued here.

In order to fully understand the advantages brought about by binary RDM, it is convenient to recall the basic construction of DM, which was proposed by Chen and Wornell in [3]. The procedure can be summarized by defining the shifted lattices

$$\Lambda_b \triangleq 2\Delta\mathbb{Z} - b\Delta/2, \quad b = -1, 1 \quad (4)$$

which describe the centroids for the respective quantizers $Q_{-1}(\cdot)$ and $Q_1(\cdot)$, here assumed to be based on Euclidean distances. Given the kth information symbol $b_k \in \{-1, 1\}$, embedding in the kth sample is performed using the rule $y_k = Q_{b_k}(x_k), k = 1, \ldots, N$.

Having observed the possibly attacked sample z_k, the kth decoded bit $\hat{b}_k$ is decided according to a minimum Euclidean distance rule

$$\hat{b}_k = \arg\min_{-1,1} |z_k - Q_{b_k}(z_k)|, \quad k = 1, \ldots, N \quad (5)$$

which can be also seen to be equivalent to quantizing z_k with a quantizer with step-size Δ.

Now, consider an FGA with no additive noise; therefore, the vector at the input of the decoder can be written as $\mathbf{z} = \rho\mathbf{y}$, which is equivalent to scaling the output of the embedder by ρ. Unfortunately, the quantization bins at the decoder are not scaled accordingly, thus producing a mismatch between embedder and decoder that dramatically affects performance, even in the absence of attacking noise. In fact, substituting $z_k = \rho y_k$ into (5), it is easy to see that the decoded bit depends on ρ. However, a crucial observation is that if the quantizers were such that $Q_b(\rho y_k) = \rho Q_b(y_k), b = -1, 1$, then it would be possible to factor out ρ of the absolute value in (5), and hence, the decoder output would become invariant with ρ. Unluckily, the quantizers in (4) do not hold this property. Identical conclusions can be drawn when the attack also encompasses noise addition.

Inspired by the preceding discussion, we propose a simple modification of the basic DM idea, in which the quantization step is made dependent in a causal way on the watermarked samples, as illustrated in Fig. 1. A broader class of schemes following the same paradigm is briefly described in Section VII. The current method is summarized by the following embedding and decoding rules:

$$y_k = |y_{k-1}| Q_{b_k}\left(\frac{x_k}{|y_{k-1}|}\right) \quad (6)$$

$$\hat{b}_k = \arg\min_{-1,1} \left| \frac{z_k}{|z_{k-1}|} - Q_{b_k}\left(\frac{z_k}{|z_{k-1}|}\right) \right| \quad (7)$$

for $k = 1, \ldots, N$, where, as noted in Section II, it is assumed without loss of generality that $y_0 = z_0 = 1$.

Taking into account that the proposed method relies on computing the ratio of two quantities, and given the fact that it resembles differential modulations used in communications (where a subtraction is used instead of a ratio), we have named it *Rational Dither Modulation*. It is interesting to notice that the implementation of the generic RDM amounts to small modifications to the DM scheme; in embedding, it is necessary to divide x_k by $|y_{k-1}|$ prior to performing the quantization, whereas in decoding, the divisor becomes $|z_{k-1}|$. This is due to the unavailability of y_{k-1} at decoding so that z_{k-1} becomes an estimate.

Now, it is possible to see why RDM is insensitive to gain attacks: Substituting $z_k = \rho(y_k + n_k)$ and $z_{k-1} = \rho(y_{k-1} + n_{k-1})$ into (7), it can be readily verified that ρ cancels out in the expression, and consequently, the decision $\hat{b}_k$ does not depend on ρ. Despite the similarities between RDM and DM, it is obvious that they make use of distinct decision variables; therefore, the expected performance of the two schemes will be different. This motivates the need for a careful analysis of RDM, which we address in the next sections, except for the BER computation, which is postponed for Section V.

A. Embedding Distortion

Instead of seeing the embedding problem as quantizing $x_k/|y_{k-1}|$ with a quantizer with step 2Δ, for the purposes of this section, it is more convenient to equivalently consider x_k as being quantized with a variable step quantizer with step-size $2\Delta|y_{k-1}|$. In addition, even though the previous discussion applies to both deterministic and stochastic signals, here, we deal with the latter because we are mainly interested in computing statistical quantities. Then, if we write

$$Y_k = X_k + W_k \tag{8}$$

it is clear that, in a first approximation, for large DWRs X_k can be considered as having an almost flat pdf within each quantization bin so that W_k is uniformly distributed, i.e., $W_k \sim U[-\Delta|y_{k-1}|, \Delta|y_{k-1}|)$. Moreover, from this same assumption, it is easy to show that $E\{X_k W_k\} = 0$ so that X_k and W_k are uncorrelated. Then, from (8), we can write

$$\mathrm{Var}\{Y_k\} \approx \sigma_x^2 + \mathrm{Var}\{W_k\} = \sigma_x^2 + \frac{\Delta^2}{3}\mathrm{Var}\{Y_{k-1}\}. \tag{9}$$

An immediate conclusion of (9) is that the process $\{Y_k\}$ is nonstationary (even in the wide sense). Expression (9) describes a difference equation for the variances; therefore, it is interesting to analyze the steady-state, that is, $k \to \infty$. It is easy to show that under the flat assumption, in steady-state

$$\sigma_y^2 \approx \frac{\sigma_x^2}{1 - \Delta^2/3} \tag{10}$$

where $\sigma_y^2 \triangleq \lim_{k\to\infty} \mathrm{Var}\{Y_k\}$. Equation (10) is valid only for $\Delta < \sqrt{3}$; otherwise, there would be no convergence. Defining σ_w^2 as the steady-state variance for the watermark and using (10) together with (9) yields $\sigma_w^2 \approx \sigma_x^2/(3/\Delta^2 - 1)$. Note that by carefully choosing $\Delta \geq 0$, it is possible to attain any desired level of steady-state variance σ_w^2. In practice, it has been observed that the settling time is small; therefore, it is reasonable to consider that most watermark samples can be modeled by a stationary process with variance σ_w^2, allowing us to write the embedding distortion $D_w \approx \sigma_w^2$.

We remark that the results given in this section are based on the assumption of W_k being uniformly distributed and, thus, are good approximations for small Δ (i.e., large DWR). This means, for instance, that the previous condition on Δ for the convergence of the variance is not exact. In Section III-C, a more rigorous development leads to the condition $\Delta < 2$. In any case, it is worth mentioning that numerical simulations for DWRs larger than 10 dB show that the approximation to σ_w^2 given in the previous paragraph differs from the true value by less than 1%.

B. Derivation of the Stationary Probability Density Function

One relevant question regarding RDM is what is the distribution of the watermarked signal. Recall that for embedding symbol b_k at the kth sample, the set of possible centroids is $2\Delta|y_{k-1}|\mathbb{Z} + b_k\Delta|y_{k-1}|/2$; therefore, it is clear that the BER will depend on the pdf of Y_{k-1}. In fact, since we are quantizing to a discrete set of centroids, it turns out that the pdf of Y_k conditioned on a transmitted symbol b_k and on $Y_{k-1} = y_{k-1}$ is discrete and has the form

$$\begin{aligned} f_{Y_k \,|\, Y_{k-1}, B_k}(y_k \,|\, y_{k-1}, b_k) &= \sum_m p_m(y_{k-1}, b_k)\delta(y_k - \Delta|y_{k-1}|(2m + b_k/2)) \\ &= \sum_m \frac{p_m(y_{k-1}, b_k)}{|2m + b_k/2|\Delta}\delta\left(|y_{k-1}| - \frac{y_k}{(2m + b_k/2)\Delta}\right) \end{aligned} \tag{11}$$

where δ denotes Dirac's delta, and $p_m(y_{k-1}, b_k)$ is the probability that X_k takes its value in the interval $[\Delta|y_{k-1}|(2m - 1 + b_k/2), \Delta|y_{k-1}|(2m + 1 + b_k/2))$.

If the conditions for the convergence of Y_k are met ($\Delta < 2$; see Section III-C), then, there will exist $Y = \lim_{k\to\infty}\{Y_k\}$, with pdf $f_Y(y)$.[2] Next, we derive an equilibrium equation for this pdf. Assuming that B_k takes the values ± 1 with equal probability, it follows that

$$\begin{aligned} f_{Y_k}(y_k) &= \frac{1}{2}\int_{-\infty}^{\infty} f_{Y_k \,|\, Y_{k-1}, B_k}(y_k \,|\, y_{k-1}, b_k = -1) f_{Y_{k-1}}(y_{k-1})\, dy_{k-1} \\ &\quad + \frac{1}{2}\int_{-\infty}^{\infty} f_{Y_k \,|\, Y_{k-1}, B_k}(y_k \,|\, y_{k-1}, b_k = +1) \\ &\quad \times f_{Y_{k-1}}(y_{k-1})\, dy_{k-1}. \end{aligned} \tag{12}$$

If the pdf of the host $f_X(x)$ is symmetric about the origin, then it is not difficult to show that $p_m(y, +1) = p_{-m}(y, -1) \triangleq p_m(y)$, with

$$p_m(y) = \int_{\frac{\Delta}{2}|y|(4m-1)}^{\frac{\Delta}{2}|y|(4m+3)} f_X(x)\, dx \tag{13}$$

and that $p_m(y) = p_m(-y)$.

Consider now both integrals in (12), which we will denote by I_1 and I_2, respectively. Substituting (11) into (12) and performing the integration for the case $y_k \geq 0$, we obtain

$$\begin{aligned} I_1 &= \frac{1}{2}\sum_{m=1}^{\infty} \frac{1}{|2m - 1/2|\Delta} p_m\left(\frac{y_k}{(2m - 1/2)\Delta}, -1\right) \\ &\quad \times f_{Y_{k-1}}\left(\frac{y_k}{(2m - 1/2)\Delta}\right) \\ &\quad + \frac{1}{2}\sum_{m=1}^{\infty} \frac{1}{|2m - 1/2|\Delta} \cdot p_m\left(\frac{-y_k}{(2m - 1/2)\Delta}, -1\right) \\ &\quad \times f_{Y_{k-1}}\left(\frac{-y_k}{(2m - 1/2)\Delta}\right), \quad y_k \geq 0 \end{aligned} \tag{14}$$

and

$$\begin{aligned} I_2 &= \frac{1}{2}\sum_{m=0}^{\infty} \frac{1}{|2m + 1/2|\Delta} p_m\left(\frac{y_k}{(2m + 1/2)\Delta}, 1\right) \\ &\quad \times f_{Y_{k-1}}\left(\frac{y_k}{(2m + 1/2)\Delta}\right) \\ &\quad + \frac{1}{2}\sum_{m=0}^{\infty} \frac{1}{|2m + 1/2|\Delta} \cdot p_m\left(\frac{-y_k}{(2m + 1/2)\Delta}, 1\right) \\ &\quad \times f_{Y_{k-1}}\left(\frac{-y_k}{(2m + 1/2)\Delta}\right), \quad y_k \geq 0. \end{aligned} \tag{15}$$

[2]Here, as in the rest of the paper, we consider that *for sequences of random variables*, convergence implies *convergence in distribution*.

For $y_k < 0$, the results are identical to (14) and (15), with the sums ranging from $m = -\infty$ to $m = 0$ in I_1 and from $m = -\infty$ to $m = -1$ in I_2, respectively. Combining the results of the two integrals in (12) and the properties above, and after some straightforward algebra, we arrive at the following expression for $f_{Y_k}(y_k)$:

$$\begin{aligned} f_{Y_k}(y_k) &= \frac{1}{2}\sum_{m=-\infty}^{\infty}\frac{1}{|2m+1/2|\Delta}p_m\left(\frac{y_k}{(2m+1/2)\Delta}\right) \\ &\quad\times f_{Y_{k-1}}\left(\frac{-y_k}{(2m+1/2)\Delta}\right) \\ &\quad+\frac{1}{2}\sum_{m=-\infty}^{\infty}\frac{1}{|2m+1/2|\Delta}\cdot p_m\left(\frac{y_k}{(2m+1/2)\Delta}\right) \\ &\quad\times f_{Y_{k-1}}\left(\frac{y_k}{(2m+1/2)\Delta}\right). \end{aligned} \tag{16}$$

From (16), it is possible to write the equilibrium equation that the stationary distribution must satisfy (see Section III-C), by simply replacing $f_{Y_k}(\cdot)$ and $f_{Y_{k-1}}(\cdot)$ by $f_Y(\cdot)$. An immediate consequence of such an equilibrium equation is that $f_Y(y) = f_Y(-y)$. Thus, it can be written in a compact form as

$$\begin{aligned} f_Y(y) = \sum_m \frac{1}{|2m+1/2|\Delta}p_m\left(\frac{y}{(2m+1/2)\Delta}\right) \\ \times f_Y\left(\frac{|y|}{|2m+1/2|\Delta}\right). \end{aligned} \tag{17}$$

Noticing that for all $m \in \mathbb{Z}$, $x \in \mathbb{R}, p_m(x) \geq 0$ and that for at least one $m_0, p_{m_0}(x) \neq 0$, it is possible to conclude from (17) that $f_Y(0) = 0$.

Equation (17) obviously is specialized for any particular host distribution. The case of a Gaussian pdf is discussed in Appendix A, where it is shown that in such case the stationary pdf of Y_k can be well approximated by the following mixture:

$$\begin{aligned} f_Y(y) \approx \frac{4|y|}{\pi\sigma_y^2\Delta}\sum_{m=0}^{\infty}\frac{1}{(2m+1)^2} \\ \times\exp\left(-\frac{2y^2}{\sigma_y^2\Delta^2(2m+1)^2}-\frac{y^2}{2\sigma_y^2}\right) \end{aligned} \tag{18}$$

where σ_y^2 is approximately given by (10). The method given in Appendix A can be readily extended to other host pdfs. In Section IV, we compare the analytical pdf with measured distributions for the first-order case.

C. Stationarity of the Probability Density Function

Let us consider the set of host samples X_k as i.i.d. random variables as above. Thus, the quantized values Y_k are a sequence of random variables, which can be considered as a Markov process in which the probability distribution of Y_k is a function solely of X_k, B_k, and Y_{k-1}. The quantization rule (6) is such that Y_k can take any of the following values: $\{\pm|y_{k-1}|(\Delta/2), \pm|y_{k-1}|(3\Delta/2), \pm|y_{k-1}|(5\Delta/2), \pm|y_{k-1}|(7\Delta/2), \ldots\}$, depending on the host value x_k and the bit b_k. In fact, the dependence of y_k on the initial value y_0 can be traced back as follows:

$$y_k = y_0\left(\frac{\Delta}{2}\right)^k(2m+1) \tag{19}$$

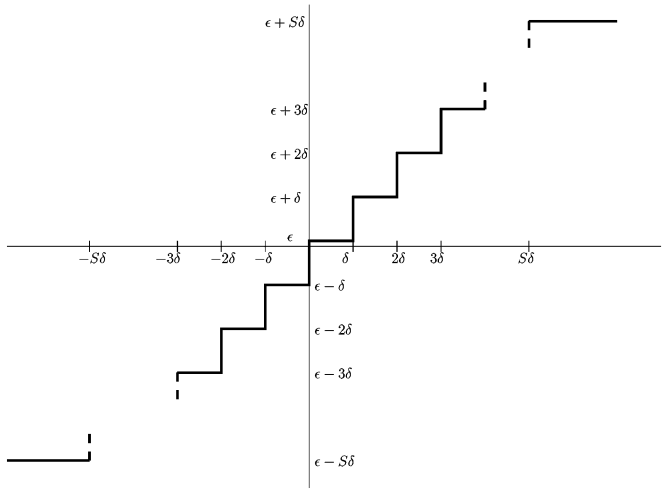

Fig. 2. Quantization function $Q_\epsilon(\cdot)$ for the Markov chain.

for some integer m. It is easy to conclude that the Markov chain that characterizes the evolution of y_k has an infinite number of states given by the previous set. This set is countable for a fixed y_0 and yields a discretization as dense as desired of the real line, provided that $\Delta < 2$; under this restriction, which necessarily applies throughout the entire paper, and for k and m large enough in (19), we can approximate any real value with arbitrary accuracy. In the case that the host samples X_k follow a finite support distribution, for example, X_k being a uniform random variable, it can be readily seen that the pdf of the output signal Y_k will be also restricted to approximately lie within the same support. In such a case, only the states with values y_k contained in this support must be considered.

For the sake of the analysis, we will consider that the output values are uniformly quantized, slightly modifying the original embedding rule:

$$\hat{y}_k = Q_\delta\left(|\hat{y}_{k-1}|Q_{b_k}\left(\frac{x_k}{|\hat{y}_{k-1}|}\right)\right), \quad k > 1 \tag{20}$$

where $\hat{y}_k$ denotes the quantization of y_k with a uniform quantizer with output values $l\delta + \epsilon$, corresponding to the quantization intervals $[l\delta, (l+1)\delta), -S < l < S$ (see Fig. 2) and output values $-S\delta + \epsilon$ and $S\delta + \epsilon$ for $(-\infty, -S\delta]$ and $[S\delta, \infty)$, respectively. This is a quantizer with saturation, for which S can be chosen arbitrarily large. The offset ϵ is a positive value such that $\epsilon < \delta$. Noticeably, this offset does not impose any restriction to our analysis since the quantization step δ can be made arbitrarily small. Furthermore, this corresponds to the actual implementation of a real system, for which the possible values depend on the number of available bits. It can be readily seen that this new setting evolves also as a Markov chain with a finite number of states. In Appendix B, we prove that, for $\Delta < 2$, this Markov chain is *irreducible* and *aperiodic*. In other words, all states are reachable from any given state in finite time, in such a way that movement through states is aperiodic. As a consequence, there exists a unique stationary probability distribution describing the occupation probabilities of all the states. Moreover, this stationary distribution happens to be the limiting distribution, regardless of the initial state, due to the aperiodic character of the chain [14].

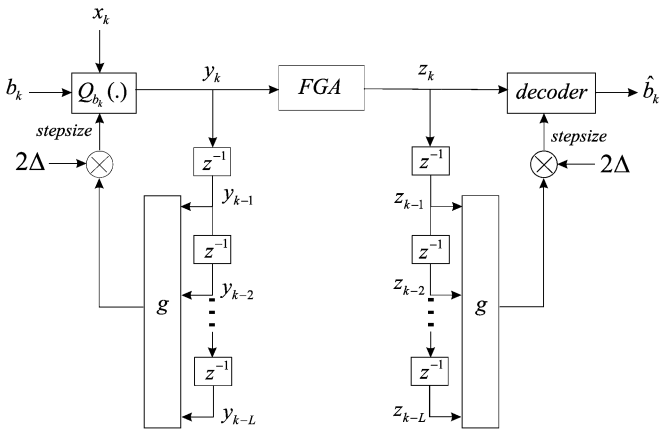

Fig. 3. Block-diagram of Lth-order RDM.

Similar conclusions were obtained by Gersho [15], who studied adaptive quantizers in a source coding setting. The assumptions made there are, however, different from those used in our work, and therefore, the methodology that we have followed also differs.

IV. General-Order RDM

One of the drawbacks of the RDM scheme is the large influence that the attacking noise has in the decoding quantization step-size, which will become manifest when we analyze its performance in Section V. This can be largely alleviated by constructing a more robust quantization step, which can be achieved by using some suitable function of several past samples, as we discuss in the following. Again, we focus on the binary scalar case, noticing that a generalization to multilevel constellations is also possible. The complete embedding and decoding process is represented in Fig. 3.

Let $\mathbf{y}_{k-1} \triangleq (y_{k-1}, y_{k-2}, \ldots, y_{k-L})^T$, with a similar definition for $\mathbf{z}_{k-1}$. Then, in Lth–order RDM, the kth binary information symbol $b_k \in \{\pm 1\}$ is embedded as

$$y_k = g(\mathbf{y}_{k-1}) Q_{b_k}\left(\frac{x_k}{g(\mathbf{y}_{k-1})}\right) \tag{21}$$

where the quantizers $Q_{-1}(\cdot)$ and $Q_1(\cdot)$ are, respectively, induced by the shifted lattices Λ_{-1} and Λ_1 in (4), and $g \in \mathcal{G}$, i.e., g is a function satisfying property (3). Then, given z_k and $\mathbf{z}_{k-1}$, decoding is carried out by following a generalization of (7), that is

$$\hat{b}_k = \arg\min_{-1,1} \left| \frac{z_k}{g(\mathbf{z}_{k-1})} - Q_{b_k}\left(\frac{z_k}{g(\mathbf{z}_{k-1})}\right) \right|. \tag{22}$$

Again, it can be shown that the decoder output is invariant to fixed gain attacks: If $\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N})$, it follows that $z_k = \rho(y_k + n_k)$ and $g(\mathbf{z}_{k-1}) = g(\rho(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})) = \rho g(\mathbf{y}_{k-1} + \mathbf{n}_{k-1})$, where the last equality follows from $g \in \mathcal{G}$. Now, it is easy to see that the ρ in the numerator and the denominator cancel out and the result is identical, irrespective of the value of ρ.

From (21) and (22), it is clear that the function $g \in \mathcal{G}$ plays a significant role in the definition of this data-hiding scheme. The problem of choosing a particular function is very involved due to the intrinsic nonlinearity of the quantization process; this suggests approaching the problem by considering simple elements of $\mathcal{G}$. One such subset is that based on Hölder or ℓ_p vector-norms:

$$g(\mathbf{y}_{k-1}) = \left(\frac{1}{L}\sum_{m=k-L}^{k-1} |y_m|^p\right)^{1/p}, \quad p \geq 1 \tag{23}$$

which can be extended to consider weighted-norms. The selection of parameter p is tackled in Section V-C. The case described in Section III corresponds to $L = 1$ in (23). As we will soon corroborate, for large L, $g(\mathbf{y}_{k-1})$ and $g(\mathbf{z}_{k-1})$ become very close, and thus, the embedding and decoding step-sizes are almost identical, with the consequent improvement in performance.

For the theoretical developments of the foregoing sections, we need to define the random vectors $\mathbf{Y}_k = (Y_k, \ldots, Y_{k-L+1})^T$ and $\hat{\mathbf{Y}} = \lim_{k\to\infty}\{\mathbf{Y}_k\}, \hat{\mathbf{Y}} \in \mathbb{R}^L$, provided that this limit exists. Similar definitions follow for $\mathbf{Z}_k$ and $\tilde{\mathbf{Z}}$. However, we will keep using both Y and Z to denote the corresponding random variables for the case $L = 1$ for coherence reasons with Section III.

A. Derivation of the Stationary Probability Density Function

The stationary statistical characterization of the watermarked signal for an arbitrary order can be considered as a generalization of the expression (16). Similar considerations to those made in Section III-C can be invoked to prove for $\Delta < 2$ the existence of a unique asymptotic distribution by just considering vectors $\hat{\mathbf{y}}_k \triangleq (\hat{y}_k, \ldots, \hat{y}_{k-L+1})^T$ as the states of the Markov chain, with

$$\hat{y}_k = Q_\delta\left(g(\hat{\mathbf{y}}_{k-1}) Q_{b_k}\left(\frac{x_k}{g(\hat{\mathbf{y}}_{k-1})}\right)\right), \quad k > 1.$$

It can be shown that the implicit relation that defines this distribution now becomes

$$f_Y(y) = \frac{1}{2}\sum_{m=-\infty}^{\infty} \frac{1}{|2m + 1/2|\Delta} p_m\left(\frac{y}{(2m+1/2)\Delta}\right) \times f_{g(\tilde{\mathbf{Y}})}\left(\frac{|y|}{|2m+1/2|\Delta}\right) \tag{24}$$

where $p_m(y)$ is defined in (13), and $f_{g(\tilde{\mathbf{Y}})}(s)$ denotes the pdf of $g(\tilde{\mathbf{Y}})$. This expression is not amenable to analysis for an arbitrary setting, although for large L, further developments are possible, as we show next. First, since for large DWRs the absolute moments of Y_k and X_k will be very similar (see below), it is reasonable to assume that $E\{|Y_k|^p\} < \infty$, provided that the pth absolute moment of X_k is bounded. Then, if the latter condition on Y_k holds, and $\{Y_k\}$ converges to Y, it follows that $E\{|Y_k|^p\} \to E\{|Y|^p\}$ as $k \to \infty$; see [16, pp. 251–252]. Thus, defining $R_k \triangleq g^p(\mathbf{Y}_{k-1})$, we can conclude that

$$E\{R_k\} \to E\{|Y|^p\} \triangleq M_{yp},$$
$$\mathrm{Var}\{R_k\} \to \frac{1}{L}E\{|Y|^{2p}\} - \frac{1}{L}E^2\{|Y|^p\} \triangleq \sigma_r^2.$$

Additionally, since $g^p(\cdot)$ is continuous, and $\{Y_k\}$ converges, it is possible to affirm that $\{R_k\}$ also converges to, say, R. On the other hand, if the DWR is large and the components of $\{X_k\}$

are independent, we can also consider the components of $\{Y_k\}$ as approximately independent; therefore, the vector $\tilde{\mathbf{Y}}$ will have approximately i.i.d. components. Consequently, for large L, we can invoke the Central Limit Theorem (CLT) to state that R will be well-approximated by a Gaussian with mean M_{yp} and variance σ_r^2.

Finally, from this pdf and the relation $g(\mathbf{Y}_{k-1}) = |R_k|^{1/p}$, it is clear that $\{g(\mathbf{Y}_{k-1})\} \rightarrow |R|^{1/p} = g(\tilde{\mathbf{Y}})$, and hence, it is possible to approximate $f_{g(\tilde{\mathbf{Y}})}(s)$ for large L as follows:

$$f_{g(\tilde{\mathbf{Y}})}(s) \approx \frac{ps^{p-1}}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(s^p - M_{yp})^2}{2\sigma_r^2}\right), \quad s \geq 0. \quad (25)$$

One can see in (25) how $g(\tilde{\mathbf{Y}})$ is concentrated around $M_{yp}^{1/p}$. Interestingly, for L becoming very large, the pdf $f_{g(\tilde{\mathbf{Y}})}(s)$ approaches a delta function, namely

$$f_{g(\tilde{\mathbf{Y}})}(s) \approx \delta\left(s - M_{yp}^{1/p}\right)$$

and $f_Y(y)$ in (24) can be simplified as

$$\begin{aligned} f_Y(y) &\approx \frac{1}{2}\sum_{m=-\infty}^{\infty} \frac{1}{|2m+1/2|\Delta} p_m\left(\frac{y}{(2m+1/2)\Delta}\right) \\ &\quad \times \delta\left(\frac{|y|}{|2m+1/2|\Delta} - M_{yp}^{1/p}\right) \\ &= \frac{1}{2}\sum_{m=-\infty}^{\infty} p_m\left(\frac{y}{(2m+1/2)\Delta}\right) \\ &\quad \cdot \delta\left(|y| - |2m+1/2|\Delta M_{yp}^{1/p}\right). \end{aligned}$$

The appropriate manipulation of the absolute values within the delta functions yields

$$\begin{aligned} f_Y(y) &\approx \frac{1}{2}\sum_{m=-\infty}^{\infty} p_m\left(M_{yp}^{1/p}\right)\left(\delta\left(y - (2m+1/2)\Delta M_{yp}^{1/p}\right)\right. \\ &\quad \left. + \delta\left(y + (2m+1/2)\Delta M_{yp}^{1/p}\right)\right) \\ &= \frac{1}{2}\sum_{m=-\infty}^{\infty} p_m\left(M_{yp}^{1/p}\right)\delta\left(y - (4m+1)\frac{\Delta}{2}M_{yp}^{1/p}\right) \\ &\quad + \frac{1}{2}\sum_{m=-\infty}^{\infty} p_{-m}\left(M_{yp}^{1/p}\right)\delta\left(y - (4m-1)\frac{\Delta}{2}M_{yp}^{1/p}\right). \end{aligned}$$

Let us now define

$$q_m \triangleq \frac{1}{2}\int_{\frac{\Delta}{2}M_{xp}^{1/p}(2m-1)}^{\frac{\Delta}{2}M_{xp}^{1/p}(2m+3)} f_X(x)\,dx \quad (26)$$

such that $\sum_{m=-\infty}^{\infty} q_m = 1$. Note that q_m has been written as a function of $M_{xp}^{1/p}$. For large L, it can be analytically justified that $M_{yp} \approx M_{xp}$: a claim whose validity is supported by simulations for all L, especially for small values of p and Δ. For $p = 2$, the latter approximation is tantamount to saying that $\sigma_y^2 \approx \sigma_x^2$, which is in good agreement with (10) for small Δ or, equivalently, for large DWR values. Thus, from (13), it follows that $p_m(M_{yp}^{1/p})/2 \approx q_{2m}$ and $p_{-m}(M_{yp}^{1/p})/2 \approx q_{2m-1}$, and as a consequence

$$\begin{aligned} f_Y(y) &\approx \sum_{m=-\infty}^{\infty} q_{2m}\delta\left(y - (4m+1)\frac{\Delta}{2}M_{xp}^{1/p}\right) \\ &\quad + \sum_{m=-\infty}^{\infty} q_{2m-1}\delta\left(y - (4m-1)\frac{\Delta}{2}M_{xp}^{1/p}\right) \\ &= \sum_{m\ \text{even}} q_m\delta\left(y - (2m+1)\frac{\Delta}{2}M_{xp}^{1/p}\right) \\ &\quad + \sum_{m\ \text{odd}} q_m\delta\left(y - (2m+1)\frac{\Delta}{2}M_{xp}^{1/p}\right). \end{aligned}$$

Finally, the pdf of Y is given for large L by

$$f_Y(y) \approx \sum_{m=-\infty}^{\infty} q_m\delta\left(y - (2m+1)\frac{\Delta}{2}M_{xp}^{1/p}\right), \quad L \gg 1. \quad (27)$$

Fig. 4 illustrates the statistical distribution of the watermarked signal, showing the remarkable accuracy in the predicted probability density functions. For $L = 1$, we have used the results of Section III-B, whereas for the other cases, we have resorted to the large L approximation by inserting (25) in (24). For $L = 500$, the pdf clearly resembles that of a DM signal (i.e., the distribution of a uniformly quantized signal); see (27).

Noticeably, the preceding analysis also serves to establish the embedding distortion for large L and high DWRs, as we discuss next. First, from (27), it is clear that for large L, RDM behaves as if Y were obtained by quantizing X with a quantizer with step-size $2\Delta M_{xp}^{1/p}$. Then, assuming a large DWR, which leads to X having an almost flat pdf within each quantization bin, we can easily write $D_w \approx (\Delta^2 M_{xp}^{2/p}/3)$. This result is for small values of p and for moderate and large values of L in good agreement with simulations and can be used to establish the operating DWR, which becomes

$$\text{DWR} = \frac{3\sigma_x^2}{\Delta^2 M_{xp}^{2/p}}. \quad (28)$$

Note that for $p = 2$ and a Gaussian host, the DWR is approximately $3/\Delta^2$ (as long as this DWR is large enough to justify our assumptions). Equation (28) can be inverted to determine Δ for a desired level of DWR. Moreover, for large WNRs and large L, $M_{zp} \approx M_{xp}$, these calculations can be repeated at the decoder to determine the value of Δ with a small discrepancy if the target DWR is known at both embedder and decoder. As we have just seen, this discrepancy does not even exist for Gaussian hosts.

V. Analytical Derivation of the Bit Error Rate

As discussed in Section II, we assume that the watermarked signal $\mathbf{Y}$ is sent through an FGA channel, producing a vector $\mathbf{Z} = \rho(\mathbf{Y} + \mathbf{N})$. As noted previously, RDM is invariant to gain attacks; therefore, the BER analysis can be carried out by setting $\rho = 1$.

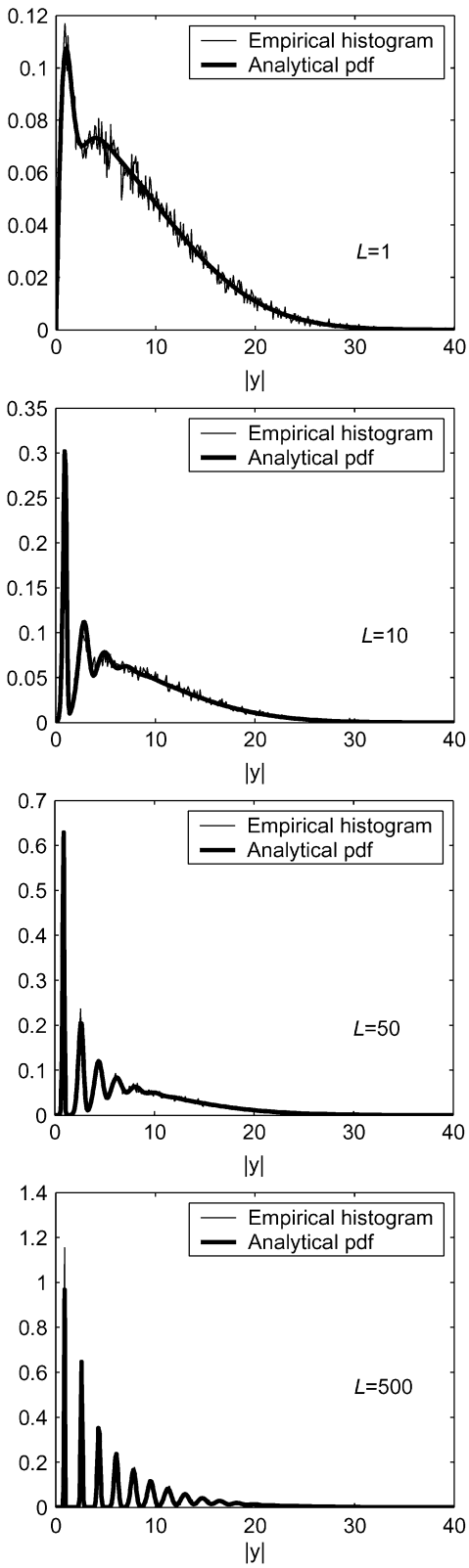

Fig. 4. Analytical pdf versus one-realization histogram for a Gaussian host DWR $= 20$ dB, $\sigma_x^2 = 100, p = 2$.

As decoding errors will occur at the same rate for $b_k = 1$ and $b_k = -1$, we will compute the probability of decoding $\hat{b}_k = -1$ when $b_k = 1$. Let us define $P_e[s, y]$ as the probability of $Z = Y + N$ falling in the set of intervals $\bigcup_{l=-\infty}^{\infty}[(2l+1)\Delta s, (2l+2)\Delta s)$, when $Y = y$:

$$P_e[s,y] \triangleq \sum_{l=-\infty}^{\infty} \int_{(2l+1)\Delta s}^{(2l+2)\Delta s} f_N(z-y)\, dz$$

where $f_N(n)$ is the pdf of the additive noise. As opposed to DM, RDM does not use a fixed discrete grid due to the variable step-size,[3] and the evaluation of the probability of error is more involved. In order to determine the bit error probability P_e, we must take the expectation of $P_e[s, y]$ with respect to the joint pdf of $g(\tilde{\mathbf{Z}})$ and Y so that

$$P_e = \int_{-\infty}^{\infty} \int_{0}^{\infty} P_e[s,y] f_{g(\tilde{\mathbf{Z}}),Y}(s,y)\, ds\, dy$$

which is equivalent to

$$P_e = \int_0^{\infty} f_{g(\tilde{\mathbf{Z}})}(s) \int_{-\infty}^{\infty} P_e[s,y] f_{Y|g(\tilde{\mathbf{Z}})}(y \,|\, s)\, dy\, ds. \quad (29)$$

Next, we make specific refinements of the above expression for those cases for which there exist closed-form statistical characterizations of the signals involved, that is, $L = 1$ and large L. These can be regarded as extreme cases, insomuch as performance will always improve as L increases, asymptotically approaching that of DM.

A. Large L

For large L, and due to the low variance of $g(\tilde{\mathbf{Y}})$, and correspondingly of $g(\tilde{\mathbf{Z}})$, $f_{Y|g(\tilde{\mathbf{Z}})}(y \,|\, s)$ can be safely approximated by $f_{Y|g(\tilde{\mathbf{Y}})}(y \,|\, s)$, i.e., the difference in the quantization steps at the embedder and the decoder will have a small impact in the final result; in such a case, we have that

$$f_{Y|g(\tilde{\mathbf{Y}})}(y \,|\, s) = \sum_{m=-\infty}^{\infty} p_m(s) \delta\left(y - \frac{4m+1}{2}\Delta s\right)$$

leading to

$$P_e = \int_0^{\infty} f_{g(\tilde{\mathbf{Z}})}(s) \sum_{m=-\infty}^{\infty} p_m(s) \times P_e\left[s, \frac{4m+1}{2}\Delta s\right] ds, \quad L \gg 1$$

with

$$f_{g(\tilde{\mathbf{Z}})}(s) \approx f_{g(\tilde{\mathbf{Y}})}(s) \approx \frac{p s^{p-1}}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{(s^p - M_{yp})^2}{2\sigma_r^2}\right), \quad s \geq 0. \quad (30)$$

This approximation is supported by the results in Appendix C for large DNR. Recognizing that the summation in (30) is nothing but the probability of error of DM for a step-size $2\Delta s$,

[3] In the limit, for $L = \infty$, the embedding quantization step is fixed and equal to $2\Delta M_{yp}^{1/p}$.

here denoted by $P_{\mathrm{DM}}(2\Delta s)$, P_e can be rewritten in a more compact form as follows:

$$P_e = \int_0^{\infty} f_{g(\tilde{\mathbf{Z}})}(s) P_{\mathrm{DM}}(2\Delta s)\, ds. \tag{31}$$

In [17], the probability of error of DM was shown to be

$$P_{\mathrm{DM}}(2\Delta s) = \sum_{l=-\infty}^{\infty} \int_{\Delta s(2l+1/2)}^{\Delta s(2l+3/2)} f_N(n)\, dn$$

which can be put in closed-form when N follows either a Gaussian or a uniform distribution [17]. Therefore, from (31), it is possible to conclude that for large L, the performance of RDM is equivalent to averaging that of DM for all the possible values of the step size. In particular, if $L \to \infty$, we have that $f_{g(\tilde{\mathbf{Z}})}(s) \to \delta(s - M_{yp}^{1/p})$, and $P_e \to P_{\mathrm{DM}}(2\Delta M_{yp}^{1/p})$.

B. $L = 1$

For low and moderate memory sizes L, and in particular, for $L = 1$, the previous assumption that $f_{Y\,|\,g(\tilde{\mathbf{Z}})}(y\,|\,s) \approx f_{Y\,|\,g(\tilde{\mathbf{Y}})}(y\,|\,s)$ no longer holds due to the difference between the quantization steps at embedding and decoding. As a consequence, we must handle $f_{Y\,|\,g(\tilde{\mathbf{Z}})}(y\,|\,g(\tilde{\mathbf{z}}))$ without neglecting the effect of the "jitter" associated with the decoder step-size estimation. To keep the analysis simple, but at the same time trying to preserve the differences in the step-sizes caused by the additive noise, we make the following first-order approximation:

$$g(z) = |z| = |y + n| \approx |y| + n.$$

This approximation turns out to work quite well for practical predictions and shows that the attacking additive noise leaks completely into the receiver step-size estimation. This effect adds to the degradation with respect to DM due to the spread of the step sizes employed at decoding. From (29), the performance is now given by

$$P_e = \int_0^{\infty} f_{g(Z)}(s) \sum_{m=-\infty}^{\infty} p_m(s) \cdot \int_{-\infty}^{\infty} P_e[s, y] \times \frac{2}{|4m+1|\Delta} f_N\left(\frac{y}{\frac{4m+1}{2}\Delta} - s\right) dy\, ds. \tag{32}$$

Note that $f_{g(Z)}(s) = f_{|Z|}(s)$ can be approximated by the probability density function of $|Y|$ for large DNR values, as justified in Appendix C. The first integral in (32) expresses the degradation due to the spread of the step sizes employed by the decoder, whereas the integral in y takes into account the jitter of the step size estimated by the decoder.

C. Selection of the Parameter p

One interesting question concerns the selection of the parameter p that determines the ℓ_p-norm used in the definition of $g(\mathbf{z}_{k-1})$ in (23). An appealing approach is to determine the mean and variance of the random variable $g(\tilde{\mathbf{Z}})$, here denoted as $\mu_g(p)$ and $\sigma_g^2(p)$, respectively. Obviously, $g(\tilde{\mathbf{z}}) = \mu_g(p) + \epsilon$, where ϵ is a zero-mean noise term with variance $\sigma_g^2(p)$. Recall that the decoding rule amounts to quantizing z_k with a uniform quantizer with step size $g(\mathbf{z}_{k-1})\Delta = \mu_g(p)\Delta + \epsilon_k\Delta$. Then, if the discrepancy between the quantization steps used at embedding and decoding is small (which occurs for large L, as we have just seen), the only difference with the standard DM decoder will be due to the presence of the term ϵ_k.

It can be shown that the minimization of the power of this noise for a fixed embedding distortion is equivalent to minimizing the ratio $\sigma_g^2(p)/\mu_g^2(p)$. Moreover, when Z_k follows a *generalized Gaussian distribution* with shape parameter $c \in [1, \infty)$ and variance σ_z^2, i.e.,

$$f_{Z_k}(z_k) = \frac{c}{2\sigma_z} \frac{\Gamma^{1/2}(3/c)}{\Gamma^{3/2}(1/c)} \exp\left(-\frac{|z_k|^c}{\sigma_z^c \left(\frac{\Gamma(1/c)}{\Gamma(3/c)}\right)^{c/2}}\right) \tag{33}$$

it is possible to formally prove that the optimal value of p, which is denoted by p^*, is such that $p^* \to c$, as $L \to \infty$. The rather lengthy proof is omitted here due to space restrictions. Noting that for large L and moderate p the absolute moments of Z_k are very close to those of X_k (see Appendix C) and knowing that the proof is solely based on those moments, it is possible to conclude that a judicious choice of p is to match the shape parameter of the host feature signal, provided that this can be modeled as a generalized Gaussian random variable. For $c < 1$, it turns out that a reasonable election is $p = 1$.

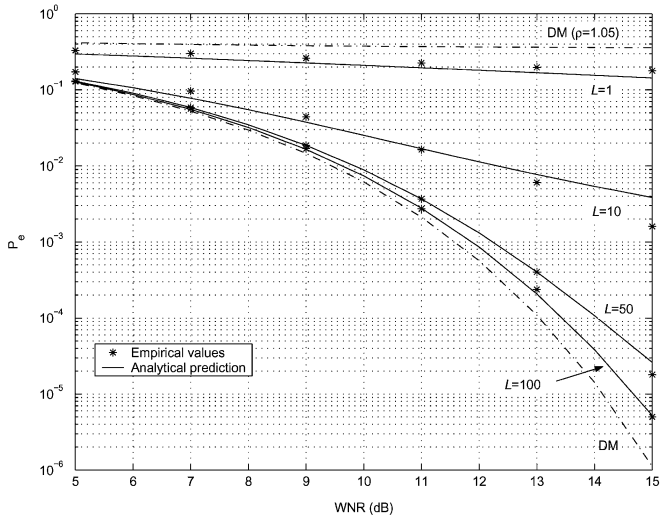

Fig. 5. Empirical and analytical values of the probability of error for different values of the memory size L. Gaussian host DWR $= 25$ dB, $c = 2, p = 2$.

VI. Experimental Results

Our first goal in this section is to illustrate how effective RDM is in approaching the performance of ideal DM as the order L increases. In this context, ideal DM is associated with the absence of gain attack (i.e., $\rho = 1$), thus assuming that the step size at both embedding and decoding is the same. Fig. 5 shows the empirical and analytical values of RDM in the Gaussian case

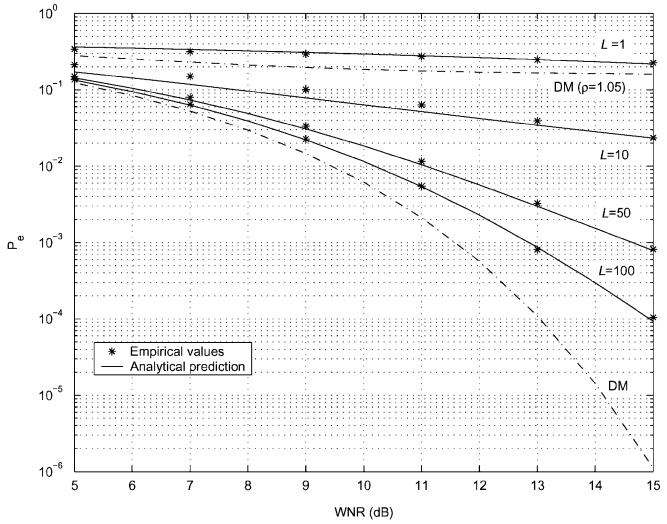

Fig. 6. Empirical and analytical values of the probability of error for different values of L. Generalized Gaussian host DWR $= 25$ dB, $c = 0.5, p = 1$.

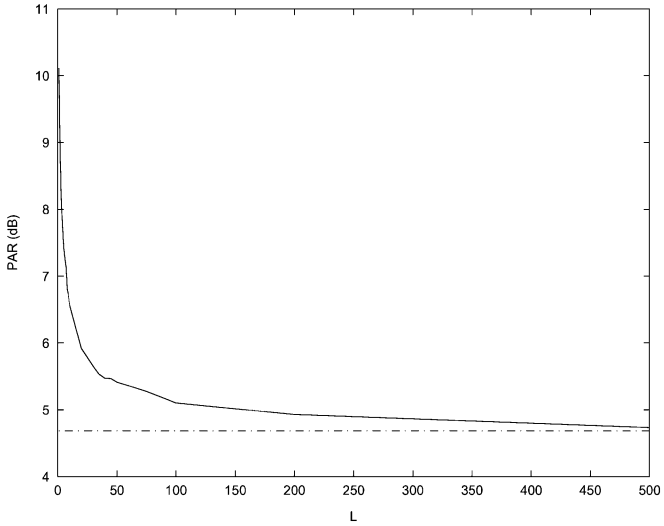

Fig. 7. Empirical embedding peak-to-average ratio (PAR), as a function of the memory size L, for a Gaussian host. DWR $= 25$ dB, $p = 2$.

$(c = 2)$, if we use the ℓ_2 norm in the definition of normalization function $g(\cdot)$.[4] The performance of DM in the ideal case, and assuming a gain attack of 5%, i.e., $\mathbf{Z} = 1.05(\mathbf{Y} + \mathbf{N})$, is plotted as a reference. The measured probability of error was established after averaging the results of several simulations with 50 000 bits each. In all cases, the number of simulations for each WNR and L was large enough to accumulate at least 50 bit errors for statistical significance. Several comments are in order. First, the analytical expressions of Section V for the case $L = 1$ and the large L setting have been used. Although the predictions are good, there are discrepancies at some points as a consequence of the approximations made in Section V. In particular, the use of $f_{g(\tilde{\mathbf{Z}})}(s)$ in (30) loses validity for L getting smaller. Moreover, we must keep in mind the approximate modeling of the noise-driven jitter in the estimated step size $g(\tilde{\mathbf{Z}})$ for $L = 1$. Second, it is important to note that we can reduce the performance loss with respect to DM as much as desired for L sufficiently high. This is especially relevant if we consider that DM is useless beyond a small degree of gain attack, unless some other measures are taken into account, as shown in the figure for a channel gain of $\rho = 1.05$. From Fig. 6, we can draw similar conclusions for a non-Gaussian host following a Generalized Gaussian distribution with shape parameter $c = 0.5$; see (33). Owing to the non-Gaussianity of the host, we cannot use the analytical distribution of the watermarked signal Y derived in Appendix A, and therefore, the pdf of $g(Z)$, which is needed for the $L = 1$ case, is not available in closed-form, its determination constituting, in fact, an open line of research. Therefore, we have used the empirical distribution of $g(Z)$ to compute the probability of error in (32). The agreement is again quite good between the analytical values and the empirically measured probability of error.

Due to its varying step size, one reasonable concern about RDM is the peak embedding distortion, which may produce a considerable perceptual impact in practical applications. As is well known, for large values of the DWR, DM produces a watermark that is uniformly distributed in $[-\Delta, \Delta)$; therefore, its peak-energy value is Δ^2; however, in RDM, the step size may be momentarily large, thus leading to a watermark sample that will also be too large. On the other hand, one should expect that by increasing L, the step-size variation will be smoothed and the watermark peak-energy reduced. In order to measure this effect, we have borrowed from communications the so-called *peak-to-average ratio* (PAR), which is defined as

$$\text{PAR} = \frac{w_+^2}{\sigma_w^2}$$

where w_+ is the magnitude exceeded by 1% of the watermark samples,[5] i.e., $\Pr\{|W| \geq w_+\} = 0.01$. For a uniformly distributed watermark in $[-\Delta, \Delta)$, we have $w_+ = 0.99\Delta$, and therefore, PAR $= (0.99)^2 \cdot 3$, i.e., 4.68 dB. Fig. 7 shows the embedding PAR as a function of L, for a Gaussian host and a DWR of 25 dB, demonstrating the benefits of large values of L in decreasing the embedding PAR. As expected, when $L \to \infty$, the PAR tends to 4.68 dB because in the limit, the pdf of the watermark will be uniform. Remarkably, for $L = 4$, the PAR is already at less than 3 dB from this asymptote (also shown in the figure). This means that for $L = 4$, a *back-off* of 3 dB in the embedding distortion (achieved by reducing Δ) will yield approximately the same peak energy as conventional DM. Of course, this back-off can be arbitrarily reduced by increasing L, as shown in Fig. 7.

Fig. 8 represents the probability of bit error for different values of p when the host samples are drawn from a generalized Gaussian pdf with shape parameter $c = 1.5$. The minimum of P_e is achieved for p close to 1.5, thus confirming that $p = c$ is a good choice as long as $c \geq 1$, as discussed in Section V-C. In any case, the variation in the BER is small for the values of p considered here, showing that performance is quite robust against mismatches with c.

Regarding the comparison of RDM with other methods that are also designed to cope with gain attacks, it must be said

[4]Evidently, for the case $L = 1$, the parameter p of the ℓ_p norm is irrelevant.

[5]The amplitudes of such small set of samples can be clipped to w_+ with little impact on the final performance.

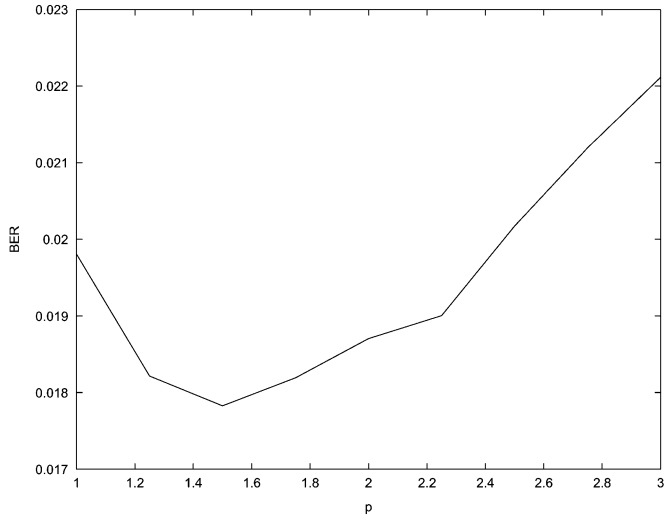

Fig. 8. BER versus p when the host follows a generalized Gaussian with shape parameter $c = 1.5$. DWR $= 45$ dB, WNR $= 10$ dB, $L = 20$.

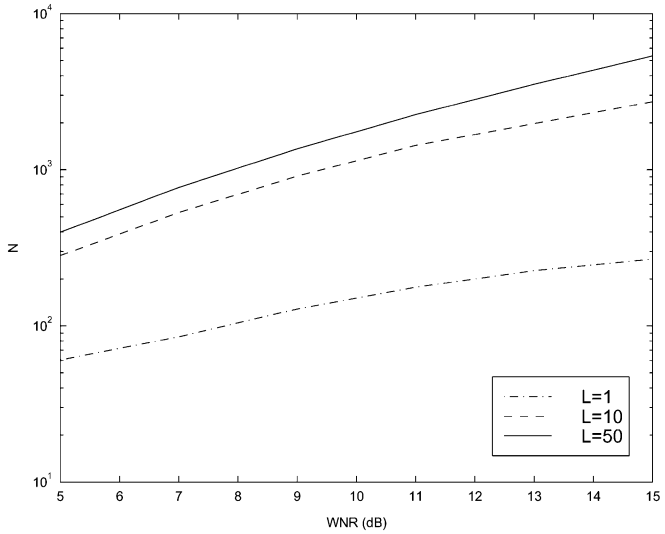

Fig. 9. Spreading factor N needed in ISS to achieve the same BER as in scalar RDM, for different values of L. Gaussian host, DWR $= 25$ dB, $p = 2$.

that it is quite difficult to find a set of fair conditions to establish such a comparison. For instance, those methods based on spherical codewords cannot be implemented in scalar form; in a similar vein, pilot-based schemes waste some payload (on the order of 1000 samples in [7]) for embedding some signals that are deterministically known to both encoder and decoder: a loss RDM does not incur. Nevertheless, RDM can be evenly likened with the recently proposed Improved Spread Spectrum (ISS) [18], which can be seen as spread-spectrum with distortion-compensation, and potentially can work in a scalar fashion. ISS has been proposed as a simple and gain-invariant alternative to QIM (DM) methods. However, contrary to what is stated in [18], it can be shown that the performance of ISS and DM is similar only for very small WNRs, whereas for moderate and large WNRs, DM outperforms ISS (but is prone to gain attacks). Indeed, as shown in [19], for ISS to work reasonably well, a very small (i.e., much less than 0 dB) effective DWR is necessary, something that can only be achieved with spreading. This practically implies that for the scalar binary case and a DWR of 25 dB, the probability of bit error afforded by ISS is always larger than 0.43 for the range of WNRs considered in Fig. 5. Since ISS needs some amount of spreading for attaining a satisfactory performance, it is interesting to plot the spreading factor N that is required to achieve *the same* probability of bit error as RDM for different values of the memory size L. This is plotted in Fig. 9 for a DWR of 25 dB, where it can be seen that for $L = 10$, and depending on the WNR, a spreading factor between 300 and 3000 is necessary. This simply means that for a WNR of 15 dB, ISS reduces the payload by a factor of 3000 compared with RDM to achieve the same BER.

It is important to remark that the results for ISS presented here have been obtained using the optimal distortion compensation strategy (which requires prior knowledge at the embedder about the attacking strength), while we are *not* using distortion compensation (a source of potential improvements). Furthermore, whereas RDM can be combined with any multilevel constellation, for ISS to be robust against gain attacks, binary antipodal symbols are mandatory.

After our manuscript was submitted, a method related to RDM proposed by Oostven *et al.* [20] was brought to our attention. This method can be regarded as replacing $g(\mathbf{y}_{k-1})$ in (21) by $g'(\mathbf{x}_{k+L}) \triangleq (1/2L+1)\sum_{i=-L}^{L} x_{k-i}$ and $g(\mathbf{z}_{k-1})$ in (22) by $g'(\mathbf{z}_{k+L}) \triangleq (1/2L+1)\sum_{i=-L}^{L} z_{k-i}$, which can both be recognized as moving averages. As it can be readily checked, this scheme is also gain invariant and has the additional advantage of not needing to impose a strict-causality constraint on the argument of function $g'(\cdot)$. Unfortunately, even in the absence of channel noise, the method in [20] presents a nonzero probability of error due to using $\mathbf{x}$ (instead of $\mathbf{y}$) in embedding, whereas $\mathbf{y} = \mathbf{x} + \mathbf{w}$ is employed in decoding. Moreover, when the mean of the host signal is small compared with its standard deviation, experiments reveal that Oostven *et al.*'s algorithm is noticeably outperformed by RDM because for the former, the mean of $g'(\tilde{\mathbf{Z}})$ will also be small compared to its standard deviation, resulting in a significant amount of step-size jitter. On the other hand, for host mean values larger than the standard deviation, both schemes achieve similar performance.

RDM has been conceived to fight fixed gain attacks. Then, one logical matter of interest is the degradation of RDM performance under varying gain attacks. From its construction, it can be expected that RDM should be effective in handling slow variations in the gain ρ_k. In order to establish how slow this variations would be, we have considered the following model for ρ_k:

$$\rho_k = 1 + 0.1\cos\left(\frac{2\pi k}{L \cdot M}\right) \tag{34}$$

in which the variation is controlled by M. It must be stressed that unless they are modified to track these gain variations, at the expense of considerable increase of complexity, those methods that estimate the *fixed* quantization step size will fail against a varying gain attack such as (34). On the other hand, those schemes that use binary antipodal constellations, like spread-spectrum or ISS, will show much better performance. As for RDM, in Fig. 10, the BER values obtained experimentally for

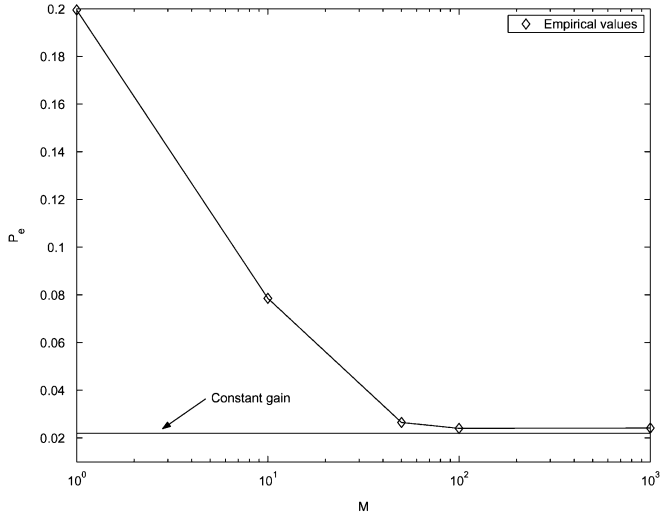

Fig. 10. BER performance of RDM for a varying gain attack ρ_k as in (34), where N is a parameter that expresses the degree of change. DWR $= 25$ dB, WNR $= 15$ dB, $L = 10, p = 1, c = 0.5$.

fixed DWR and WNR, when $L = 10$, and M is allowed to change, are shown. The results were obtained after averaging 10 realizations of 25 000 bits each. As can be clearly seen, for values of M larger than 100, the achieved performance is almost the same as with a fixed gain attack. Similar conclusions can be drawn for other values of L: The degradation coming from the proposed varying gain model becomes negligible for M sufficiently high. Therefore, there will typically be a compromise between performance (which improves with increasing L) and the ability to cope with faster gain variations (which requires L being small). We can conclude that RDM has a good degree of robustness against varying gains, provided they are sufficiently slow. In any case, an in-depth discussion of the behavior of RDM for varying-gain attacks is out of the scope of the paper.

VII. General Setup for Gain-Invariant Quantization-Based Methods

Here, we give a general formulation of quantization-based methods that are invariant to fixed gain attacks, which considerably generalize RDM. We do this by modifying the original QIM proposal by Chen and Wornell [3] to define a Gain-Invariant QIM (GI-QIM). We assume that the message to be embedded m is chosen from the set $\mathcal{M} = \{1, 2, \ldots, |\mathcal{M}|\}$. Let $\mathbf{y}_k^f \triangleq (y_k, y_{k-1} \cdots, y_1)^T, k = 1, \ldots, N$ and with $\mathbf{y}_0^f = (1, 1, \ldots, 1)^T$. Consider now the following embedding rule

$$y_k = g_k\left(\mathbf{y}_{k-1}^f\right) Q_m^{(k)}\left(t_k\left(\mathbf{x}, \mathbf{y}_{k-1}^f\right)\right), \quad k = 1, \ldots, N \tag{35}$$

where the functions g_k and t_k are such that $g_k : \mathbb{R}^{k-1} \to \mathbb{R}, t_k : \mathbb{R}^{N+k-1} \to \mathbb{R}^N$, and $Q_m^{(k)}(\cdot)$ is a message-dependent quantizer that for each k takes a sample in $\mathbb{R}^N$ and outputs a scalar. Formula (35) is very general and, in fact, includes QIM as a particular case in which $g_k = 1$ and $t_k(\mathbf{x}, \mathbf{y}_{k-1}^f) = \mathbf{x}$. As in [3], we will consider only minimum Euclidean distance decoding, although more sophisticated, e.g., Maximum Likelihood (ML) algorithms can be employed as well. Given the received signal $\mathbf{z} \in \mathbb{R}^N$, the minimum Euclidean distance decoder uses the following rule:

$$\hat{m} = \arg\min_{m \in \mathcal{M}} \sum_{k=1}^{N} \left(z_k - g_k\left(\mathbf{z}_{k-1}^f\right) Q_m^{(k)}\left(t_k\left(\mathbf{z}, \mathbf{z}_{k-1}^f\right)\right)\right)^2 \tag{36}$$

where $\mathbf{z}_k^f$ is defined in a similar way as $\mathbf{y}_k^f$.

As we did in Section III with DM, it is not difficult to show that in general, the data-hiding scheme just described is not robust against gain attacks. To see this, it is enough to check that the output of the decoder in (36) depends on ρ when the the received signal is $\mathbf{z} = \rho(\mathbf{y} + \mathbf{n})$.

In order to achieve a truly gain-invariant scheme, one must require that certain constraints upon functions g_k and t_k above are satisfied, namely, for any $\rho > 0, g_k(\rho \mathbf{y}_k^f) = \rho g_k(\mathbf{y}_k^f), t_k(\rho \mathbf{x}, \rho \mathbf{y}_k^f) = t_k(\mathbf{x}, \mathbf{y}_k^f)$. A simple way of achieving this is to simultaneously impose the following set of sufficient conditions:

$$\begin{aligned} g_k\left(\rho \mathbf{y}_k^f\right) &= \rho g_k\left(\mathbf{y}_k^f\right) \\ t_k\left(\rho \mathbf{x}, \mathbf{y}_k^f\right) &= \rho t_k\left(\mathbf{x}, \mathbf{y}_k^f\right) \\ t_k\left(\mathbf{x}, \rho \mathbf{y}_k^f\right) &= \rho^{-1} t_k\left(\mathbf{x}, \mathbf{y}_k^f\right) \end{aligned} \tag{37}$$

for all $k = 1, \ldots, N$. Now, substituting conditions (37) into (36), it is possible to show that

$$\begin{aligned} &\arg\min_{m \in \mathcal{M}} \sum_{k=1}^{N} \left(\rho y_k - g_k\left(\rho \mathbf{y}_{k-1}^f\right)\right. \\ &\quad \left. \times Q_m^{(k)}\left(t_k\left(\rho \mathbf{y}, \rho \mathbf{y}_{k-1}^f\right)\right)\right)^2 \\ &= \arg\min_{m \in \mathcal{M}} \rho \sum_{k=1}^{N} \left(y_k - g_k\left(\mathbf{y}_{k-1}^f\right)\right. \\ &\quad \left. \times Q_m^{(k)}\left(t_k\left(\mathbf{y}, \mathbf{y}_{k-1}^f\right)\right)\right)^2 \end{aligned}$$

and therefore, the decoder output does not depend on $\rho, \rho > 0$.

This construction can be specialized to many cases of practical interest. For instance, a general GI-DM follows by designing quantizers $Q_m^{(k)}(\cdot)$ by simply shifting a base quantizer $Q^{(k)}(\cdot)$. Particular cases of GI-DM other than the binary RDM can then be designed: multilevel scalar RDM, block-based RDM, and spread-transform RDM, all of which can be combined with distortion compensation. For the sake of brevity, the details on these schemes are omitted and will be found elsewhere.

VIII. Conclusion

Quantization-based methods are becoming increasingly popular in watermarking and data-hiding applications, owing to their improved performance, as compared with spread-spectrum-based schemes. Unfortunately, the fixed gain attack has revealed itself as the Achilles' heel of quantization-based algorithms, due to its overwhelming simplicity and devastating

effects. In this paper, we have presented RDM, which is a novel data-hiding method that is invariant to fixed gain attacks and does not require estimating the step-size, as most existing methods do. RDM constructs a gain-invariant domain in which quantization takes place, and it does so in a extremely simple way, amounting to minor modifications of the standard DM method. In fact, RDM may work in a scalar fashion, thus largely reducing the complexity of those algorithms based on spherical codewords, which are quite difficult to deal with.

We have also provided a thorough statistical analysis of RDM, which, due to its asymptotic stationarity, demands the usage of formal tools to determine the stationary pdf of the watermarked signal. This paved the way for analytically assessing the performance of RDM. As we have seen, the basic RDM formulation can be improved by increasing the memory of the system in such a way that the performance of RDM asymptotically tends to that of DM. Experimental results have validated our analysis and confirmed the predicted behavior of RDM. Furthermore, we have shown that RDM is moderately resilient to varying gain attacks. Finally, a family of gain-invariant methods that retains the spirit of the QIM paradigm has been briefly discussed and will be reported elsewhere in more detail.

The current RDM proposal has proven its merits in a highly theoretical context, which is largely independent of the host nature; needless to say, a considerable amount of work is necessary to tune RDM to the demands of practical applications. For instance, our assumption of a stationary host is far from holding with real signals. A partial relief would be to pseudorandomly permute the host samples to create a "pseudo-stationary" signal, but in turn, this may affect RDM's resilience to slow-varying gains. Other open implementation issues that need be taken into consideration include the use of varying embedding strengths (induced by human perception), the study of practical effects of the embedding PAR, the selection of the quantization step, and the initialization of the function g.

Besides the practical implementation of RDM with multimedia signals, ongoing research covers the design of the function g controlling the step size, with the possible inclusion of weighted ℓ_p norms, the combination of RDM with distortion compensation and channel coding, and the adaptation of RDM to deal with faster gain variations.

APPENDIX A
STATIONARY PDF OF $\{Y_k\}$ FOR GAUSSIAN HOSTS

For a Gaussian pdf with zero mean and variance σ_x^2, we have

$$p_m(y,+1) = \mathcal{Q}\left(\frac{\Delta|y|(4m-1)}{2\sigma_x}\right) - \mathcal{Q}\left(\frac{\Delta|y|(4m+3)}{2\sigma_x}\right) \tag{38}$$

where $\mathcal{Q}(x) \triangleq (1/\sqrt{2\pi})\int_x^\infty e^{-t^2/2}\,dt$. Substituting (38) into (17) gives

$$f_Y(y) = \sum_m \frac{2}{|4m+1|\Delta}\left(\mathcal{Q}\left(\frac{|y|(4m-1)}{|4m+1|\sigma_x}\right) - \mathcal{Q}\left(\frac{|y|(4m+3)}{|4m+1|\sigma_x}\right)\right) f_Y\left(\frac{|y|}{|2m+1/2|\Delta}\right). \tag{39}$$

Exactly solving the functional equation (39) seems to be an extremely difficult task. Some approximations that allow us to arrive at more manageable expressions are presented next. A first approximation arises after considering that when $y_1 - y_0 \ll \sigma$ and $y_0 < y_1$

$$\mathcal{Q}\left(\frac{y_0}{\sigma}\right) - \mathcal{Q}\left(\frac{y_1}{\sigma}\right) \approx \frac{(y_1-y_0)}{\sqrt{2\pi}\sigma} e^{-(y_0+y_1)^2/8\sigma^2}.$$

This allows to write (39) as

$$f_Y(y) \approx \frac{8|y|}{\sqrt{2\pi}\sigma_x\Delta} e^{-y^2/2\sigma_x^2} \times \sum_m \frac{1}{|4m+1|^2} f_Y\left(\frac{2|y|}{|4m+1|\Delta}\right)$$

which, noting that the absolute value of the terms involving m is being taken, leads to

$$f_Y(y) \approx \frac{8|y|}{\sqrt{2\pi}\sigma_x\Delta} e^{-y^2/2\sigma_x^2} \times \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} f_Y\left(\frac{2|y|}{(2m+1)\Delta}\right). \tag{40}$$

Although it is in a more explicit form, this approximate functional equation still looks very difficult to solve. However, by virtue of the aperiodicity of the associated Markov chain (see Section III-C), asymptotic convergence to the stationary pdf is attained for any initial pdf by just iterating on (40), provided that the conditions for convergence are met. Unfortunately, iterating on (40) quickly becomes infeasible because of the combinatorial explosion evident in (40). Alternatively, one may try to initialize the recursion with a pdf that is sufficiently close to the stationary solution and then perform a single iteration on (40). Noticing that our interest lies in high DWRs (which imply small values of Δ) and that, for very small values of Δ, the pdf of Y should resemble that of X, we start by conjecturing a Gaussian distribution for Y, which we will denote by $f_{Y,0}(y)$, with zero-mean and variance σ_x^2. Substituting this pdf into the right-hand side of (40), we obtain the pdf after the first iteration $f_{Y,1}(y)$

$$f_{Y,1}(y) = \frac{4|y|}{\pi\sigma_x^2\Delta} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} \times \exp\left(-\frac{2y^2}{\sigma_x^2\Delta^2(2m+1)^2} - \frac{y^2}{2\sigma_x^2}\right). \tag{41}$$

For small values of Δ, the series in (41) converges very rapidly, allowing us to keep just a few terms for practical computations. In addition, it is not difficult to show that in the limit, when $\Delta \to 0, f_{Y,1}(Y) = f_X(Y)$, thus confirming the fact that for very small Δ, Y_k practically holds the same distribution as X_k.

Next, we show that $f_{Y,1}(y)$ corresponds to a valid pdf. Obviously, $f_{Y,1}(y) \geq 0$, for all $y \in \mathbb{R}$. In addition, it is possible to show that

$$\int_{-\infty}^{\infty} f_{Y,1}(y)\,dy = \tanh(\pi/\Delta)$$

which is very close to 1 for small $\Delta(\Delta < 0.5)$.

It is also illustrative to compute the variance that corresponds to $f_{Y,1}(y)$ and compare it with σ_y^2 in (10). In this case, the following can be shown:

$$\int_{-\infty}^{\infty} y^2 f_{Y,1}(y)\,dy = \sigma_x^2\left(\frac{\pi}{\Delta}\mathrm{sech}^2(\pi/\Delta) + \tanh(\pi/\Delta)\right)$$

which is very close to σ_x^2 for small $\Delta(\Delta < 0.5)$. Thus, the variance of the random variable with pdf $f_{Y,1}(y)$ is approximately σ_x^2, regardless of Δ, as opposed to (10). This suggests a further refinement for the approximation to $f_Y(y)$ by using $\sigma_y^2 = \sigma_x^2/(1-\Delta^2/3)$ instead of σ_x^2 in (41). With all this, we propose the approximation given in (18).

For $\Delta > 0.5$, it turns out that setting a two-sided Rayleigh as the initial state, i.e., $f_{Y,0}(y) = (|y|/\sigma_x^2)e^{-y^2/2\sigma_x^2}$, provides more accurate results when substituted into the right-hand side of (40) to produce $f_{Y,1}(y)$.

Appendix B
Irreducibility and Aperiodicity of the Markov Chain

We show here that from an arbitrary state $\hat{y}_k$ of the Markov chain, it is possible to reach any other state in a finite number of steps. First, we will prove that for $\hat{y}_k = \epsilon$, i.e., $\hat{y}_k$ corresponding to the positive quantization interval closest to zero, it is possible to reach any other quantization interval. Second, we will show that this first quantization interval can be reached from any other, including both positive and negative intervals. Thus, the first quantization interval is the hub guaranteeing all the possible transfers among the states of the Markov chain, and as a consequence, the Markov chain given by (20) is irreducible.

Given $\hat{y}_k = \epsilon$, we have that $\hat{y}_{k+1}$ is the result of quantizing $\epsilon(\Delta/2)(2m+1)$ for some integer m. If the offset ϵ is such that $\epsilon\Delta < \delta$, then there exists an integer m for which $\epsilon(2m+1)\Delta/2$ falls within an arbitrary quantization interval $[l\delta, (l+1)\delta)$, with $|l| < S$. Therefore, it is possible to jump from $\hat{y}_k = \epsilon$ to $\hat{y}_{k+1} = l\delta + \epsilon$ in one transition, or what is truly important, in a finite number of steps. For the second part of the proof, we show that it is possible to achieve the first quantization interval from any other: If $\hat{y}_n = l\delta + \epsilon, l > 0$, then one of the possible transitions is given by the quantization of $\hat{y}_n\Delta/2 = (l\delta+\epsilon)\Delta/2$, which will produce a state closer to zero provided that

$$\Delta < \frac{2l\delta}{l\delta + \epsilon}. \tag{42}$$

For $l < 0$, similar considerations apply after the transition given by the quantization of $(-\Delta/2)\hat{y}_n$. Thus, (42) holds true for $l \neq 0$ as long as the RDM quantization step Δ is less than 2, as ϵ can be chosen arbitrarily small. Therefore, for finite l, the state closest to zero can be reached in at most l steps. In addition, for $l = 0$, it is possible to remain at this state, i.e., to achieve $\hat{y}_{k+1} = \hat{y}_k$. The existence of this aperiodic state guarantees the aperiodicity of the whole Markov chain [14].

Appendix C
Influence of the Additive Noise on the Absolute Moments

We want to analyze the impact of the additive noise on the absolute moments of $Z = Y + N$ with respect to those of Y or, equivalently, of X if we use the approximation $M_{yp} \approx M_{xp}$. The analysis will apply to the asymptotic case $L \to \infty$ for which we can write $f_Y(y)$, as in (27):

$$f_Y(y) \approx \sum_{m=-\infty}^{\infty} q_m \delta\left(y - (2m+1)\frac{\Delta}{2}M_{xp}^{1/p}\right).$$

We want to compute the pth absolute moment of Z, i.e., $M_{zp} = \int_{-\infty}^{\infty}|z|^p f_Z(z)\,dz$, or, equivalently

$$M_{zp} = \sum_{m=-\infty}^{\infty} q_m \int_{-\infty}^{\infty}|z|^p f_N\left(z - (2m+1)\frac{\Delta}{2}M_{xp}^{1/p}\right)dz. \tag{43}$$

As we will show below, for Gaussian distributions, the above integral can be handled straightforwardly. Otherwise, we can use a second-order expansion of $|z|^p$ around the points $(2m+1)\Delta/2M_{xp}^{1/p}$. If $\sigma_n \ll \Delta M_{xp}^{1/p}$, then the pdf $f_N(z - \Delta M_{xp}^{1/p}/2)$ will be small at $z = 0$, weighting down the errors of the approximation of $|z|^p$ near the origin.

$$\begin{aligned}|z|^p \approx{}& \left|\frac{2m+1}{2}\Delta\right|^p M_{xp} \\ &+ \mathrm{sgn}(2m+1)p\left|\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right|^{p-1} \\ &\times\left(z - \frac{2m+1}{2}\Delta M_{xp}^{1/p}\right) \\ &+ \frac{p(p-1)}{2}\left|\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right|^{p-2} \\ &\times\left(z - \frac{2m+1}{2}\Delta M_{xp}^{1/p}\right)^2.\end{aligned}$$

This series expansion makes it easy to approximate the integrals in (43) as

$$\begin{aligned}&\int_{-\infty}^{\infty}|z|^p f_N\left(z - \frac{2m+1}{2}\Delta M_{xp}^{1/p}\right)dz \\ &\quad\approx \left|\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right|^p \\ &\qquad + \frac{p(p-1)}{2}\left|\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right|^{p-2}\sigma_n^2.\end{aligned}$$

The pth absolute moment of Z is then approximately given by

$$\begin{aligned}M_{zp} \approx \sum_m q_m \Bigg(&\left|\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right|^p \\ &+\frac{p(p-1)}{2}\left|\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right|^{p-2}\sigma_n^2\Bigg).\end{aligned}$$

For small Δ, we can use that q_m in (26) is well approximated by

$$q_m \approx \Delta M_{xp}^{1/p} f_X\left(\frac{2m+1}{2}\Delta M_{xp}^{1/p}\right)$$

and substitute the sums in M_{zp} by integrals

$$\begin{aligned} M_{zp} &\approx \int_{-\infty}^{\infty} f_X(x)|x|^p\,dx \\ &\quad + \int_{-\infty}^{\infty} f_X(x)\frac{p(p-1)}{2}|x|^{p-2}\sigma_n^2\,dx \\ &= M_{xp}\left(1+\frac{p(p-1)}{2}\frac{\sigma_n^2}{M_{xp}}M_{x(p-2)}\right). \end{aligned} \tag{44}$$

Therefore, the pth absolute moment of Z reads as

$$M_{zp} \approx M_{xp}\left(1+\frac{p(p-1)}{2}\frac{\sigma_n^2}{M_{xp}}M_{x(p-2)}\right).$$

Note that for $p=2$, we have

$$E\{|Z|^2\} = \sigma_x^2+\sigma_n^2 = \sigma_x^2\left(1+\frac{1}{\text{DNR}}\right) \tag{45}$$

as expected from the sum of two independent random variables and the approximation $\sigma_y^2 \approx \sigma_x^2$, which is valid for small Δ.

The approximation for M_{zp} degrades as p gets away from 2. Thus, for large p, the second-order polynomial expansion of $|z|^p$ is no longer valid. Even more, for low p, the second integral in (44) is a poor representation of the corresponding sum, given the singularity of $|x|^{p-2}$ at $x=0$. For Gaussian distributions, the integrals in (43) can be handled analytically:

$$\begin{aligned} \int_{-\infty}^{\infty} |z|^p f_N(z-\Delta[m])\,dz &= \sigma_n^p\Gamma\left(\frac{1+p}{2}\right)\frac{2^{p/2}}{\sqrt{\pi}} \\ &\quad \times e^{-\Delta^2[m]/2\sigma_n^2}{}_1F_1\left(\frac{1+p}{2},\frac{1}{2},\frac{\Delta^2[m]}{2\sigma_n^2}\right) \end{aligned}$$

where $\Delta[m] \triangleq (2m+1/2)\Delta M_{xp}^{1/p}$, and ${}_1F_1(\cdot,\cdot,\cdot)$ is the Kummer confluent hypergeometric function [21]. As a consequence, the pth absolute moment of Z is

$$\begin{aligned} M_{zp} &= \sigma_n^p\Gamma\left(\frac{1+p}{2}\right)\frac{2^{p/2}}{\sqrt{\pi}} \\ &\quad \times \sum_{m=-\infty}^{\infty} q_m e^{-\Delta^2[m]/2\sigma_n^2}{}_1F_1\left(\frac{1+p}{2},\frac{1}{2},\frac{\Delta^2[m]}{2\sigma_n^2}\right) \end{aligned}$$

where no approximation has been performed besides the large L assumption. Now, for a Gaussian pdf, $f_X(x)$, q_m is approximately given by $\Delta M_{xp}^{1/p} f_X(\Delta[m])$ and M_{zp} by

$$\begin{aligned} M_{zp} &\approx \frac{\sigma_n^p}{\sigma_x}\Gamma\left(\frac{1+p}{2}\right)\frac{2^{(p-1)/2}}{\pi} \\ &\quad \times \int_{-\infty}^{\infty} e^{-(1/\sigma_x^2+1/\sigma_n^2)x^2/2}{}_1F_1\left(\frac{1+p}{2},\frac{1}{2},\frac{x^2}{2\sigma_n^2}\right)dx \end{aligned}$$

where the sum has been substituted by an integral. The integration of the product of the exponential and hypergeometric function admits an analytical solution since

$$\begin{aligned} \int_{-\infty}^{\infty} e^{-(1/\sigma_x^2+1/\sigma_n^2)x^2/2}{}_1F_1\left(\frac{1+p}{2},\frac{1}{2},\frac{x^2}{2\sigma_n^2}\right)dx \\ = \sqrt{2\pi}\sigma_x\left(1+\frac{\sigma_x^2}{\sigma_n^2}\right)^{p/2} \end{aligned}$$

yielding a final expression for M_{zp} as follows:

$$M_{zp} \approx M_{xp}\left(1+\frac{\sigma_n^2}{\sigma_x^2}\right)^{p/2} = M_{xp}\left(1+\frac{1}{\text{DNR}}\right)^{p/2}$$

which, for $p=2$, boils down to the previous result (45). M_{xp} denotes the pth absolute moments of a Gaussian random variable, which turn out to be

$$M_{xp} = \frac{1}{\sqrt{2\pi}}\sigma_x^p 2^{(1+p)/2}\Gamma\left(\frac{1+p}{2}\right).$$

REFERENCES

[1] I. J. Cox, M. L. Miller, and A. L. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul. 1999.

[2] B. Chen and G. Wornell, "Achievable performance of digital watermarking schemes," in *Proc. IEEE Int. Conf. Multimedia Comput. Syst.*, vol. 1, Florence, Italy, June 1999, pp. 13–18.

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

[4] S. I. Gelf'and and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Contr. Inf. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

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

[6] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*. New York: Springer-Verlag, 1988.

[7] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Trans. Signal Process.*, vol. 4, no. 51, pp. 1003–1019, Apr. 2003.

[8] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jul. 2002.

[9] F. Bartolini, M. Barni, and A. Piva, "Performance analysis of spread transform dither modulation (STDM) watermaking in presence of non-additive attacks," *IEEE Trans. Signal Process.*, to be published.

[10] J. J. Eggers, R. Bäuml, and B. Girod, "Estimation of amplitude modifications before SCS watermark detection," in *Proc. SPIE Security Watermarking Multimedia Contents*, P. W. Wong and E. J. Delp, Eds. San Jose, CA, Jan. 2002, vol. 4675, pp. 387–398.

[11] K. Lee, D. S. Kim, T. Kim, and K. A. Moon, "EM estimation of scale factor for quantization-based audio watermarking," in *Proc. Second Int. Workshop Digital Watermarking*, I. Cox, T. Kalker, and Y. M. Ro, Eds. Seoul, Korea: Springer-Verlag, Oct. 2003, pp. 316–327.

[12] M. L. Miller, G. J. Doerr, and I. J. Cox, "Applying informed coding and embedding to design a robust, high capacity, watermark," *IEEE Trans. Image Process.*, vol. 13, no. 6, pp. 792–807, Jun. 2004.

[13] A. Abrardo and M. Barni, "Orthogonal dirty paper coding for informed watermarking," in *Proc. SPIE Security, Steganography, Watermarking Multimedia Contents*, P. W. Wong and E. J. Delp, Eds. San Jose, CA, Jan. 2004, vol. 5306.

[14] J. R. Norris, *Markov Chains*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[15] D. J. Goodman and A. Gersho, "Theory of an adaptive quantizer," *IEEE Trans. Commun.*, vol. COM-22, no. 8, pp. 1037–1045, Jul. 1974.

[16] W. Feller, *An Introduction to Probability Theory and Its Applications*, Second ed. New York: Wiley, 1971.
[17] L. Pérez-Freire, F. Pérez-González, and S. Voloshynovskiy, "An accurate analysis of scalr quantization-based data hiding," *IEEE Trans. Inf. Forensics Security*, to be published.
[18] H. S. Malvar and D. A. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Process.*, vol. 4, no. 51, pp. 898–905, Apr. 2003.
[19] L. Pérez-Freire, F. Pérez-Gonzalez, and S. Voloshynovskiy, "Revisiting scalar quantization-based data hiding: Exact analysis and results," *IEEE Trans. Signal Processing*, submitted for publication.
[20] J. Oostven, T. Kalker, and M. Staring, "Adaptive quantization watermarking," in *Proc. SPIE Security, Steganography, Watermarking Multimedia Contents VI*, vol. 5306, P. W. Wong and E. J. Delp, Eds., San Jose, CA, Jan. 2004, pp. 296–303.
[21] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Fifth ed. San Diego, CA: Academic, 1994.

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

He joined the faculty of the School of Telecommunications Engineering, University of Vigo, as an assistant professor in 1990, where he is currently Professor. 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, robust control, 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 three special sections of the EURASIP journal *Signal Processing* devoted to signal processing for communications and digital watermarking, as well as Guest Editor of a Feature Topic of the IEEE COMMUNICATIONS MAGAZINE on digital watermarking. He has also co-authored over 25 papers in leading international journals and more than 70 papers published in various conference proceedings.

Prof. Perez-Gonzalez was the Chairman of the Fifth and Sixth Baiona Workshops on Signal Processing in Communications, held in Baiona, Spain, in 1999 and 2003, respectively.

Carlos Mosquera was born in Vigo, Spain, in 1969. He received the undergraduate degree from the Universidad de Vigo, the M.S. degree from Stanford University, CA, in 1994, and the Ph.D. degree from the Universidad de Vigo in 1998, all in electrical engineering.

In 1999, he spent six months with the European Space Agency at ESTEC in the Netherlands. He is currently an Associate Professor at the Universidad de Vigo. His interests lie in the area of signal processing applied to communications.

Mauro Barni (S'90–M'96) graduated in electronic engineering in 1991 and received the Ph.D. degree in informatics and telecommunications in October 1995, both from the University of Florence, Florence, Italy.

From 1991 through 1998, he was with the Department of Electronic Engineering, University of Florence. Since September 1998, he has been with the Department of Information Engineering, University of Siena, Siena, Italy, where he is an associate professor. His main interests are in the field of digital image processing and computer vision. His current research activity is focused on the application of image processing techniques to copyright protection and authentication of multimedia data (digital watermarking). He is author/co-author of more than 130 papers published in international journals and conference proceedings and holds three patents in this field. He is on the editorial board of the *EURASIP Journal of Applied Signal Processing.*

Dr. Barni serves as associate editor of the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE SIGNAL PROCESSING MAGAZINE (Column and Forum section). He is a member of the IEEE Multimedia Signal Processing Technical Committee (MMSP-TC).

Andrea Abrardo received the Dr.Ing. degree in electronic engineering from the University of Florence, Florence, Italy, in April 1993 and the Ph.D. degree, also from the University of Florence, in June 1998, with a thesis on "Telecommunications in Medicine."

From January to November 1994, he was with the Image Processing and Communications Laboratory, Department of Electronic Engineering, University of Florence, collaborating with the Tuscany region for the development of broadband network infrastructures. Since August 1998, he has been with the Department of Information Engineering, University of Siena, Siena, Italy, where he is an Assistant Professor. Presently, he is involved in the activities of the Secure Mobile PAYments and Services On Chip (SMPAYSOC) IST Projects within the fifth Research Framework of the European Commission. His main research interests include the field of computer and communication networks with an emphasis on code-division multiple access for third-generation wireless communications and radio resource management for B3G and *ad hoc* networks. He is also involved in the field of digital watermarking, with a particular emphasis on the design of channel coding techniques for robust information hiding.